

©Copyright 2012

Jonathan Gruhl

Bayesian Modeling for Multivariate Mixed Outcomes with Applications to Cognitive Testing Data

Jonathan Gruhl

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Elena A. Erosheva, Chair

Peter D. Hoff

Thomas S. Richardson

Program Authorized to Offer Degree:
Statistics

University of Washington

Abstract

Bayesian Modeling for Multivariate Mixed Outcomes
with Applications to Cognitive Testing Data

Jonathan Gruhl

Chair of the Supervisory Committee:
Associate Professor Elena A. Erosheva
Statistics And Social Work

This dissertation studies parametric and semiparametric approaches to latent variable models, multivariate regression and model-based clustering for mixed outcomes. We use the term mixed outcomes to refer to binary, ordered categorical, count, continuous and other ordered outcomes in combination. Such data structures are common in social, behavioral, and medical sciences. We first review existing parametric approaches to mixed outcomes in latent variable models before developing extensions to accommodate outcome types specific to cognitive testing data. We subsequently develop two new regression approaches for mixed outcome data, the semiparametric Bayesian latent variable model and the semiparametric reduced rank multivariate regression model. In contrast to the existing parametric approaches, these models allow us to avoid specification of distributions for each outcome type. We apply the latent variable and multivariate regression models to investigate the association between cognitive outcomes and MRI-measured regional brain volumes using data from a study of dementia and compare results from the different models. Finally, we develop a new semiparametric correlated partial membership model for model-based clustering of mixed outcome data that also allows us to avoid specification of outcome distributions. We demonstrate our semiparametric approach to model-based clustering on NBA player data from the 2010-2011 season as well as on cognitive testing data from a study of dementia.

TABLE OF CONTENTS

	Page
List of Figures	iii
List of Tables	vii
Chapter 1: Introduction	1
Chapter 2: Bayesian Methods Background	7
2.1 Introduction	7
2.2 Hybrid Monte Carlo	7
2.3 Software Validation	11
2.4 Posterior Predictive Model Checking	12
Chapter 3: Introduction to the SIVD study	21
Chapter 4: Item Response Theory Model for Mixed Outcomes	29
4.1 Introduction	29
4.2 Adapting IRT Models for Mixed Outcomes	32
4.3 Estimation	40
4.4 A Simulated Data Example	43
4.5 Application to the SIVD Study	45
4.6 Discussion	53
Chapter 5: Semiparametric Latent Variable Model for Mixed Outcomes	56
5.1 Introduction	56
5.2 Semiparametric Latent Variable Model	57
5.3 Estimation	62
5.4 Hierarchical Formulation and Estimation	70
5.5 A Simulated Data Example	72
5.6 Application to the SIVD Study	78
5.7 Discussion	85

Chapter 6:	Semiparametric Multivariate Regression for Mixed Outcomes	91
6.1	Introduction	91
6.2	Model	93
6.3	Estimation	97
6.4	A Simulated Data Example	100
6.5	Application to the SIVD Study	102
6.6	Discussion	106
6.7	Postscript: Model Review	108
Chapter 7:	Semiparametric Correlated Bayesian Partial Membership Model	115
7.1	Introduction	115
7.2	The Bayesian Partial Membership Model	117
7.3	Comparison of Partial Membership and Mixed Membership Structures	119
7.4	Semiparametric Correlated Bayesian Partial Membership Model: Derivation and Estimation Algorithms	131
7.5	A Simulated Data Example	136
7.6	Application to 2010-11 NBA Player Data	144
7.7	Application to the SIVD Study	157
7.8	Discussion	163
Chapter 8:	Conclusion and Future Work	165
8.1	Conclusion	165
8.2	Future Work	167
Bibliography		175
Appendix A:	Induced Priors	188
Appendix B:	Comparison of Correlation Matrix Distance Measures	192

LIST OF FIGURES

Figure Number	Page
2.1 Sample Posterior Predictive Marginal Distribution Checks	14
2.2 Example Pairwise Correlation Plots	16
2.3 Correlation Distance Scatterplot Examples	19
2.4 Sample Eigenvalue Plots	20
3.1 Histograms of SIVD Executive Functioning Outcomes.	25
3.2 Histograms of SIVD MRI-measured Brain Volumes.	26
3.3 Bivariate plots of SIVD executive functioning outcome against MRI-measured brain volumes.	27
4.1 Marginal Distribution PPMC For Parametric IRT Mixed Outcome Model Simulated Dataset	45
4.2 Correlation Distance Scatterplot For Parametric IRT Mixed Outcome Model Simulated Dataset	46
4.3 Eigenvalue Plot For Parametric IRT Mixed Outcome Model Simulated Dataset	47
4.4 Pairwise Rank Correlation Plots For Parametric IRT Mixed Outcome Model Simulated Dataset	48
4.5 Histograms of scores for Mattis Dementia Rating Scale E and W.	48
4.6 Marginal Distribution PPMC for IRT Mixed Outcome model of SIVD data .	51
4.7 Correlation Distance and Eigenvalue Plot PPMC for IRT Mixed Outcome model of SIVD data	52
4.8 Pairwise Correlation Plots PPMC for IRT Mixed Outcome model of SIVD data	53
5.1 Extended Rank Likelihood Applied To Gaussian Copula	60
5.2 Extended Rank Likelihood Applied To Factor Analysis	60
5.3 Trace plots for $\lambda_{14,1}$ and $\lambda_{4,3}$ for the three different approaches.	75
5.4 Marginal distribution PPMC plots for semiparametric latent variable model estimation of simulated data.	76
5.5 Pairwise correlation plot for semiparametric latent variable model estimation of simulated data.	77

5.6	Correlation distance and eigenvalue plots for semiparametric latent variable model estimation of simulated data.	78
5.7	Histograms of scores for MDRS E and W items.	80
5.8	Histograms of the observed scores for the Verbal Fluency A and MDRS E. The black points indicate the mean count across replicated datasets for each score. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.	82
5.9	Correlation Distance and Eigenvalue Plot PPMC.	89
5.10	Pairwise correlation plots for the single factor ($Q = 1$) and bifactor models ($Q = 4$). Each pairwise correlation plot depicts the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) for Kendall's τ values calculated using replicated data. Kendall's τ values computed from the observed data are denoted by a black "X".	90
6.1	Marginal distribution PPMC plots for semiparametric multivariate regression model estimation of simulated data.	102
6.2	Correlation distance and eigenvalue plots for semiparametric multivariate regression model estimation of simulated data.	103
6.3	Histograms of the observed scores for the Verbal Fluency A and Mattis Dementia Rating Scale E. The black points indicate the mean count across replicated datasets for each grade. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.	105
6.4	Correlation Distance and Eigenvalue Plot PPMC	106
6.5	Pairwise correlation plots for Mattis Dementia Rating Scale J and Visual Span Backwards. Each pairwise correlation plot depicts the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) for Kendall's τ values calculated using replicated data. Kendall's τ values computed from the observed data are denoted by a black "X".	107
6.6	Diagram of the IRT Model for Mixed Outcomes. The circles represent latent variables and the squares represent observed variables with solid borders corresponding to outcomes and dashed borders corresponding to covariates. The one-sided solid arrows represent regression coefficients.	109
6.7	Diagram of the Semiparametric Latent Variable Model with Bifactor Structure for Mixed Outcomes. The circles represent latent variables and the squares represent observed variables with solid borders corresponding to outcomes and dashed borders corresponding to covariates. The one-sided solid arrows represent regression coefficients.	110

6.8	Diagram of the Semiparametric Multivariate Regression Model for Mixed Outcomes. The circles represent latent variables and the squares represent observed variables with solid borders corresponding to outcomes and dashed borders corresponding to covariates. The one-sided solid arrows represent regression coefficients while the two-sided dashed lines represent residual correlations.	111
6.9	Histograms of the observed scores for the Verbal Fluency A (top row) and Mattis Dementia Rating Scale E (bottom row) by model. The black points indicate the mean count across replicated datasets for each score. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.	112
6.10	Correlation distance and eigenvalue plots for the IRT model, the single factor and bifactor semiparametric latent variable models, and the semiparametric multivariate regression model. The top row of plots present scatterplots of $d_{sld}(C^{obs}, C^{rep,m})$ versus $d_{sld}(C^{rep,m}, C^{rep,m'})$ for all replicated datasets. The grey line represents the 45 degree line. The bottom row of plots display the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) of the largest ten eigenvalues calculated using replicated data. Eigenvalues computed from the observed data are denoted by a black "X".	113
7.1	Simulated data according to different individual-level mixture model formulations. The green points represent the class centers and the green ellipse represents a 1SD contour. Left: MM. Center: BPM with diagonal covariance matrix. Right: BPM with full covariance matrix.	121
7.2	Simulated data according to different individual-level mixture model formulations with membership scores generated using scale parameter $a = 10$. Left: MM. Center: BPM with diagonal covariance matrix. Right: BPM with full covariance matrix.	122
7.3	Simulated data according to different individual-level mixture model formulations with membership scores generated using scale parameter $a = 1/10$. Left: MM. Center: BPM with diagonal covariance matrix. Right: BPM with full covariance matrix.	123
7.4	Simulated data according to different individual-level mixture model formulations. The green points represent the class centers and the green ellipse represents a 1SD contour. Left: MM. Center: BPM with diagonal covariance matrix. Right: BPM with full covariance matrix.	124
7.5	Simulated data according to different individual-level mixture model formulations with membership scores generated with scale parameter $a = 10$. Left: MM. Center: BPM with diagonal covariance matrix. Right: BPM with full covariance matrix.	125

7.6	Simulated data according to different individual-level mixture model formulations with membership scores generated with scale parameter $a = 1/10$. Left: MM. Center: BPM with diagonal covariance matrix. Right: BPM with full covariance matrix.	126
7.7	Marginal Probability Plots for Scenarios 1-5 in Table 7.6. Scenario 1 is the leftmost plot and Scenario 5 is the rightmost.	130
7.8	Trace Plots for BPM Simulated Data Example	138
7.9	Histograms of the observed scores for simulated outcomes 14 and 19 under BPM model.	142
7.10	Correlation distance and eigenvalue plots for varying K for simulated data under BPM model	143
7.11	Trace Plots for BPM NBA Example	147
7.12	Histograms of the observed scores for the Turnover Rate and Field Goal Attempts (FGA) At The Rim per 40 minutes. The black points indicate the mean count across replicated datasets for each interval. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.	148
7.13	Correlation Distance and Eigenvalue Plot PPMC	149
7.14	Histograms of the observed scores for the Verbal Fluency A and MDRS E. The black points indicate the mean count across replicated datasets for each score. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.	157
7.15	Correlation Distance and Eigenvalue Plot PPMC	161
7.16	Pairwise correlation plots for the $K = 2$ and $K = 6$ models. Each pairwise correlation plot depicts the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) for Kendall's τ values calculated using replicated data. Kendall's τ values computed from the observed data are denoted by a black "X".	162

LIST OF TABLES

Table Number	Page
3.1 Summary Statistics for SIVD Executive Functioning Outcomes	23
4.1 Data-generating values and posterior summaries for discrimination parameters, $\mathbf{a}$, from simulated data examples using the IRT model for mixed outcomes.	44
4.2 Summary Statistics for SIVD Executive Functioning Outcomes	49
4.3 Posterior summaries for regression coefficients estimated using the IRT model for mixed outcomes.	51
5.1 Summary of prior distributions and estimation approaches.	68
5.2 Posterior summaries of factor loadings from three estimation approaches using the semiparametric latent variable model.	74
5.3 Summary of effective sample sizes for $\mathbf{\Lambda}$ parameters by estimation approach.	75
5.4 Summary statistics for $I = 341$ responses to 19 SIVD executive functioning outcomes as well as outcome type assignment. ‘RC Count’ denotes a right-censored count outcome.	79
5.5 Proposed factor structure for SIVD executive functioning outcomes. * indicates a non-zero factor loading to be estimated.	84
5.6 Posterior summaries for regression coefficients for single factor, $Q = 1$, and bifactor, $Q = 4$, models.	85
6.1 Posterior summary of $\mathbf{\lambda}$	101
6.2 Posterior summary of $\mathbf{\beta}$	102
6.3 Posterior summaries for regression coefficients, $\mathbf{\beta}$	104
7.1 Means by class for outcomes 1 and 2.	120
7.2 Variances by class for outcomes 1 and 2.	122
7.3 For $K = 3$ and $(p_1 = .4, p_2 = .5, p_3 = .6)$, we calculate the Bernoulli probability of a success for $x \mathbf{g}$ under the mixed membership model (p_{mm}) and the partial membership model (p_{pm}) for varying $\mathbf{g}$	127
7.4 For $K = 3$ and $(p_1 = .02, p_2 = .15, p_3 = .90)$, we calculate the Bernoulli probability of a success for $x \mathbf{g}$ under the mixed membership model (p_{mm}) and the partial membership model (p_{pm}) for varying $\mathbf{g}$	128

7.5	Notation for cell and marginal probabilities for two binary variables, X_1 and X_2	128
7.6	λ values used to generate marginal probability plots in Figure 7.7.	130
7.7	Summary of simulated class memberships.	137
7.8	Posterior summaries for class means for classes 2 and 3.	140
7.9	Posterior summary for membership score means, $\boldsymbol{\rho}$	141
7.10	Posterior summary for membership score covariances, $\boldsymbol{\Sigma}$	141
7.11	Variables, abbreviations and formulas.	145
7.12	Posterior means for class means in the $K = 4$ model. Bolded figures indicate that the corresponding posterior 95% CI did not include 0.	150
7.13	Players with top ten membership scores for class 1 (Well-Rounded Big Men).	151
7.14	Players with top ten membership scores for class 2 (Ball Handlers).	152
7.15	Players with top ten membership scores for class 3 (Spot-up Shooters).	153
7.16	Players with top ten membership scores for class 4 (Defensive Big Men).	154
7.17	Membership scores for commonly regarded dynamic players	154
7.18	Cluster labels and typical members found in Lutz (2012).	155
7.19	Membership scores for players typifying the Lutz (2012) combo guard cluster.	155
7.20	Posterior mean of membership score correlations.	156
7.21	Posterior summary for class means for class 2 in the $K = 2$ model.	159
7.22	Posterior means for class means for classes 2-6 in the $K = 6$ model. Bolded figures indicate that the corresponding posterior 95% CI did not include 0.	160
B.1	Comparison of dissimilarity measures for matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$	193

ACKNOWLEDGMENTS

This dissertation would not have been possible without the effort and support of many people. First, I would like to thank my advisor, Elena Erosheva. From my point of view, she deftly struck the often elusive balance between providing encouragement and demanding more of my work. Her guidance and teaching has made me a more deliberate and technically sound statistician as well as a better communicator of statistical concepts in writing and speaking. She has also been a valuable resource on less statistically-oriented issues ranging from managing one's career to raising kids. I am very grateful to have worked with her.

Second, I would like to thank Paul Crane, a member of my supervisory committee and on whose grant I have been funded for almost the entirety of my graduate career. His enthusiasm for his own research has been inspiring and educational. His genuine interest in and spirited promotion of my contributions, whether directly or indirectly related to his research, has always been appreciated. Moreover, I received valuable feedback from Paul's research group, the Cognitive Outcomes with Advanced Psychometrics group, including McKay Curtis, Laura Gibbons, and Joey Mukherjee.

I also appreciate the generosity of time of the remaining members of my supervisory committee, Peter Hoff and Thomas Richardson. I have often interrupted their work seeking their perspective and they repeatedly made time to speak with me, always providing useful insight on a method or problem. Furthermore, I am thankful for the time they took to review this work. Their feedback has strengthened this dissertation.

As I mentioned during my final exam, a primary reason for choosing the University of Washington as the school at which to pursue my degree was the quality of the people, both faculty and students. Time has shown my instincts to be well-honed as I have enjoyed my time in the Statistics department at the University of Washington immensely. I hesitate to name specific people as I have received the benefit of many faculty members' and students'

intellectual curiosity and friendship.

These acknowledgments be would incomplete without recognizing the sources of funding. This work was supported by Grant R01 AG 029672 from the National Institute on Aging.

I must also acknowledge a few key influences prior to my arrival at the University of Washington. Professor Robert Kohn at New York University was instrumental in directing me towards statistics and was a trusted mentor for a student considering the pursuit of a doctorate.

As with any meaningful endeavor in my life, without the encouragement and confidence of my parents and my brother and his family, I may never have pursued such an opportunity. Their faith in me is often overly optimistic but I would not know how to proceed without it.

Finally, I would like to dedicate this work to my wife, Lindsay, and my son, Jasper. They have shown unwavering patience, support and tolerance, more so than I could have reasonably expected of anyone. Although it is unlikely that I will ever be able to adequately thank them, I look forward to trying.

DEDICATION

To Lindsay and Jasper.

Chapter 1

INTRODUCTION

Medicine, psychology, economics, education, marketing, sociology and political science are examples of disciplines that employ latent variable modeling to analyze multivariate data. However, the types of data and research questions in many studies do not conform neatly with common latent variable analyses such as those using item response theory (IRT) models from educational testing. Instead, the data are often of mixed type in the sense that some of the observed responses may be binary, some may be ordered categorical, some may be counts and others yet may be continuous. In this dissertation, we consider parametric and semiparametric approaches to accommodating mixed outcomes in latent variable and multivariate regression models. These models provide great flexibility in the types of data that may be analyzed, represent multivariate data with a lower dimensional structure, and enable statistical inference on the association between multivariate outcomes and covariates of interest.

The dissertation work is motivated by medical studies of cognitive decline in the elderly. Severe cognitive deficits such as those associated with Alzheimer’s disease are a substantial health problem for the elderly and constitute both a significant societal and financial burden. I will primarily use data from an ongoing study, the Subcortical Ischemic Vascular Disease Program Progress Grant (SIVD) (Chui, Zarow, Mack, Ellis, Zheng, Jagust, Mungas, Reed, Kramer, DeCarli, et al., 2006), to illustrate the proposed methods. More specifically, we are interested in analyzing multivariate mixed data that attempt to measure cognitive functioning in SIVD study participants. We would like to relate this multivariate set of outcomes to covariates of interest, primarily MRI-measured brain volumes. Hierarchical latent variable and multivariate regression models provide means of modeling the data in a way that allows us to make inference on the associations between multivariate outcomes and predictors.

Within psychometrics and the social sciences, there have been many separate branches of multivariate methods development, including latent variable models. Psychometricians and social scientists have applied latent variable models since at least the early 1900s when Spearman is generally credited with the origin of factor analysis and its application to measure general intelligence. Bartholomew, Knott, and Moustaki (2011) note that the ideas underlying factor analysis can be traced back as far as Galton (1888). Thurstone (1947) and Holzinger and Swineford (1937) expanded upon Spearman's work to include multiple factors. In education, item response theory models were developed to model the probability of correct responses to individual test questions or items with an underlying latent variable assumed to represent a person's ability.

Whereas the original factor analysis model assumed continuous data, item response theory models were developed for binary data with later extensions for ordered categorical data. Lord, Novick, and Birnbaum (1968) are often credited with producing the seminal work on item response theory. Bock (1997) pointed out that the origins of these models can be traced back to Thurstone who was expanding upon the work of Spearman in factor analysis. Takane and de Leeuw (1987) established the close relationship between factor analysis and item response theory. Although primarily used for the development of large-scale tests administered by organizations such as the Educational Testing Service, item response theory has recently also been applied in medicine (Ehlenbach, Hough, Crane, Haneuse, Carson, Curtis, and Larson, 2010), sociology (Osgood, McMorris, and Potenza, 2002), marketing (De Jong, Steenkamp, Fox, and Baumgartner, 2008) and economics (Pitt, Khandker, and Cartwright, 2006). Item response theory models will be our point of departure for analyzing the cognitive functioning outcomes in the SIVD study. Van der Linden and Hambleton (1997) provides a comprehensive reference for item response theory models and methods.

Separate from factor analysis and item response theory, Lazarsfeld developed latent class models for sociological research and Lazarsfeld and Neil (1968) compiles much of this early work. Applied to categorical data, latent class models replace the continuous latent variables present in factor analysis and item response theory models with categorical latent variables. In social science research, latent class models are often used to describe

heterogeneous populations with unidentified groups. Heinen (1996) and Bartholomew et al. (2011) discussed the relationship between latent class models and models with continuous latent variable models. Grade of membership models (Woodbury, Clive, and Garson Jr, 1978) relax the assumptions of latent class analysis to allow for partial membership in multiple groups or subpopulations. Heller, Williamson, and Ghahramani (2008) provide a different type of partial membership model that relaxes the assumptions of a different but equivalent formulation of the latent class model.

Despite the disparate origins of these developments in multivariate analysis and latent variable modeling, statistics has taken an increasing role in the theoretical and methodological development of these models, sometimes directly and other times unknowingly in parallel. Bartholomew et al. (2011) provide a unifying treatment of latent variable models from a statistical modeling perspective. Furthermore, as noted by Bartholomew et al. (2011), “latent variables have a long history in statistics, but their ubiquity has been obscured by variations in terminology.” Bartholomew et al. (2011) cite random effects, missing data, and potential outcomes as a few examples of the use of latent variables in statistics by another name. Skrondal and Rabe-Hesketh (2004) also discuss a wide variety of statistical models that may also be considered latent variable models.

While the interest of psychometricians and social scientists may focus on the models’ ability to estimate latent constructs or identify hidden groups, statisticians have focused on the models as methods for multivariate analysis and dimensionality reduction. Latent class models, for example, can be thought of as a subset of finite mixture models whose statistical origins can be traced back to Pearson (1894). Statisticians and researchers in other fields have long used mixture models to model multimodal and skewed distributions and to perform model-based cluster analysis.

Early statistical involvement in latent variable modeling focused on the development of theory and estimation for inference. Lawley (1940) and Anderson and Rubin (1956) were among the first to provide more statistical treatments of factor analysis. Anderson (1954) was early to address statistical inference issues in latent class analysis. Jöreskog (1969), Jöreskog (1970) and Jöreskog and Goldberger (1975) developed methods of estimation for confirmatory factor analysis and structural equation models. More recently, statistics re-

search on latent variable modeling has included developments relating to models for mixed data (Sammel, Ryan, and Legler, 1997; Moustaki and Knott, 2000; Dunson, 2003), methods for the analysis of high-dimensional data (Carvalho, Chang, Lucas, Nevins, Wang, and West, 2008), and hierarchical models (Fox and Glas, 2001; Dunson, 2003). Recent developments in estimation have focused on semiparametric methods (Fahrmeir and Raach, 2007) and Bayesian estimation (Shi and Lee, 1998; Fox, 2010). Moreover, latent variable models are one of the many areas that lie at the intersection of statistical and computer science research (Erosheva and Fienberg, 2005; Blei, Ng, and Jordan, 2003; Knowles and Ghahramani, 2011).

These recent areas of research inform and motivate many of the proposed methods in this dissertation. We develop new parametric and semiparametric approaches to handle mixed outcomes in latent variable models and extend these models hierarchically to relate latent variables to covariates. We also consider a multivariate regression model as an alternative to the hierarchical latent variable models.

We use Bayesian methods for model formulation and estimation. Bayesian methods allow for ready extension of models hierarchically and easily computable estimates of uncertainty for parameters and latent variables using posterior distributions. They also facilitate one-stage estimation of hierarchical models and thus avoid the biased estimates and failure to propagate uncertainty that plague two-stage estimation procedures (Lu, Thomas, and Zumbo, 2005; Moustaki, Jöreskog, and Mavridis, 2004).

This dissertation is organized as follows. In Chapter 2, we provide a brief review of the Hybrid (Hamiltonian) Monte Carlo method and introduce the posterior predictive model checking methods that will be used extensively throughout this dissertation. In Chapter 3, we present the motivating data example, the Subcortical Ischemic Vascular Dementia (SIVD) study. We briefly describe the study and the construction of the analytic sample and provide selected results from exploratory data analysis.

In Chapter 4, we focus on extending and applying existing parametric approaches for mixed data in latent variable models in the context of item response theory models. We extend the mixed data methods for latent variable models to handle new types of outcomes specifically encountered in cognitive testing situations like the SIVD study. We employ a hierarchical model formulation to incorporate covariates, discuss Bayesian estimation of

the model and demonstrate these methods on simulated data before applying them to our motivating data example from the SIVD study.

In Chapter 5, we adopt a semiparametric approach for accommodating mixed outcomes in latent variable models. Hoff (2007) proposed the semiparametric approach to estimate a Gaussian copula model for modeling the dependence among mixed outcomes. We no longer restrict ourselves to the item response theory framework and apply the method of Hoff (2007) in a more general latent variable modeling framework. As in Chapter 4, we present a hierarchical version of the model to include covariates. We explore the use of a parameter expansion approach to estimation to aid in more efficient sampling and to overcome high autocorrelations among Markov chain Monte Carlo draws. We demonstrate the model on simulated data and compare different estimation approaches. In applying the model to the SIVD data, we use the posterior predictive model checking methods to determine whether a proposed latent structure is sufficient to adequately model the dependence structure among the outcomes.

In Chapter 6, we introduce a reduced rank regression model that is able to employ the semiparametric approach of Hoff (2007) but forego the latent variable model in favor of a multivariate regression approach. This approach allows for the direct estimation of the correlations among the outcomes while retaining the same mean structure as the semiparametric latent variable for the mixed outcomes. While such an approach sacrifices some of the dimension reduction benefits of the latent variable model in modeling the dependence structure among the outcomes, there is no longer a need to identify an appropriate latent structure in order to approximate the dependence structure. We discuss the different identifiability constraints necessary for employing a reduced rank formulation and sampling correlation matrices during estimation. We demonstrate the reduced rank regression model proposed on simulated data and on the SIVD study data.

Given the utility of the semiparametric approach for latent variable and multivariate regression models for mixed outcomes, we consider in Chapter 7 its application with another method for the analysis of multivariate data, model-based clustering. The Bayesian partial membership model originally proposed by Heller et al. (2008) provides a means for soft clustering of a variety of types of outcomes. As a member of the family of individual-level

mixture models, the model is related to the mixed membership model (Erosheva and Fienberg, 2005) and we compare and contrast the two models. Deriving the model formulation for data generated by any exponential family distribution, Heller et al. (2008) demonstrated the model on binary data. We show how the model may be applied to normally distributed data. We then extend the model to allow for more flexible correlations among class memberships and once again leverage the semiparametric approach of Hoff (2007) to allow for mixed outcomes. We demonstrate the model on executive functioning data from the SIVD study and on NBA player data from the 2010-11 season. In the latter example, we compare and contrast our results with those of a cluster analysis on the same data (Lutz, 2012).

Finally, we conclude with a summary of contributions and results as well as a discussion of future directions in Chapter 8.

Chapter 2

BAYESIAN METHODS BACKGROUND

2.1 Introduction

In this dissertation, we rely exclusively on the Bayesian approach to statistical inference. Here we briefly review and, in the case of some of the posterior predictive checks, introduce a few of the methods that will be extensively used throughout this dissertation. We do assume general familiarity with Bayesian methods and do not intend to provide an overview. For a review of Bayesian methods, see Hoff (2009) and Gelman, Carlin, Stern, and Rubin (2004). This chapter is organized as follows. In Section 2.2, we review Hybrid Monte Carlo (HMC), an alternative to the more commonly applied Gibbs and Metropolis Hastings (MH) sampling algorithms for approximating posterior distributions. Section 2.3 briefly discusses the validating procedures for the software programs implementing the proposed models. Finally, in Section 2.4, we introduce the posterior predictive model checking methods for evaluating fit of multivariate models with mixed outcomes.

2.2 Hybrid Monte Carlo

One of the disadvantages of the most common MCMC algorithms, the Gibbs and MH samplers, is their random walk nature. One means of overcoming the random walk behavior is to use gradient information to enable proposals far away from the current state that still have a high probability of being accepted. Hybrid (Hamiltonian) Monte Carlo (HMC), applicable to continuous state spaces only, achieves this improvement by borrowing from molecular dynamics and incorporating gradient information through the addition of a momentum variable p . We provide a brief description of the algorithm here. Duane, Kennedy, Pendleton, and Roweth (1987) were the first to integrate the molecular dynamics approach with MCMC. Neal (2010) provides a thorough review, MacKay (2003) provides a more succinct description, and Ishwaran (1999) and Chen, Qin, and Liu (2001) are examples of its

application to common statistics problems.

Let $\boldsymbol{\theta}$ be a D -length vector of parameters, let $\mathbf{Y}$ be the observed data and let $p(\boldsymbol{\theta}, \mathbf{Y}|\boldsymbol{\psi})$ be the joint density of the data and parameters. If we specify an auxiliary “momentum” variable, p_j , for each parameter $\theta_j, j = 1, \dots, D$, then we define the Hamiltonian as

$$H(\boldsymbol{\theta}, \mathbf{p}) = U(\boldsymbol{\theta}) + K(\mathbf{p}), \quad (2.1)$$

where the kinetic energy, $U(\boldsymbol{\theta})$, and the potential energy, $K(\mathbf{p})$ are,

$$U(\boldsymbol{\theta}) = -\log p(\boldsymbol{\theta}, \mathbf{Y}|\boldsymbol{\psi}), \quad (2.2)$$

$$K(\mathbf{p}) = \frac{1}{2} \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}. \quad (2.3)$$

$\mathbf{M}$ is the “mass” matrix and is symmetric and positive definite. Thus $K(\mathbf{p})$ is proportional to a negative log Gaussian distribution with zero mean and covariance matrix $\mathbf{M}$. $\mathbf{M}$ is commonly assumed to be diagonal.

In Hamiltonian dynamics, the variables $\boldsymbol{\theta}$ and $\mathbf{p}$ evolve according to the Hamiltonian equations for $j = 1, \dots, D$:

$$\begin{aligned} \frac{d\theta_j}{dt} &= (\mathbf{M}^{-1} \mathbf{p})_j, \\ \frac{dp_j}{dt} &= -\frac{dU}{d\theta_j}. \end{aligned} \quad (2.4)$$

These “forces” push the parameters to regions of higher probability. The HMC algorithm then proceeds in two main steps. First, the momentum variables are drawn according to a $N(\mathbf{0}, \mathbf{M})$ distribution. Second, we update the parameters $\boldsymbol{\theta}$ and the auxiliary variables $\mathbf{p}$ using the Hamiltonian equations (2.4). The dynamics described by the Hamiltonian equations (2.4) must be approximated in discrete steps. For HMC, the leapfrog method is typically used for this approximation. The Hamiltonian dynamics are simulated by running leapfrog for L steps with each individual step as follows

$$p_j(t + \epsilon/2) = p_j(t) - \frac{\epsilon}{2} \frac{dU}{d\theta_j}(\theta_j(t)), \quad (2.5)$$

$$\theta_j(t + \epsilon) = \theta_j(t) + \epsilon \frac{p_j(t + \epsilon/2)}{m_j}, \quad (2.6)$$

$$p_j(t + \epsilon) = p_j(t + \epsilon/2) - \frac{\epsilon}{2} \frac{dU}{d\theta_j}(\theta_j(t + \epsilon)), \quad (2.7)$$

under the assumption $\mathbf{M}$ is diagonal and m_j is the j -element on the diagonal of $\mathbf{M}$. At the end of L steps, we have a new proposal $(\boldsymbol{\theta}^*, \mathbf{p}^*)$ for the parameters and auxiliary variables.

Here, ϵ is a step-size parameter that is set by the user. Moreover, the number of leapfrog steps L is also set by the user. Note that HMC performance can be sensitive to these parameters so they must be tuned.

To adjust for bias introduced by the discretization, the proposal $(\boldsymbol{\theta}^*, \mathbf{p}^*)$ put forth by the simulated Hamiltonian dynamics is accepted with probability

$$\min [1, \exp (-H(\boldsymbol{\theta}^*, \mathbf{p}^*) + H(\boldsymbol{\theta}, \mathbf{p}))] \quad (2.8)$$

$$= \min [1, \exp (-U(\boldsymbol{\theta}^*) + U(\boldsymbol{\theta}) - K(\mathbf{p}^*) + K(\mathbf{p}))]. \quad (2.9)$$

If the proposal is rejected, then the new draw is set to the same values as the previous draw. The momentum variables are discarded. This full algorithm is presented in Algorithm 1. Note that step 5(d) in iteration l can be combined with step 5(a) in $l + 1$ for $l \neq L$ as they are the same.

As Neal (2010) notes, it may be preferable in some cases to combine HMC with other types of sampling such as Gibbs sampling. Combining HMC with other types of sampling may be desirable when some parameters are discrete or it is not possible (or computationally difficult) to calculate the derivatives of the log probability. In these cases, HMC is used to sample a subset of the parameters while the remaining parameters are drawn using alternative sampling methods. In Chapters 6 and 7, we combine HMC with Gibbs sampling steps.

Again, the performance of the HMC algorithm is heavily influenced by the values of L , ϵ , and $\mathbf{M}$. The values of L and ϵ are commonly tuned so that proposals are accepted at a targeted rate. Neal (2010) concludes that an optimal acceptance rate is around 65%. Some options for tuning the HMC parameters including varying the number of steps L , using multiple stepsizes, and using adaptive values for $\mathbf{M}$ that depend on $\boldsymbol{\theta}$ as proposed by Girolami and Calderhead (2011). Neal (2010) discusses a number of alternatives for tuning and discusses variations on HMC that may further improve performance.

Algorithm 1 Hybrid Monte Carlo algorithm

For each iteration:

1. Sample $\mathbf{p} \sim N(\mathbf{0}, \mathbf{M})$.
 2. Calculate current value for $H(\boldsymbol{\theta}, \mathbf{p})$.
 3. Set $\mathbf{p}^* = \mathbf{p}$, $\boldsymbol{\theta}^* = \boldsymbol{\theta}$.
 4. Calculate gradient for current value of $\boldsymbol{\theta}^*$, $\frac{dU}{d\boldsymbol{\theta}}(\boldsymbol{\theta}^*)$.
 5. For l in $1 : L$
 - (a) Update momentum, $\mathbf{p}^* = \mathbf{p}^* - \frac{\epsilon}{2} \frac{dU}{d\boldsymbol{\theta}}(\boldsymbol{\theta}^*)$.
 - (b) Update parameters, $\boldsymbol{\theta}^* = \boldsymbol{\theta}^* + \epsilon \mathbf{M}^{-1} \mathbf{p}^*$.
 - (c) Calculate gradient for current value of $\boldsymbol{\theta}^*$, $\frac{dU}{d\boldsymbol{\theta}}(\boldsymbol{\theta}^*)$.
 - (d) Update momentum, $\mathbf{p}^* = \mathbf{p}^* - \frac{\epsilon}{2} \frac{dU}{d\boldsymbol{\theta}}(\boldsymbol{\theta}^*)$.
 6. For proposed values $(\boldsymbol{\theta}^*, \mathbf{p}^*)$, calculate $H(\boldsymbol{\theta}^*, \mathbf{p}^*)$.
 7. Draw $u \sim \text{Uniform}(0, 1)$.
 8. If $u < \min[1, \exp(-H(\boldsymbol{\theta}^*, \mathbf{p}^*) + H(\boldsymbol{\theta}, \mathbf{p}))]$, set $\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^*$. Else set $\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}$.
-

2.3 Software Validation

We rely on Markov chain Monte Carlo (MCMC) methods for Bayesian estimation of our models. We implement these methods through custom-developed code. As a result, there is naturally a need to test the custom-developed code implementing the models. We follow the common strategy of simulating datasets from the model we wish to fit. As we know the data-generating parameter values, we may compare these values to model estimates after having fit the model to a simulated dataset. More specifically, we compare posterior means and medians to the corresponding true parameters to see how well they match up. In addition, we check the frequency with which posterior intervals contain the data-generating parameters with the expectation that, for example, 95% posterior credible intervals will contain the data-generating values approximately 95% of the time.

Cook, Gelman, and Rubin (2006) propose a more systematic approach to testing code implementing Bayesian methods but it is potentially computationally intensive. Their approach calculates posterior quantiles and compares them to a Uniform(0,1) distribution. If the posterior quantiles deviate from a Uniform(0,1) distribution, then the software may have been incorrectly implemented. Interestingly enough, Monahan and Boos (1992) use a closely related procedure to determine whether a likelihood is proper in the sense that the posterior distribution based on the likelihood is valid by coverage. Monahan and Boos (1992) developed this method in order to evaluate the quality of approximations by alternate likelihoods such as those based on ranks (Pettitt, 1982).

Although results by Hartigan (1966) established the frequentist coverage validity to $O(n^{-1})$ of two-sided equal-tailed posterior credible intervals for a single parameter under regularity conditions for any prior and Nicolaou (1993) and Datta (1996) among others have provided results for intervals in the presence of nuisance or additional parameters, these results were developed for the case where the likelihood function is the density function of the data. In the semiparametric methods we apply in Chapters 5-7, the likelihood is not the density function but rather an alternative function based on the orderings of the data. Therefore, the results of Hartigan (1966) and others may not necessarily hold for the semiparametric methods we apply in Chapters 5 - 7. As a result, in evaluating the

correctness of our code, we may not be able to use posterior intervals in a strict sense to evaluate the validity of our program.

2.4 Posterior Predictive Model Checking

Posterior predictive model checking is a diagnostic method for Bayesian models in which the observed data are compared to the posterior predictive distribution under the proposed model. If the model is able to approximate the data-generating process reasonably well, then the observed data should appear consistent with the posterior predictive distribution. Posterior predictive model checks can be implemented by drawing simulated values from the posterior distribution of replicated data. One can then compare these simulated values to the observed data directly or use test quantities that describe important data features. Differences between the observed and replicated data indicate potential model misspecification.

Let $\mathbf{y}^{obs}$ denote the observed data and let $\boldsymbol{\psi}$ denote the set of parameters in the model. Let $\mathbf{y}^{rep}$ denote replicated data values drawn from the posterior predictive distribution,

$$p(\mathbf{y}^{rep}|\mathbf{y}^{obs}) = \int p(\mathbf{y}^{rep}|\mathbf{y}^{obs}, \boldsymbol{\psi})p(\boldsymbol{\psi}|\mathbf{y}^{obs}) d\boldsymbol{\psi}. \quad (2.10)$$

The test quantity or discrepancy measure $T(\mathbf{y}, \boldsymbol{\psi})$ is typically a scalar summary of data (observed or replicated) and parameters. When restricted to a summary of data only, $T(\mathbf{y})$ is often referred to as a statistic. The summary $T(\cdot)$ is chosen to highlight features of the data that we would like to approximate well in our model.

The test quantity can be employed to provide graphical or numerical evidence regarding model misfit. Graphically, one might compare $T(\mathbf{y}^{obs}, \boldsymbol{\psi})$ and $T(\mathbf{y}^{rep}, \boldsymbol{\psi})$ using scatterplots or histograms as described in Gelman et al. (2004). Lack of fit can be detected numerically using posterior predictive p-values. This p-value is defined as the probability that the test quantity for the replicated data is more extreme than the corresponding quantity for the observed data,

$$p = \Pr(T(\mathbf{y}^{rep}, \boldsymbol{\psi}) \geq T(\mathbf{y}^{obs}, \boldsymbol{\psi})|\mathbf{y}^{obs}). \quad (2.11)$$

When we sample from the posterior using Monte Carlo methods, posterior predictive model checking methods are straightforward to implement. We draw M sets of parameter values

$\psi^1, \dots, \psi^M$ from the posterior $p(\psi|\mathbf{y}^{obs})$; that is, we may just use the values drawn in estimating the model. We then draw replicated data $\mathbf{y}^{rep,m}$ from $p(\mathbf{y}^{rep,m}|\psi^m)$ for each $m = 1, \dots, M$. Finally, we can compute $T(\mathbf{y}^{rep,m}, \psi^m)$ for each m , and then estimate equation (2.11) by comparing $T(\mathbf{y}^{rep,m}, \psi^m)$ to $T(\mathbf{y}^{obs}, \psi^m)$.

Robins, van der Vaart, and Ventura (2000) and Bayarri and Berger (2000) point out that the posterior predictive p-value is not uniformly distributed on the interval $[0, 1]$ under the null. Although some have proposed adjustments to these posterior predictive p-values so that their behavior mimics more closely that of traditional p-values, many treat these p-values as a diagnostic tool for model fit rather than a strict hypothesis test. Moreover, Gelman (2007) and Crespi and Boscardin (2009) argue that graphical displays have more utility. We will take this point of view in evaluating the fit of our proposed models. See Gelman et al. (2004) for more discussion of posterior predictive model checking methods.

Application of posterior predictive model checking methods to latent variable, specifically item response theory (IRT), models includes work by Glas and Meijer (2003), Sinharay (2005), Sinharay and Johnson (2003), Sinharay, Johnson, and Stern (2006), Levy, Mislevy, and Sinharay (2009), and Curtis (2010). The proposed methods focus on different aspects of IRT models, examining the assumption of local independence, person fit and item fit among others. Suggested discrepancy measures include observed score distribution, item pair odds ratio, and biserial correlation coefficients. Different graphical methods are highlighted as well. Sinharay and Johnson (2003) provide an extensive set of simulation studies comparing different discrepancy measures. Fox (2010) provides an overview of posterior predictive model checking applications to IRT models. Curtis (2010) uses rank correlations to compare the observed data to the replicated data on an item by item basis and we adopted this approach as one means of checking our models.

Posterior predictive model checking methods have also been applied to models for mixed outcomes, such as by Miglioretti (2003) and by Daniels and Normand (2006). Common methods include graphical displays comparing the observed distributions and the predicted distributions for the different outcomes as well as comparing the correlations between outcomes. Following along these lines, we compare the observed marginal distribution of each item or outcome to those of the replicated data in order to assess how well we are modeling

this aspect of the data. We subsequently consider methods for evaluating the fit of the model to measures of the dependence structure in the data.

2.4.1 Marginal Distributions

Figure 2.1 presents sample model checks for the marginal distributions of two hypothetical outcomes. The histograms show the observed data for these items. For each of the replicated datasets, we tabulated the counts of scores in each histogram bin. The black point indicates the posterior predictive mean of the counts and the vertical segment spans the 2.5% to 97.5% posterior predictive quantiles of these counts. Figure 2.1(a) displays a case where the model appears to fit the data reasonably well with each of the black vertical segments covering the observed values represented by the top of each histogram bar. Figure 2.1(b) exhibits a case where the model is less successful with many of the black vertical segments failing to cover the observed values.

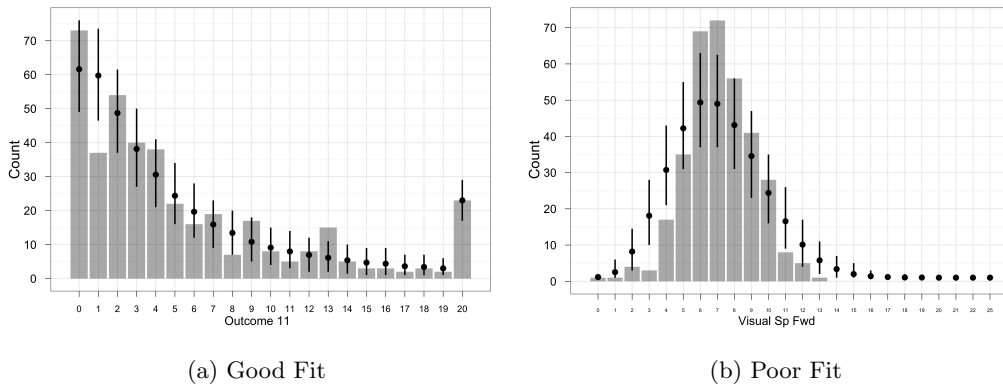


Figure 2.1: Histograms of the observed item scores in the simulated dataset. The black points indicate the mean count across replicated datasets for each response. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.

2.4.2 Item-by-Item Correlations

To examine how well our model captures the dependence structure among outcomes, we compute a rank correlation using Kendall’s τ matrix for the observed data and for each of the replicated datasets. Rank correlations such as Kendall’s τ are useful for mixed outcomes where the data are ordinal but the true association between variables may not be linear. Curtis (2010) used Kendall’s τ to assess the fit of an item response theory model on an item-by-item basis. We compare the observed values of Kendall’s τ to the replicated values. Figure 2.2 presents two sample plots of pairwise rank correlations¹ between an outcome and the other remaining outcomes in their respective multivariate sets. Each plot displays the mean predicted value and 95% posterior predictive intervals for the rank correlations from the replicated data in grey and the rank correlations from the observed data with the black “X”. In Figure 2.2(a), the model fits the observed pairwise rank correlations well with the blue segments covering the observed values. Figure 2.2(b) presents an example in which the model is less successful.

2.4.3 Correlation Matrices

While this item-by-item comparison of correlations between observed and replicated data can be useful, we also need a way to carry out a global comparison of the correlation matrices. To compare the matrices, we employ a dissimilarity measure between two matrices, similar in spirit to the approach of Crespi and Boscardin (2009). For multivariate outcome data, Crespi and Boscardin (2009) describe a discrepancy measure that relies on distance metrics such as Euclidean or Manhattan distances to compare the vectors of outcomes in the replicated data and the observed data. Let $\mathbf{y}_1^{obs}, \dots, \mathbf{y}_I^{obs}$ represent the observed responses for I participants. Further assume that we have drawn M sets of parameters from their posterior distribution. For each set of parameters, $\boldsymbol{\psi}^m, m = 1, \dots, M$, we simulate data according to the parameters $\boldsymbol{\psi}^m$ to generate $\mathbf{y}_1^{rep,m}, \dots, \mathbf{y}_I^{rep,m}$. Crespi and Boscardin (2009) then compute pairwise distances for all replicated data $d(\mathbf{y}_i^{rep,m}, \mathbf{y}_i^{rep,m'})$ for each participant i across draws $m, m' = 1, \dots, M$ as well as the distance between the replicated

¹For the remainder of the document, we use rank correlation to mean Kendall’s tau.

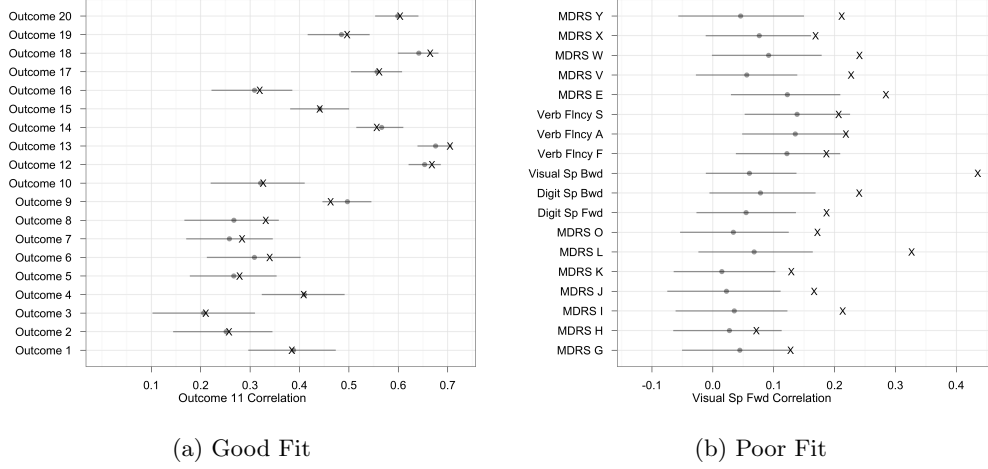


Figure 2.2: Sample Pairwise correlation plots for a single outcome against the remaining outcomes depicting the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) for Kendall's τ values calculated using replicated data. Kendall's τ values computed from the observed data are denoted by a black "X".

data and the observed data $d(\mathbf{y}_i^{rep,m}, \mathbf{y}_i^{obs})$. By their logic, if the model for the multivariate outcomes is consistent with the data, then the set of distances of $\{d(\mathbf{y}_i^{rep,m}, \mathbf{y}_i^{obs})\}$ should be largely consistent with $\{d(\mathbf{y}_i^{rep,m}, \mathbf{y}_i^{rep,m'})\}$. If the model fits poorly, then $\{d(\mathbf{y}_i^{rep,m}, \mathbf{y}_i^{obs})\}$ should be stochastically greater than $\{d(\mathbf{y}_i^{rep,m}, \mathbf{y}_i^{rep,m'})\}$. Crespi and Boscardin (2009) present histograms of the sets of distances as a graphical means of comparison. To numerically evaluate the consistency between these sets for each participant, Crespi and Boscardin (2009) suggest the use of a Mann-Whitney-Wilcoxon procedure.

Rather than compare the distance between two vectors, we are interested in comparing the distance between rank correlation matrices of observed and replicated data, $\mathbf{C}^{obs}$ and $\mathbf{C}^{rep}$. As in Crespi and Boscardin (2009), we will have two sets of discrepancies, $\{d(\mathbf{C}^{obs}, \mathbf{C}^{rep,m})\}$ and $\{d(\mathbf{C}^{rep,m}, \mathbf{C}^{rep,m'})\}$ where $m \neq m'$. We compare these two sets by plotting them in the following manner. For each replicated draw m , we have one value $d(\mathbf{C}^{obs}, \mathbf{C}^{rep,m})$ and an $M - 1$ -length vector of values $d(\mathbf{C}^m, \mathbf{C}^{m'})$. We plot $d(\mathbf{C}^{obs}, \mathbf{C}^{rep,m})$ on the vertical axis and $d(\mathbf{C}^{rep,m}, \mathbf{C}^{rep,m'})$ on the horizontal axis, leaving us with a hori-

zontal set of points at $d(\mathbf{C}^{obs}, \mathbf{C}^{rep,m})$ for each m . If the model is consistent with the data, then we expect $d(\mathbf{C}^{obs}, \mathbf{C}^{rep,m})$ to be consistent with $d(\mathbf{C}^{rep,m}, \mathbf{C}^{rep,m'})$. That is, we expect the points in the plot to concentrate in the neighborhood of a 45 degree line on the plot. We are particularly concerned about the misfit of our model when all points lie to the left of the 45 degree line, indicating that the distances among replicated rank correlation matrices are smaller than those between the observed and replicated correlation matrices.

This global method of comparing rank correlation matrices requires the specification of a dissimilarity measure/distance metric. An obvious choice is the generalization of Euclidean distance to matrices. The Euclidean (or Frobenius) distance, d_E , between two correlation matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ is

$$d_E(\mathbf{C}_1, \mathbf{C}_2) = \sqrt{\sum_{i < j} (\mathbf{C}_{1(i,j)} - \mathbf{C}_{2(i,j)})^2} \quad (2.12)$$

where $\mathbf{C}_{1(i,j)}$ and $\mathbf{C}_{2(i,j)}$ denote the i, j th element of the matrices $\mathbf{C}_1$ and $\mathbf{C}_2$.

As Davis (2008) observed, this is a localized means of comparing matrices and correlation or covariance matrices often possess more structure that methods such as principal components analysis try to exploit. Davis (2008) relies on the log determinant (LogDet) divergence to compare correlation matrices. The LogDet divergence, d_{ld} , is defined as

$$d_{ld}(\mathbf{C}_1, \mathbf{C}_2) = \text{tr}(\mathbf{C}_1 \mathbf{C}_2^{-1}) - \log |\mathbf{C}_1 \mathbf{C}_2^{-1}| - p \quad (2.13)$$

where p is the dimension of $\mathbf{C}_1$ and $\mathbf{C}_2$, $|X|$ denotes the determinant of X and $\text{tr}(X)$ the trace of X . Unlike the Euclidean distance, the LogDet divergence is not a distance metric and is not symmetric meaning that $d_{ld}(\mathbf{C}_1, \mathbf{C}_2) \neq d_{ld}(\mathbf{C}_2, \mathbf{C}_1)$. The LogDet divergence is also known as the Wishart dissimilarity or distance measure. The LogDet divergence can be formulated as a component of the Kullback-Leibler divergence between two multivariate Gaussians or as a constrained maximum likelihood using the Wishart distribution (Davis, 2008). One means of creating a symmetric LogDet divergence, d_{sld} , is setting

$$d_{sld}(\mathbf{C}_1, \mathbf{C}_2) = d_{ld}(\mathbf{C}_1, \mathbf{C}_2) + d_{ld}(\mathbf{C}_2, \mathbf{C}_1). \quad (2.14)$$

One additional possibility for a distance measure between correlation matrices is the correlation matrix distance (CMD) employed by Herdin, Czink, Ozcelik, and Bonek (2005)

to evaluate the stationarity of multiple input multiple output channels in mobile radio. They define the distance between two correlation matrices, $\mathbf{C}_1$ and $\mathbf{C}_2$, as

$$d_{CMD}(\mathbf{C}_1, \mathbf{C}_2) = 1 - \frac{\text{tr}(\mathbf{C}_1 \mathbf{C}_2)}{\|\mathbf{C}_1\|_f \|\mathbf{C}_2\|_f} \in [0, 1] \quad (2.15)$$

where $\|\cdot\|_f$ is the Frobenius norm. They note this metric can be reformulated as 1 minus the inner product of the vectorized correlation matrices

$$d_{CMD}(\mathbf{C}_1, \mathbf{C}_2) = 1 - \frac{\langle \text{vec}(\mathbf{C}_1), \text{vec}(\mathbf{C}_2) \rangle}{\|\text{vec}(\mathbf{C}_1)\|_2 \|\text{vec}(\mathbf{C}_2)\|_2} \in [0, 1]. \quad (2.16)$$

The authors note that “it becomes zero if the correlation matrices are equal up to a scaling factor and one if they differ to a maximum extent” (Herdin et al., 2005, pg. 137).

After an informal and heuristic comparison of the measures detailed in Appendix B, we ultimately settled on the symmetric LogDet discrepancy measure as it appeared to be sensitive to differences in correlation matrices that the other measures were not. Figure 2.3 provides examples of model fit assessment using the global comparison of the rank correlation matrices based on the symmetric LogDet discrepancy measure. In the figure, we display the scatterplots of $d_{sld}(\mathbf{C}^{obs}, \mathbf{C}^{rep,m})$ versus $d_{sld}(\mathbf{C}^{rep,m}, \mathbf{C}^{rep,m'})$. Figure 2.3(a) presents a case where the model does a good job of replicating the observed rank correlation matrix; we see the points are largely symmetrically around the 45 degree lines in black.

The plot on the right in Figure 2.3(b) shows the result for a case of poor model fit. In this case, we see that $d_{sld}(\mathbf{C}^{rep,m}, \mathbf{C}^{rep,m'})$ is always smaller than $d_{sld}(\mathbf{C}^{obs}, \mathbf{C}^{rep,m})$, forcing the cloud of points to the upper left of the plot. It is this type of pattern that we expect to see when there is non-trivial inconsistency between the dependence structure of the model and the data.

Another means of evaluating the model’s ability to represent the dependence structure observed in the data is to compare the eigenvalues of the observed rank correlation matrix to those of replicated data correlation matrices. The eigenvalues of correlation matrices form the basis of heuristic tests in factor analysis such as the latent root criterion (Guttman, 1954) or the scree test (Cattell, 1966). These heuristic tests are typically employed to determine the number of factors to include in the model. Figure 2.4 presents sample plots of the ten largest eigenvalues of the simulated data marked with a black “X”. The grey

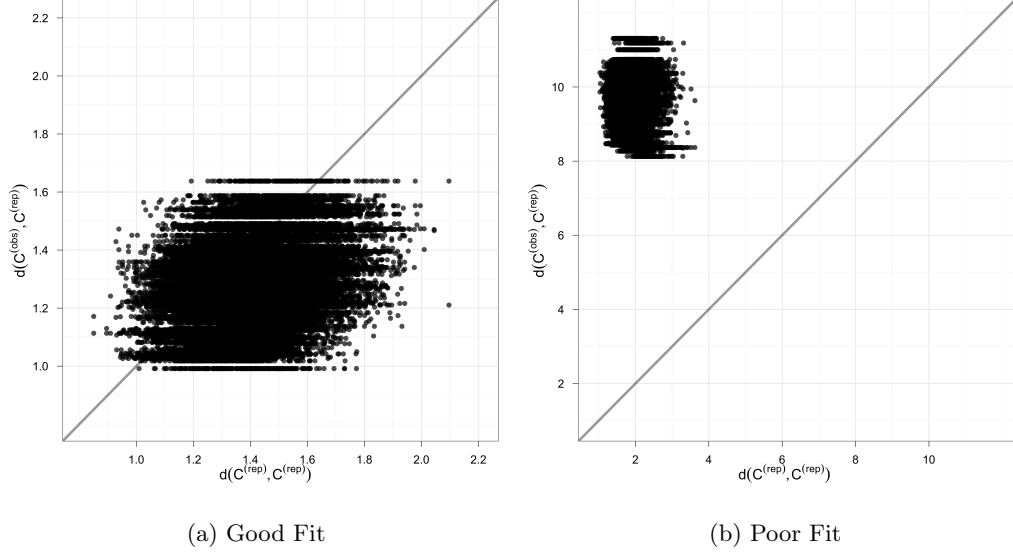
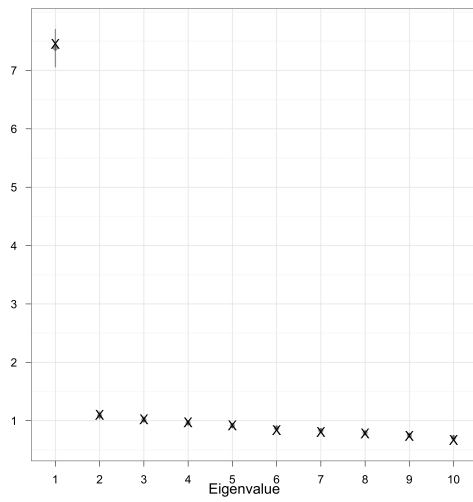
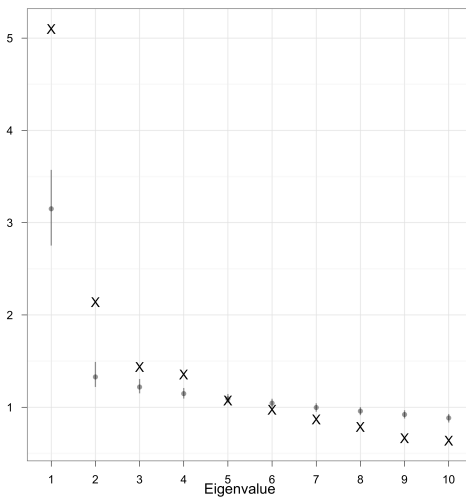


Figure 2.3: Sample Scatterplots of $d_{sld}(C^{obs}, C^{rep,m})$ versus $d_{sld}(C^{rep,m}, C^{rep,m'})$ for all replicated datasets. The grey line represents the 45 degree line.

points and intervals on the plot display the posterior predictive mean and 95% posterior predictive intervals for the eigenvalues. In Figure 2.4(a), we see a case where the model does a good job of replicating the eigenvalues of the observed rank correlation matrix. In fact, it is somewhat difficult to distinguish the observed values from the posterior predictive means. The observed values and posterior predictive values are more easily distinguished in Figure 2.4(b) where the model does not fit the data well.



(a) Good Fit



(b) Poor Fit

Figure 2.4: Sample Plots of top ten eigenvalues depicting the mean prediction (grey point) and 95% prediction intervals (grey line segment) of the eigenvalues calculated using replicated data. Eigenvalues computed from the observed data are denoted by a black “X”.

Chapter 3

INTRODUCTION TO THE SIVD STUDY

Our primary motivating example is a dataset from a large multicenter study called the Subcortical Ischemic Vascular Dementia (SIVD) Program Project Grant (Chui et al., 2006). The SIVD study follows individuals longitudinally until death, collecting serial imaging and neuropsychological data from a large group of study participants. One major goal of the study is to further elucidate relationships between brain structure and function.

Participants in the study were recruited to span a range of cognitive functioning and presence (vs absence) of lacunes. Lacunes are small areas of dead brain tissue caused by blocked or restricted blood supply. Cognitive functioning was assessed using the Clinical Dementia Rating total score, a numerical rating of dementia status based on medical history and clinical examination as well as other forms of assessment (Morris, 1993, 1997). Among the data collected in this study are the results of neuropsychological testing and standardized magnetic resonance imaging (MRI) scans of the participants' brains (Mungas, Harvey, Reed, Jagust, DeCarli, Beckett, Mack, Kramer, Weiner, Schuff, et al., 2005). A computerized segmentation algorithm classified pixels from the MRI brain scans into different components, including white matter hyperintensities (Cardenas, Ezekiel, Di Sclafani, Gomberg, and Fein, 2001).

In our applications, we focused on one particular cognitive domain, executive functioning, thought to be particularly susceptible to cerebrovascular disease (Hachinski, Iadecola, Petersen, Breteler, Nyenhuis, Black, Powers, DeCarli, Merino, Kalaria, et al., 2006). Executive functioning refers to higher order cognitive tasks ("executive" tasks) such as working memory, set shifting, inhibition, and other frontal lobe-mediated functions. In Chapters 4-6, we are concerned with relating individual levels of executive functioning at the initial SIVD study visit to the concurrent MRI-measured amount of white matter hyperintensities (WMH) located in the frontal lobe of the brain. Executive functioning capabilities may be

particularly sensitive to white matter hyperintensities in this region (Kuczynski, Targan, Madison, Weiner, Zhang, Reed, Chui, and Jagust, 2010). White matter hyperintensities are areas of increased signal intensity that are commonly associated with increasing age.

The SIVD neuropsychological battery includes 21 distinct indicators that can be conceptualized as measuring some facet of executive functioning. We refer to the executive functioning-related outcomes as “indicators” as they include some elements that are scales by themselves, and other elements that are subsets of scales. Observed responses to the SIVD neuropsychological test items are diverse in their types. In addition to binary and ordered categorical outcomes, the SIVD neuropsychological indicators include count as well as censored count data. In our analyses, we generally excluded two outcomes, Mattis Dementia Rating Scales M and N, as everyone except one participant among the analytic sample received full credit on these outcomes. As a result, we typically used 19 of the 21 executive functioning outcomes in our analyses.

Neuropsychological data in the SIVD study are available for 667 individuals, of whom 627 have a complete set of scores for the executive functioning indicators. Brain imaging data are available for 445 participants, of whom 370 have complete sets of measurements. That is, on occasion images used to estimate the occipital region volumes are cut off and prevent an accurate set of volume measurements. As a result, we only consider those participants for whom the images are complete. Finally, we restricted our analyses to participants with concurrent and complete cognitive and MRI results at their initial study visit. We defined cognitive and MRI results as concurrent when the measurements were recorded within 6 months of one another. This restriction limits the analytic sample to 341 participants. Table 4.2 displays basic information for the 19 outcomes as well as some summary statistics observed in the data for the 341 participants.

To allow the reader some familiarity with the test outcomes, we provide a brief description of selected executive functioning indicators. In the Digit Span Forward task, the participant is asked to repeat back digits in the same order they are read. The set of digits increases by one with every other trial. Testing continues until the participant is unable to repeat sets of numbers of a particular length. The total number of correctly repeated digits is the observed outcome reported in the data set. Digit Span Backward is the same as Digit

Table 3.1: Summary statistics for baseline responses of 341 participants to 19 SIVD executive functioning outcomes as well as outcome type assignment. 'RC Count' denotes a right-censored count outcome.

	Range	Mean	Median	Outcome Type
Digit Span Forward	3-12	7.69	8	Count
Digit Span Backwards	1-12	5.97	6	Count
Visual Span Forward	0-13	7.15	7	Count
Visual Span Backwards	0-12	6.18	6	Count
Verbal Fluency Letter F	1-26	11.8	12	Count
Verbal Fluency Letter A	0-40	10.2	10	Count
Verbal Fluency Letter S	0-50	12.4	12	Count
MDRS E	2-20	16.64	19	RC Count
MDRS G	0-1	0.96	1	Binary
MDRS H	0-1	0.98	1	Binary
MDRS I	0-1	0.95	1	Binary
MDRS J	0-1	0.97	1	Binary
MDRS K	0-1	0.98	1	Binary
MDRS L	0-1	0.79	1	Binary
MDRS O	0-1	0.94	1	Binary
MDRS V	9-16	14.9	16	RC Count
MDRS W	0-8	6.44	7	Ordered Cat.
MDRS X	0-3	2.66	3	Ordered Cat.
MDRS Y	0-3	2.93	3	Ordered Cat.

Span Forward except that participants are asked to repeat the number sequence backwards. For the Verbal Fluency Letter F, A, and S tests, participants are asked to name unique words starting with the specified letter. Participants are scored based on the number of

unique words listed in 1 minute.

The Mattis Dementia Rating Scale consists of questions designed to test a number of cognitive subdomains such as attention, initiation and conceptualization. We only considered a subset of these to be indicators of executive functioning. Among the Mattis Dementia Rating Scale test outcomes that we identified as executive functioning-related, Mattis Dementia Rating Scale outcome V asks a participant to examine pictures and identify similarities and differences. Mattis Dementia Rating Scale W asks a participant to compare words and identify similarities.

For Mattis Dementia Rating Scale outcome E, participants are given one minute and are asked to name as many items found in supermarkets as they can. The participant's score is the number of valid items named and the score is censored at 20. Mattis Dementia Rating Scale I instructs participants to repeat a series of alternating movements while Mattis Dementia Rating Scale L instructs participants to copy drawings of ramparts.

In Figures 3.1, we present sets of histograms for two of the executive functioning indicators. The plot on the left in each set is the histogram for all 341 study participants. The two remaining plots in each set are histograms for the sample broken down by age (years) and years of education. In Figure 3.1(a), we see the effect of the censoring described above for Mattis Dementia Rating Scale E. In fact, for most if not all of the Mattis Dementia Rating Scales, a high number of study participants attained the maximum score. In looking at the histograms by age, scores do appear to decrease for older age groups, although the effect is perhaps not as sharp as one might expect. Perhaps more pronounced is the increase of scores with education in the rightmost histograms in each set. We see similar relationships with age and education for the other outcomes.

Figure 3.2 presents the corresponding histograms for frontal white matter hyperintensity volume and total brain volume. The histogram for frontal white matter hyperintensity volume is heavily right-skewed whereas the histogram for total brain is more symmetric. The frontal white matter hyperintensity volume plots by age and education show the histogram shifting to right for older age groups and to lower values for higher educational attainment. In Figure 3.2(b), the relationships between total brain volume and age and education run in the opposite direction.

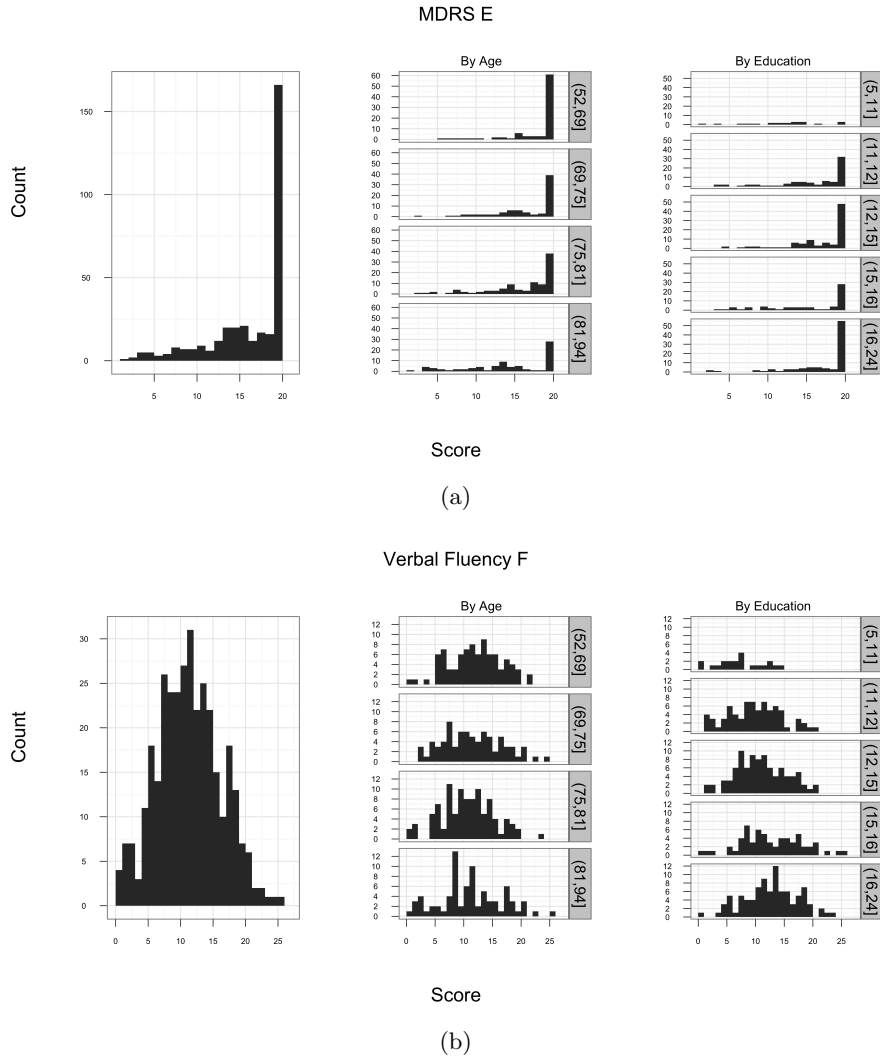


Figure 3.1: Histograms of SIVD Executive Functioning Outcomes. The plot on the left for each outcome is the histogram for the total analytic sample. The two plots in the center and on the right are histograms by age (years) and years of education.

In Figure 3.3, we display bivariate scatterplots with loess curves (local regression, Cleveland and Devlin (1988)) of some of the executive functioning indicators against the frontal white matter hyperintensity volume and total brain volume. For both Digit Span Backwards and Mattis Dementia Rating Scale W, there is a negative relationship with frontal white

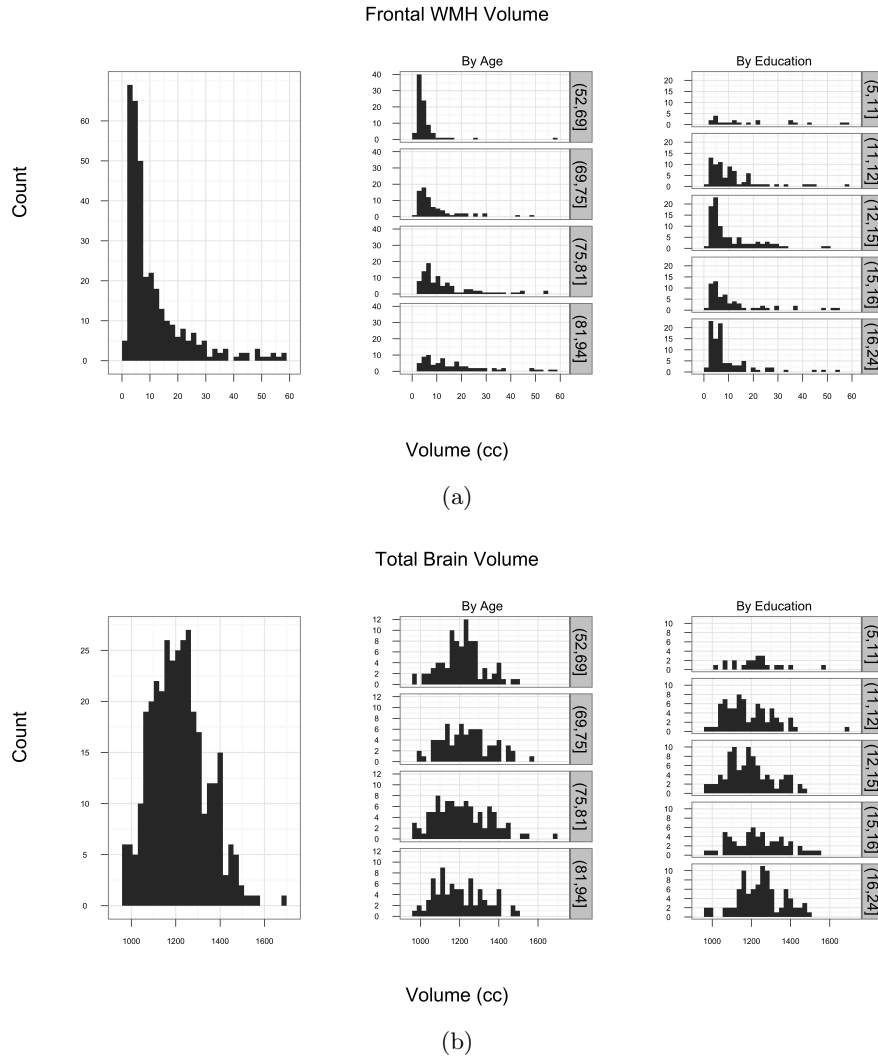


Figure 3.2: Histograms of SIVD MRI-measured Brain Volumes. The plot on the left for each outcome is the histogram for the total analytic sample. The two plots in the center and on the right are histograms by age (years) and years of education.

matter hyperintensity volume. The relationship with total brain volume is more ambiguous. Digit Span Backwards appears to possibly be positively associated with total brain volume. However, there appears to be no relationship between Mattis Dementia Rating Scale W and total brain volume.

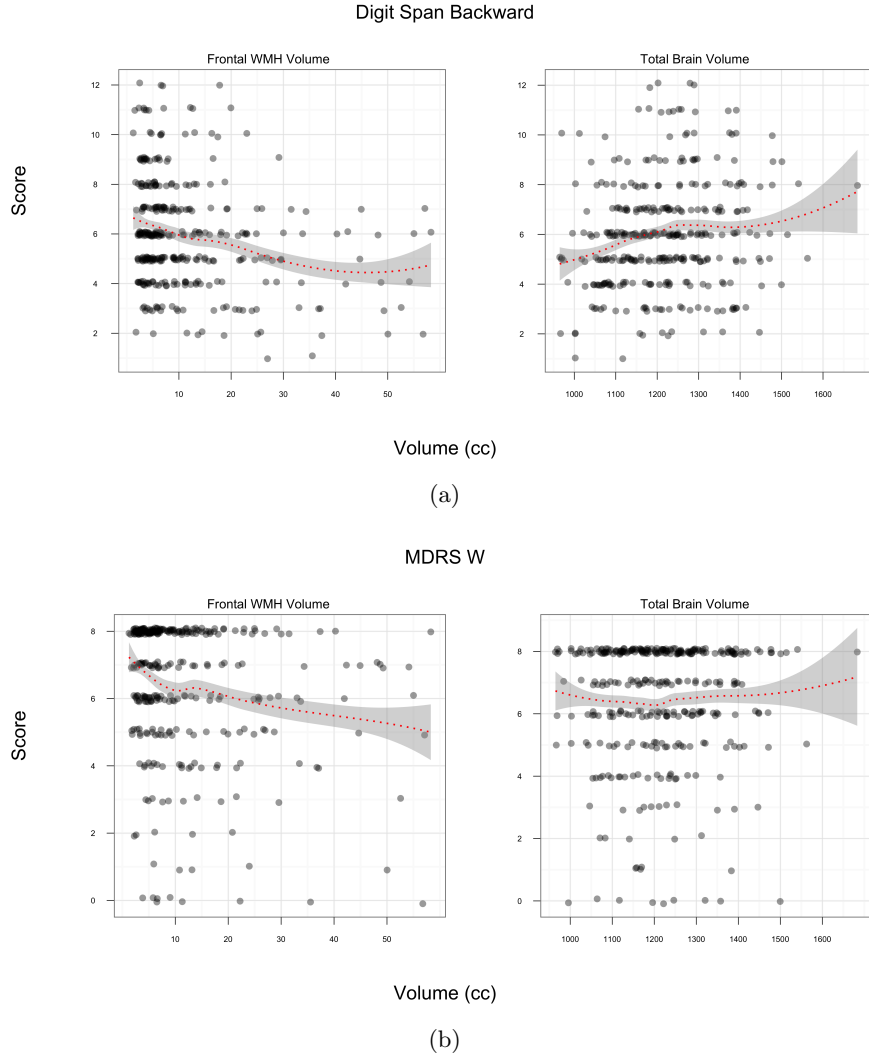


Figure 3.3: Bivariate plots of SIVD executive functioning outcome against MRI-measured brain volumes. The red dotted line is a loess curve and is surrounded by a shaded 95% pointwise confidence interval.

In Chapters 4-6, we propose models for studying the association between multivariate mixed outcomes and covariates of interest. We demonstrate these models using executive functioning indicators and frontal white matter hyperintensity volume data from the SIVD study. It should be noted that, in demonstrating our methods, we ignore possible effects of

the SIVD study design features. For example, our analyses do not account for the multicenter nature of the study. Because participants were recruited from different centers, it may be important to account for possible within-center correlations among the participants. One means of accounting for a possible multicenter effect would be a straightforward extension of the hierarchical model to include center-specific effects. Such extensions, however, are beyond the scope of this dissertation.

Chapter 4

ITEM RESPONSE THEORY MODEL FOR MIXED OUTCOMES

4.1 *Introduction*

Item response theory (IRT) is commonly applied to test questions, or items, to simultaneously estimate parameters characterizing the ‘ability’ of individuals and parameters characterizing the ‘difficulty’ of items. IRT accomplishes the dual characterization of individual and items by modeling the probability of a correct response for each individual and item conditional on a person-specific latent variable and item-specific parameters (Van der Linden and Hambleton, 1997). The most commonly applied IRT models such as the two and three parameter logistic models (Lord et al., 1968) and the graded response model (Samejima, 1970, 1972) were developed for binary and ordered categorical items. Many IRT software programs including Parscale (Muraki and Bock, 2008), BILOG-MG (Zimowski, Muraki, Mislevy, and Bock, 2008) and Testfact (Bock, Gibbons, Schilling, Muraki, Wilson, and Wood, 2003) implement estimation only for binary and ordered categorical outcomes and hence many analyses are restricted to these types of outcomes.

In cognitive testing data, such as the SIVD study introduced in Chapter 3, the testing outcomes are often mixed in the sense that some items may have binary outcomes, others have count outcomes, and others yet may be continuous. To accommodate such mixed outcomes, we apply methods from the more general category of latent variable models (Bartholomew et al., 2011), of which IRT models are a subset. We further extend these methods for additional outcome types including right-censored count encountered in our motivating example, the SIVD study and time-to-completion outcomes common in other studies involving cognitive testing such as the Alzheimer’s Disease Neuroimaging Initiative (Mueller, Weiner, Thal, Petersen, Jack, Jagust, Trojanowski, Toga, and Beckett, 2005; Weiner, Aisen, Jack, Jagust, Trojanowski, Shaw, Saykin, Morris, Cairns, Beckett, et al., 2010).

As discussed in Chapter 3, we would like to study the association between executive functioning and frontal white matter hyperintensity volume and we propose to do so through a hierarchical IRT model where the latent variable underlying the executive functioning indicators is related to the MRI measurements of frontal white matter hyperintensities. Many items used in the neuropsychological tests may not easily be classified as binary or ordinal categorical. For instance, some items in the SIVD battery closely resemble count or right-censored count outcomes. One practical approach in situations where outcomes are continuous or discrete with a large number of categories may be to bin the observations in some fashion so that we may apply the standard IRT models and software. We would like to model the respective distributions of these outcomes directly to avoid such binning.

Existing research on latent variable models for mixed outcomes is largely focused on two parametric approaches. The first approach is to specify a different generalized linear model for each outcome that best suits its type (e.g. binary, count, ordered categorical). Shared latent variables are included as predictors in each of these generalized linear models and induce dependence among the outcomes. Sammel et al. (1997) proposed this type of parametric approach to latent variable modeling for mixed outcomes in order to analyze multivariate sets of binary and continuous outcomes. Moustaki and Knott (2000) generalized the results of Sammel et al. (1997) by demonstrating the specification of a wider variety of exponential family distributions as the conditional distributions of outcomes given the latent variables. These extensions allowed for the analysis of polytomous, count and gamma distributed outcomes in addition to the binary and normally distributed outcomes analyzed by Sammel et al. (1997). The methods of Sammel et al. (1997) and Moustaki and Knott (2000) relied on maximum likelihood for inference and employed the EM algorithm for estimation. In a Bayesian framework, Dunson (2003) extended the generalized latent trait models to allow for repeated measurements, serial correlations in the latent variables and individual-specific response behavior.

The second parametric approach to analyzing mixed discrete and continuous outcomes with latent variables is the underlying latent response approach. In this approach, ordinal outcomes are assumed to have underlying latent responses that are continuous and normally distributed. As the continuous outcomes are also assumed to be normally distributed, in-

roduction of the underlying latent responses enables one to proceed with the analysis as one might for any multivariate normal data. To map the underlying latent responses to observed ordinal outcomes, one must estimate threshold parameters. The underlying latent response approach has long been in use with ordinal probit models (Aitchison and Silvey, 1957) and this approach also aids in constructing Gibbs samplers for Bayesian estimation of models for ordinal data (Johnson and Albert, 1999). In the context of latent variable models, Shi and Lee (1998) employed Bayesian estimation for factor analysis with polytomous and continuous outcomes. More recently, Fahrmeir and Raach (2007) used a similar framework that also allowed for nonlinear covariate effects through the use of penalized splines. However, as noted by Dunson (2003), the underlying latent response approach is limited in that some observed outcome types such as counts may not be easily linked to underlying continuous responses.

Generalized latent trait models, on the other hand, can be extended to account for additional outcome types. Skrondal and Rabe-Hesketh (2004) discuss a number of different outcome types that can be modeled using the generalized latent trait model framework, including censored and duration outcomes. For these reasons, we have elected to use these generalized latent trait models while retaining the two parameter IRT parameterization to obtain latent variable estimates based on observed item responses of mixed type (binary, ordered categorical, count). Motivated by data from the SIVD study and test items from other studies of cognitive functioning, we extend these models to allow for right-censored count outcomes as well as time-to-completion outcomes.

The remainder of this chapter is organized as follows. Section 4.2 discusses the generalized latent trait model cast in an IRT framework (i.e. using the IRT parameterization) as well as an extension of this framework to include outcomes of other types. Bayesian estimation of the model using a Metropolis-Hastings within Gibbs algorithm is presented in Section 4.3. We demonstrate these methods on simulated data in Section 4.4. In Section 4.5, we demonstrate these methods on data from the SIVD study, examining the association between executive functioning indicators and covariates of interest, in particular the MRI measurements of white matter hyperintensities in the frontal lobe.

4.2 Adapting IRT Models for Mixed Outcomes

4.2.1 IRT Models: Overview

We begin with a review of IRT models. For binary items, a common IRT model is the two parameter logistic form. Assume that we have observed J item responses for I individuals where a correct responses is designated with a ‘1’ and an incorrect response a ‘0’. Let the matrix $\mathbf{Y} = \{Y_{ij} : i = 1, \dots, I; j = 1, \dots, J\}$ denote the set of binary item responses. Under the two parameter logistic IRT model, the probability of a correct response on item j for participant i is the following,

$$P(Y_{ij} = 1 | \theta_i, a_j, b_j) = \frac{1}{1 + \exp(-[a_j \theta_i - b_j])}. \quad (4.1)$$

Here a_j and b_j are item specific discrimination and difficulty parameters while θ_i is an individual-specific latent trait. The latent variable θ_i is commonly referred to as an ability parameter in IRT-related literature. To remain consistent with the literature, we will use this term strictly as a name for the parameter. The difficulty parameter b_j in equation (4.1) is a location measure for the item while the discrimination parameter a_j serves as a scaling parameter. The probability of a correct response is assumed to increase with θ_i in IRT models. The scalar parameter a_j is restricted to be positive. An equivalent and perhaps more common formulation is to use the linear form $a_j(\theta_i - b_j^*)$ instead of $a_j \theta_i - b_j$ in equation (4.1) where $b_j^* = b_j/a_j$.

When working with ordinal responses of more than two categories, IRT practitioners commonly rely upon the graded response model (Samejima, 1970, 1972) or the generalized partial credit model (Muraki, 1992). As we will employ it later, we discuss the generalized partial credit model briefly. The generalized partial credit model is essentially an adjacent categories logit model. If we denote the number of categories for item j by K_j , the probability that individual i is credited with item response $k_j \in \{1, \dots, K_j\}$ is

$$P(Y_{ij} = k_j | \theta, a_j, \mathbf{b}_j) = \frac{\exp\left(\sum_{h=1}^{k_j} a_j \theta_i - b_{jh}\right)}{\sum_{l=1}^{K_j} \exp\left(\sum_{h=1}^l a_j \theta_i - b_{jh}\right)} \quad (4.2)$$

where $\mathbf{b}_j$ is now a vector of location or threshold parameters between categories. It is a

convention to set $b_{j1} = 0$ so that there are a total of $K_j - 1$ parameters. See Van der Linden and Hambleton (1997) for more background regarding IRT models.

4.2.2 Existing Generalized Latent Trait Methods Applied to IRT

We commence by presenting the approach of Moustaki and Knott (2000) and Dunson (2003) to accommodating mixed outcomes in latent variable models. This presentation proceeds in the context of IRT models by restricting the linear predictor in the latent variable model to the above two parameter IRT parameterization, $a_j\theta_i - b_j$. We follow here generally the exposition and notation of Dunson (2003).

Suppose that I participants respond to J items. The observed response on item j for participant i is represented as y_{ij} with the full set of responses for participant i denoted by the vector $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ})^T$. These outcomes may be binary, count, continuous and ordered categorical. Let J_B denote the number of binary outcome items, J_C the number of count items, J_N the number of continuous items and J_O the number of ordered categorical items so that $J = J_B + J_C + J_N + J_O$. Similarly, let $j_B = 1, \dots, J_B$ index the binary items, $j_C = 1, \dots, J_C$ index the count items and so forth so that $y_{i1}, \dots, y_{iJ_B}$ are binary responses, $y_{i1}, \dots, y_{iJ_C}$ are count responses, $y_{i1}, \dots, y_{iJ_N}$ are continuous responses and $y_{i1}, \dots, y_{iJ_O}$ are ordered categorical responses.

We further specify $\eta_{ij} = a_j\theta_i - b_j$. In the following η_{ij} , will serve as the linear predictor in the generalized latent trait model. If we designate $f_j(y_{ij}|\eta_{ij})$ as the conditional density of the response y_{ij} given the item parameters and the latent ability, we specify the densities:

$$\text{Binary: } f_{j_B}(y_{ij_B}|\eta_{ij_B}) = \frac{\exp(y_{ij_B}\eta_{ij_B})}{1 + \exp(\eta_{ij_B})}, \quad (4.3)$$

$$\text{Count: } f_{j_C}(y_{ij_C}|\eta_{ij_C}) = \exp(-\exp(\eta_{ij_C}) + \eta_{ij_C}y_{ij_C})/y_{ij_C}!, \quad (4.4)$$

$$\text{Continuous: } f_{j_N}(y_{ij_N}|\eta_{ij_N}, \tau_{j_N}) = \sqrt{\frac{\tau_{j_N}}{2\pi}} \exp\left(-\frac{\tau_{j_N}}{2}(y_{ij_N} - \eta_{ij_N})^2\right), \quad (4.5)$$

$$\text{Ordinal: } f_{j_O}(y_{ij_O}|\eta_{ij_O}) = \frac{\exp\left(\sum_{l=1}^{y_{ij_O}} \eta_{ij_O(l)}\right)}{\sum_{m=1}^{K_{j_O}} \exp\left(\sum_{l=1}^m \eta_{ij_O(l)}\right)}. \quad (4.6)$$

The parameter, τ_{j_N} , in equation (4.5) is a precision parameter (inverse of the variance) for the continuous outcome j_N . In equation (4.6), we use the generalized partial credit

model for the ordered categorical outcomes, a model that is essentially equivalent to the adjacent categories logit model. Here $\eta_{ij_{O(l)}} = a_{j_{O(l)}}\theta_i - b_{j_{O(l)}}$ where $b_{j_{O(l)}}$ specifies a threshold or transition location parameter between the l th and the $l - 1$ categories. K_{j_O} denotes the number of categories for ordinal item j_O . The above equations thus specify logistic, Poisson, normal and adjacent category logit forms for binary, count, continuous and ordinal responses.

Under the assumption of conditional independence of the responses given θ_i , the conditional joint density of observed responses is the product of the item conditional densities over all items and participants,

$$f(\mathbf{Y}|\mathbf{H}) = \prod_i \left[\prod_{j_B} f_{j_B}(y_{ij_B}|\eta_{ij_B}) \prod_{j_C} f_{j_C}(y_{ij_C}|\eta_{ij_C}) \prod_{j_N} f_{j_N}(y_{ij_N}|\eta_{ij_N}) \prod_{j_O} f_{j_O}(y_{ij_O}|\eta_{ij_O}) \right] \quad (4.7)$$

where $\mathbf{Y} = \{y_{ij} : i = 1, \dots, I; j \in \{1_B, \dots, J_B, 1_C, \dots, J_C, 1_N, \dots, J_N, 1_O, \dots, J_O\}\}$ and $\mathbf{H} = \{\eta_{ij} : i = 1, \dots, I; j \in \{1_B, \dots, J_B, 1_C, \dots, J_C, 1_N, \dots, J_N, 1_O, \dots, J_O\}\}$.

4.2.3 Extensions to Include Additional Distributions

Although existing generalized latent trait models can accommodate mixed outcomes from binary, count, continuous and ordinal distributions, other, more specialized, cases have not been demonstrated in the literature¹. For example, consider Mattis Dementia Rating Scale E, an item in the SIVD dataset where participants are asked to name as many supermarket goods as they can in 1 minute. As first presented in Chapter 3, the participant's score is the number of valid items named with 20 being the maximum score that can be attained. Using equations (4.4) and (4.6), we might consider modeling responses for Mattis Dementia Rating Scale E as count data or as ordinal data. However, there are shortcomings to both approaches. To model the responses as count data using a Poisson distribution while ignoring the restriction on scores would bias our estimates. If we rather model the outcomes as ordinal data, then we have a more flexible model form but, with 21 categories, estimating

¹Note that Skrondal and Rabe-Hesketh (2004) discuss a number of different types of outcomes that can be modeled using the generalized latent trait model framework, including censored and duration outcomes, but do not demonstrate those discussed here in applied settings.

20 transition location parameters may pose a challenge with regards to slow convergence and prior specification.

Alternatively, we might consider modeling the Mattis Dementia Rating Scale E responses using a right-censored Poisson distribution. Censored distributions are useful in cases where responses known to be outside of a specified range are assigned a value within the range. For Mattis Dementia Rating Scale E, we know that participants with a score of 19 or below on the item have not been limited by the maximum attainable score of 20. For those with scores of 20, however, we cannot determine whether the individual scored 20, or perhaps scored higher but then were credited with a score of 20 due to the item's scoring rules. As such, for participants with scores of 20, we only know definitively that participant scored higher than 19.

In general terms, let x be a random variable with density f , then z is a right-censored observation of x if

$$z = \begin{cases} x & \text{if } x \leq d_2 \\ d_2 + 1 & \text{if } x > d_2 \end{cases}$$

where d_2 is the point of censoring. Thus, for Mattis Dementia Rating Scale E, $d_2 = 19$. The random variable z then has the density f_{RC}

$$f_{RC}(z) = [1 - F(d_2)]^{\mathbb{1}_{[z > d_2]}} \cdot f(z)^{\mathbb{1}_{[z \leq d_2]}}. \quad (4.8)$$

If y_{ij} has a right-censored Poisson distribution with mean parameter λ_{ij} , then its density $f(y_{ij})$ is as follows:

$$f(y_{ij}|\lambda_{ij}) = \left[\sum_{z=d_2+1}^{\infty} \frac{\exp(-\lambda_{ij})\lambda_{ij}^z}{z!} \right]^{\mathbb{1}_{[y_{ij} > d_2]}} \cdot \left[\frac{\exp(-\lambda_{ij})\lambda_{ij}^{y_{ij}}}{y_{ij}!} \right]^{\mathbb{1}_{[y_{ij} \leq d_2]}}. \quad (4.9)$$

If we let $\log \lambda_{ij} = \eta_{ij}$ as we did for the uncensored count data model, then

$$f(y_{ij}|\eta_{ij}) = \left[\sum_{z=d_2+1}^{\infty} \frac{\exp(-\exp(\eta_{ij})) \exp(z\eta_{ij})}{z!} \right]^{\mathbb{1}_{[y_{ij} > d_2]}} \cdot \left[\frac{\exp(-\exp(\eta_{ij})) \exp(y_{ij}\eta_{ij})}{y_{ij}!} \right]^{\mathbb{1}_{[y_{ij} \leq d_2]}}. \quad (4.10)$$

Thus we can add right-censored count outcomes to the mixed outcomes IRT model discussed in the previous section. We index these items using $j_{CRC} = 1, \dots, J_{CRC}$.

We might not only consider the use of the right-censored Poisson distribution for count outcomes but also for ordinal categorical outcomes with a large number of categories where a pronounced ceiling effect is evident. This would reduce the number of parameters to estimate as the $k - 1$ parameters required for the category thresholds would be reduced to a single difficulty parameter. Moreover, one could argue that this is not a significant departure in the description of the outcome as an individual is effectively trying to attain a high count of points scored on the item. For individuals who have attained the maximum number of points, we only know that their ability is greater or equal to that necessary to achieve this maximum.

Although not encountered in the SIVD set, many cognitive assessments such as that administered in the Alzheimer’s Disease Neuroimaging Initiative include timed activity items. One such example is the Trail Making Test Parts A and B. For both, the examiner presents the participant with 25 circles distributed over a sheet of paper. In Trail Making Test Part A, the circles are labeled 1-25 and the participant is asked to draw lines to connect the numbered circles in ascending order. Part B of the test is comprised of circles labeled with both numbers (1-13) and letters (A-L) and the participant is asked to draw lines connecting the circles, alternating between numbers and letters in ascending fashion (1-A-2-B...L-13). For both parts, the examiner instructs the participant to connect the circles as quickly as possible without lifting the writing instrument from the paper. The examiner then records the score as the number of seconds it takes the patient to complete the task.

Given the continuous nature of the scores, we might consider modeling time to completion items such as Trail Making A and B with a normal distribution as discussed above. However, a distribution with strictly positive support may be more appropriate for these timed outcomes. Drawing from survival analysis and time-to-event analyses, we propose a lognormal regression model.

Additionally recall that IRT models typically assume the item response function to be a monotone function of the ability parameter. Because a greater time score may imply less ability (for example, in the case of the Trail Making Test items), the above model would

violate the standard IRT assumptions. To rectify this, we work with the inverse of the timed score. Thus if we now denote y_{ijT}^* as the original timed score, let $y_{ijT} = 1/y_{ijT}^*$ denote the inverse. If we index the timed items j_T , then we specify the density of the inverse of the timed score, y_{ijT} as

$$f_{jT}(y_{ijT}|\eta_{ijT}, \tau_{jT}) = \sqrt{\frac{\tau_{jT}}{2\pi}} \frac{1}{y_{ijT}} \exp\left(-\frac{\tau_{jT}}{2} (\log y_{ijT} - \eta_{ijT})^2\right). \quad (4.11)$$

There is one additional feature to consider when modeling the Trail Making Test items, a feature that is common to many timed items. For Trail Making Test Part B, the examiner discontinues timing after 5 minutes and records a time of 5 minutes for participants who have not finished the task. For Trail Making Test Part A, the time limit is 2.5 minutes. The recorded time is hence right-censored and one needs to consider a censored lognormal distribution. Using a left-censored lognormal regression model for the inverse timed score y_{ijT} , we have

$$\begin{aligned} f_{jT_{LC}}(y_{ijT_{LC}}|\eta_{ijT_{LC}}, \tau_{jT_{LC}}) &= \left[\int_{-\infty}^{d_1} \sqrt{\frac{\tau_{jT_{LC}}}{2\pi}} \frac{1}{z} \exp\left(-\frac{\tau_{jT_{LC}}}{2} (\log z - \eta_{ijT_{LC}})^2\right) dz \right]^{\mathbb{1}_{[y_{ijT_{LC}} < d_1]}} \\ &\quad \cdot \left[\sqrt{\frac{\tau_{jT_{LC}}}{2\pi}} \frac{1}{y_{ijT_{LC}}} \exp\left(-\frac{\tau_{jT_{LC}}}{2} (\log y_{ijT_{LC}} - \eta_{ijT_{LC}})^2\right) \right]^{\mathbb{1}_{[y_{ijT_{LC}} \geq d_1]}} \end{aligned} \quad (4.12)$$

where jT_{LC} indexes the left-censored inverse timed score and d_1 is the point at which the data are censored on the left. We can hence model censored timed activity items in our IRT for mixed outcomes framework.

4.2.4 Item Parameter Interpretation

Although we have relied upon the same linear expression $\eta_{ij} = a_j\theta_i - b_j$ in the modeling of each outcome, it is important to note that the item parameters a_j and b_j do not hold the same interpretation across different types of outcomes. Whereas a_j and b_j for a binary item define the shape of curve representing the probability of a correct response across levels of ability, the interpretation of these parameters will be different for a count item modeled with a Poisson distribution.

To better explicate this, we rely on the interpretation from generalized linear models. For a binary outcome, for example, we might interpret the discrimination parameter a_j as the increase in the log odds of answering that item correctly that is associated with a 1-unit increase in the latent ability or, equivalently, we might interpret $\exp(a_j)$ as the multiplicative increase in the odds of answering that item correctly associated with a 1-unit increase in ability. Meanwhile, for a count outcome, we would interpret $\exp(a_j)$ as the multiplicative increase in the expected count for that item associated with a 1-unit increase in ability. Similar differences in interpretations for other types of outcomes can be described by using the generalized linear models analogy.

Moustaki and Knott (2000) provide some means by which one might standardize the item parameters a_j for normal, binary and ordinal categorical items. As discussed in Section 4.1, our primary interest is in relating executive functioning indicators to MRI-measured brain volumes via a latent variable so interpretation of the item parameters may be considered secondary. However, for those interested in comparing items, such differences in interpretation should be acknowledged.

4.2.5 Hierarchical IRT Model For Mixed Outcomes

Because we are interested in relating the level of cognitive functioning to covariates including MRI brain volume measurements, we extend the IRT model for mixed outcomes hierarchically with the linear equation for θ_i ,

$$\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim N(0, 1). \quad (4.13)$$

Here, the vector $\mathbf{x}_i = (x_{i1}, \dots, x_{ip}, \dots, x_{iP})$ represents the covariate measurements for participant i and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p, \dots, \beta_P)$ the population coefficients. We may then make inference on a parameter of interest β_p to investigate the association between a covariate of interest and the ability parameter or, indirectly through the latent variable, the general association between the covariate of interest and the outcomes. In the case of the SIVD analysis, we relate volume measurements of white matter hyperintensities in the frontal lobe to executive functioning ability as measured by item responses of mixed types (binary, count, continuous, etc.).

Extension of the model hierarchically to include equation (4.13) follows the multiple indicator multiple cause (MIMIC) model (Goldberger and Hauser, 1971; Jöreskog and Goldberger, 1975) and the multilevel IRT model described by Fox and Glas (2001) where covariates are related to outcomes through the latent variable or ability parameter only, rather than also relating covariates directly to outcomes as in other formulations (Dunson, 2003). This formulation specifies that two individuals, i and i' , with the same value for the ability parameter ($\theta_i = \theta_{i'}$) have the same response probability. In applying equation (4.13), we are assuming that the difficulty or discriminating nature of the item does not vary for individuals with the same ability parameter values but with different covariate values.

Consider the logit probability of a binary item using equations (4.13) and (4.1),

$$\text{logit}(P(Y_{ij}|\theta_i, a_j, b_j)) = a_j\theta_i - b_j \quad (4.14)$$

$$= a_j(\mathbf{x}_i^T\boldsymbol{\beta} + \epsilon_i) - b_j \quad (4.15)$$

$$= a_j\mathbf{x}_i^T\boldsymbol{\beta} + a_j\epsilon_i - b_j. \quad (4.16)$$

Here, all individual-level variation, whether explained ($\mathbf{x}_i^T\boldsymbol{\beta}$) or unexplained (ϵ_i), is scaled by a_j and affects the response probability in the same manner.

We could relax this restriction and modify equation (4.16) so that

$$\text{logit}(P(Y_{ij}|\mathbf{x}_i, \boldsymbol{\beta}, \epsilon_i, \omega_j, a_j, b_j)) = \omega_j\mathbf{x}_i^T\boldsymbol{\beta} + a_j\epsilon_i - b_j, \quad (4.17)$$

where ω_j is not restricted to equal a_j . The case where the response probability varies for individuals with different covariate values but equal ability parameter values is often referred to as test item bias. In our analyses of the SIVD dataset in this chapter and Chapter 5, we assume no test item bias is present and incorporate covariates into the model using equation (4.13). The semiparametric multivariate regression model used in Chapter 6 does not use this assumption.

To ensure identifiability of an IRT model, one often specifies the latent variable as normally distributed with a fixed variance and mean. The variance of the latent distribution is still fixed in the hierarchical formulation listed above. The mean term is identifiable because we do not include an intercept term α in equation (4.13). If we consider the linear

predictor used in the likelihood,

$$\eta_{ij} = a_j \theta_i - b_j, \quad (4.18)$$

we can incorporate equation (4.13) with an intercept included and see that

$$\begin{aligned} \eta_{ij} &= a_j(\alpha + \beta_1 x_{i1} + \dots \beta_i x_{ip} + \epsilon_i) - b_j \\ &= a_j(\alpha + \beta_1 x_{i1} + \dots \beta_i x_{ip} + \epsilon_i) + a_j c - b_j - a_j c \\ &= a_j((\alpha + c) + \beta_1 x_{i1} + \dots \beta_i x_{ip} + \epsilon_i) - (b_j + a_j c) \\ &= a_j(\alpha^* + \beta_1 x_{i1} + \dots \beta_i x_{ip} + \epsilon_i) - b_j^*, \end{aligned} \quad (4.19)$$

where c is an arbitrary constant. Due to this lack of identifiability, we do not include an intercept, α , in our model. As we are focused on the association between the latent variable and covariates of interest, the exclusion of this parameter is not important.

4.3 Estimation

Estimation of the hierarchical IRT model for mixed outcomes proceeds using Bayesian methods relying upon Markov chain Monte Carlo (MCMC) methods to approximate the posterior distributions. This section follows the computational framework laid out in Patz and Junker (1999b) and Patz and Junker (1999a) for standard IRT models and extends it to estimate the IRT model for mixed outcomes. Their estimation approach employs a Metropolis-Hastings (MH) within Gibbs sampling approach.

4.3.1 Prior Distributions

We use Lognormal($0, \sigma_a^2$) and $N(0, \sigma_b^2)$ prior distributions for the item parameters a_j and b_j respectively. For the precision parameters τ_j related to outcomes modeled with normal or lognormal distributions, we assigned Gamma(ν, ϕ) distributions. Finally, we specified $N(0, \sigma_\beta^2)$ prior distributions for the regression coefficients in equation (4.13). The complete

hierarchical model is

$$y_{ij} \sim \begin{cases} F_j(a_j, b_j, \theta_i), & j \in \{j_B, j_C, j_O, j_{RC}\} \\ F_j(a_j, b_j, \tau_j, \theta_i), & j \in \{j_N, j_T, j_{LC}\} \end{cases} \quad (4.20)$$

$$a_j \sim \text{Lognormal}(0, \sigma_a^2), \quad \forall j \quad (4.21)$$

$$b_j \sim \text{N}(0, \sigma_b^2), \quad \forall j \quad (4.22)$$

$$\tau_j \sim \text{Gamma}(\nu, \phi), \quad j \in \{j_N, j_T, j_{LC}\} \quad (4.23)$$

$$\theta_i \sim \text{N}(\mathbf{x}_i^T \boldsymbol{\beta}, 1) \quad (4.24)$$

$$\beta_p \sim \text{N}(0, \sigma_\beta^2). \quad (4.25)$$

Note that for ordinal categorical items, each threshold parameter $b_{j_O(l)}$ is also assumed to have a $\text{N}(0, \sigma_b^2)$ prior. In the non-hierarchical version of the model, the prior for θ_i in equation (4.24) is replaced by

$$\theta_i \sim \text{N}(0, 1). \quad (4.26)$$

4.3.2 Sampling From Posterior Distributions

The MH within Gibbs sampling scheme proceeds as follows. Let the superscript (m) for a parameter denote the m th draw of that parameter while sampling. For MH steps, we generate proposals for the ability parameter θ and the item parameter b using random walk proposals from a normal distribution while proposals for the item parameter a are generated using a lognormal distribution.

1. **Sample ability θ .** For each participant $i = 1, \dots, I$, draw $\theta_i^{(m)}$ from $p\left(\theta_i^{(m)} | \theta_i^{(m-1)}, \mathbf{y}_i, \mathbf{a}^{(m-1)}, \mathbf{b}^{(m-1)}, \boldsymbol{\tau}^{(m-1)}, \mathbf{x}_i, \beta^{(m-1)}\right)$ via a MH step.
2. **Sample item parameters $\mathbf{a}, \mathbf{b}$.** For items where the responses are modeled using normal and lognormal distributions, we can obtain conditional distributions for the difficulty parameters in a closed form. For these item responses modeled with normal distributions,

$$b_j^{(m)} \sim \text{N}\left(\frac{\tau_j^{(m-1)} \sigma_b^2}{I \tau_j^{(m-1)} \sigma_b^2 + 1} \left[a_j^{(m-1)} \sum_i \theta_i - \sum_i y_{ij} \right], \frac{\sigma_b^2}{I \tau_j^{(m-1)} \sigma_b^2 + 1}\right), \quad (4.27)$$

where σ_b^2 is the variance of the prior distribution specified for b_j . For lognormal item responses,

$$b_j^{(m)} \sim N \left(\frac{\tau_j^{(m-1)} \sigma_b^2}{I \tau_j^{(m-1)} \sigma_b^2 + 1} \left[a_j^{(m-1)} \sum_i \theta_i - \sum_i \log y_{ij} \right], \frac{\sigma_b^2}{I \tau_j^{(m-1)} \sigma_b^2 + 1} \right). \quad (4.28)$$

Subsequently, $a_j^{(m)}$ is drawn from $p(a_j^{(m)} | a_j^{(m-1)}, b_j^{(m)}, \mathbf{y}_j, \boldsymbol{\theta}^{(m)}, \boldsymbol{\tau}^{(m-1)})$ via a MH step.

For each outcome $j = 1, \dots, J$ that has not been modeled as a normally or lognormally distributed outcome, $(a_j^{(m)}, b_j^{(m)})$ are drawn jointly from $p(a_j^{(m)}, b_j^{(m)} | a_j^{(m-1)}, b_j^{(m-1)}, \mathbf{y}_j, \boldsymbol{\theta}^{(m)}, \boldsymbol{\tau}^{(m-1)})$ via a MH step.

3. **Sample precision parameters $\boldsymbol{\tau}$.** This step only occurs for item responses modeled as normal or lognormal outcomes, including left- and right-censored lognormal outcomes. For the censored outcomes, $\tau_j^{(m)}$ is drawn from $p(\tau_j^{(m)} | \tau_j^{(m-1)}, a_j^{(m)}, b_j^{(m)}, \mathbf{y}_j, \boldsymbol{\theta}^{(m)})$ using MH. For the normally distributed outcomes, we draw from

$$\tau_j^{(m)} \sim \text{Gamma} \left(I/2 + \nu, \frac{2\phi}{\phi \sum_i \left(y_{ij} - \eta_{ij}^{(m)} \right)^2 + 2} \right) \quad (4.29)$$

where $\eta_{ij}^{(m)} = a_j^{(m)} \theta_i^{(m)} - b_j^{(m)}$ and ν, ϕ are the hyperparameters specified for the prior distribution for τ . For the item responses modeled with lognormal distributions,

$$\tau_j^{(m)} \sim \text{Gamma} \left(I/2 + \nu, \frac{2\phi}{\phi \sum_i \left(\log y_{ij} - \eta_{ij}^{(m)} \right)^2 + 2} \right). \quad (4.30)$$

4. **Sample coefficients $\boldsymbol{\beta}$.** For the last step of our sampling scheme, the conditional distribution for $\boldsymbol{\beta}$ is available in a closed form. With $\mathbf{X}$ denoting an $I \times p$ matrix of covariates, we sample the parameters jointly from a multivariate normal distribution specified as follows,

$$\boldsymbol{\beta}^{(m)} \sim N \left(\left(\mathbf{X}^T \mathbf{X} + \frac{1}{\sigma_\beta^2} \mathbf{I}_{p \times p} \right)^{-1} \mathbf{X}^T \boldsymbol{\theta}^{(m)}, \left(\mathbf{X}^T \mathbf{X} + \frac{1}{\sigma_\beta^2} \mathbf{I}_{p \times p} \right)^{-1} \right). \quad (4.31)$$

4.4 A Simulated Data Example

To demonstrate and test our model, we generated data according to the model and examined the ability of the model estimation process to recover the data-generating parameter values. We subsequently considered the fit of the model to the simulated data using the posterior predictive model checking methods introduced in Section 2.4.

We tested the non-hierarchical version of the IRT model for mixed outcomes (equations 4.20-4.23, 4.26) using simulated data. We simulated responses for 400 hypothetical individuals and 20 hypothetical items. The items consist of 7 binary, 5 ordinal categorical, 3 count, 3 right-censored count and 2 left-censored positive continuous (e.g. timed) outcomes. These different types of responses were generated using the conditional distributions, including the right-censored Poisson and left-censored lognormal distributions. We then used the same model to estimate the parameter values. We drew 40,000 samples according to the MH within Gibbs algorithm presented in section 4.3.2, conservatively discarding the first half as burn-in.

Table 4.1 displays both the data-generating values and the posterior summaries for the discrimination parameters, **a**. The posterior summaries include the bounds of the 95% posterior credible interval. We saw that the model does a reasonable job of recovering the data-generating parameters. The posterior mean was generally close to the true value and the 95% posterior credible interval contains the true value in all cases. For the item difficulty parameters, the 95% posterior credible interval contained the true value in 36 of 39 (92.3%) cases and, for the latent trait parameters, in 96% of the cases. Next, we utilized the posterior predictive model checks to examine the fit of the model to the simulated data.

We first examined the model's ability to replicate the marginal distributions of the responses. Figure 4.1 presents histograms for simulated items 12 and 20. Item 12 was generated using a right-censored Poisson distribution and item 20 using the generalized partial credit model. In both cases, the model fit the marginal distributions reasonably well with one exception for a score of 8 on item 12 (Figure 4.1(a)). Examining the plots of the marginal distributions for other outcomes, we found that the model fit other marginals similarly well.

Table 4.1: Data-generating values and posterior summaries for discrimination parameters, $\mathbf{a}$, from simulated data examples using the IRT model for mixed outcomes.

Item	Posterior			
	Truth	Mean	Median	95% CI
1	1.37	1.32	1.31	(1.02, 1.64)
2	1.04	0.84	0.83	(0.56, 1.13)
3	0.62	0.56	0.56	(0.35, 0.80)
4	1.73	1.43	1.43	(1.11, 1.77)
5	0.92	0.81	0.80	(0.57, 1.06)
6	1.02	0.97	0.97	(0.71, 1.26)
7	1.34	1.21	1.21	(0.84, 1.61)
8	0.53	0.60	0.60	(0.45, 0.76)
9	0.68	0.61	0.61	(0.54, 0.68)
10	0.74	0.68	0.67	(0.54, 0.82)
11	1.15	1.12	1.11	(1.01, 1.22)
12	1.04	0.98	0.98	(0.89, 1.07)
13	1.76	1.63	1.62	(1.45, 1.81)
14	1.18	1.11	1.11	(1.00, 1.23)
15	1.40	1.40	1.39	(1.22, 1.58)
16	0.53	0.48	0.48	(0.36, 0.60)
17	1.00	0.91	0.91	(0.73, 1.10)
18	1.86	1.68	1.67	(1.37, 2.03)
19	1.58	1.49	1.48	(1.19, 1.81)
20	1.24	1.45	1.44	(1.08, 1.90)

Next, we reviewed the model's ability to replicate the rank correlations (Kendall's τ) observed in the data. Figures 4.2 and 4.3 present a correlation distance plot and an eigenvalue

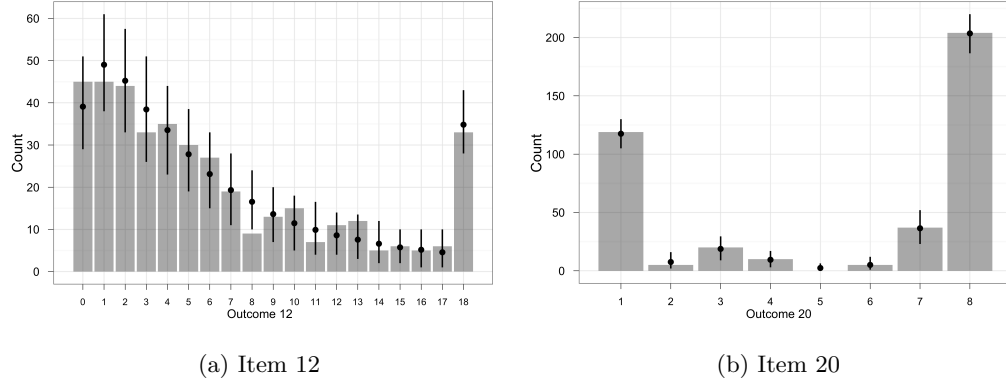


Figure 4.1: Histograms of the observed item scores in the simulated dataset. The black points indicate the mean count across replicated datasets for each score. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.

plot. The model appeared to do a good job of reproducing the rank correlations observed in the data.

Finally, we examined the pairwise rank correlations of individual items. Figure 4.4 presents the pairwise correlation plots for items 3 and 14. In both cases, the model fit the observed pairwise rank correlations well. Plots for other outcomes showed similar success. Overall, we observed good recovery of data-generating parameters from the simulated data.

4.5 Application to the SIVD Study

We now demonstrate our IRT model for mixed outcomes using executive functioning data from the SIVD study. Recall that we are interested in relating the individual's level of executive functioning to the amount of white matter hyperintensities located in the frontal lobe at the first study visit. Table 4.2 displays labels as well as some summary statistics for the 19 items observed in the data for $I = 341$ participants. For this analysis, we consider only participants with a complete set of response to the executive functioning items as well as a concurrent set of brain MRI measurements. Concurrent brain measurements here are defined as within six months (before or after) of the neuropsychological testing date.

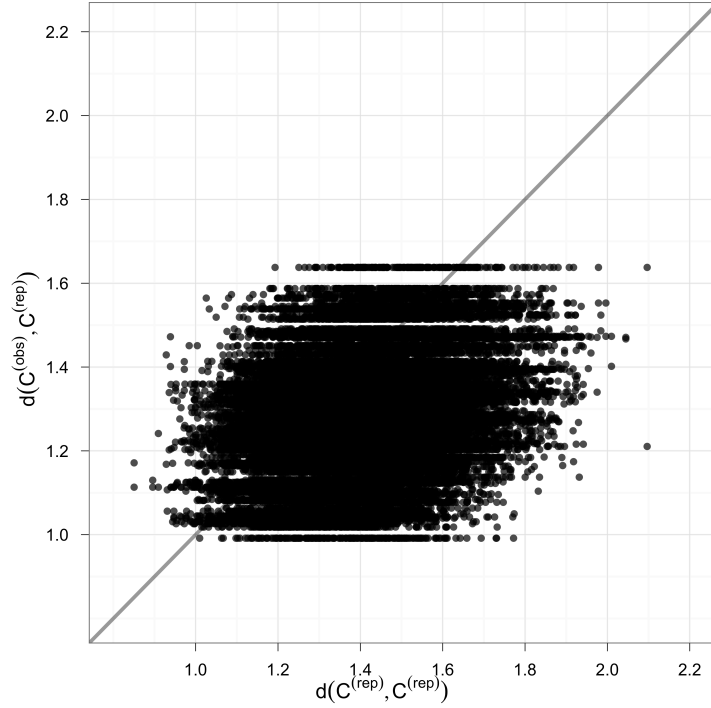


Figure 4.2: Scatterplots of $d_{std}(C^{obs}, C^{rep, m})$ versus $d_{std}(C^{rep, m}, C^{rep, m'})$ for all replicated datasets. The grey line represents the 45 degree line.

On the far right of the table, we have also listed the outcome type specified in our model for each item. As one can see from the summary statistics, the items vary greatly in their number of categories as well as in their difficulty. For many of the binary items as well as the Mattis Dementia Rating Scale E and V, the mean and median scores are very close to the maximum score. In the latter two outcomes, this indicates a possible ceiling effect.

Now consider Mattis Dementia Rating Scale E discussed in Section 4.2 as a motivating example for modeling right-censored outcomes. Recall that participants are given 1 minute to name as many supermarket goods as they can. The participant's score is the number of valid goods named and the score is censored at 20. A histogram of scores for this item is presented in Figure 4.5(a). One can see evidence of a ceiling effect for this item. Given the description of this item and its distribution of responses, treatment of the item as a

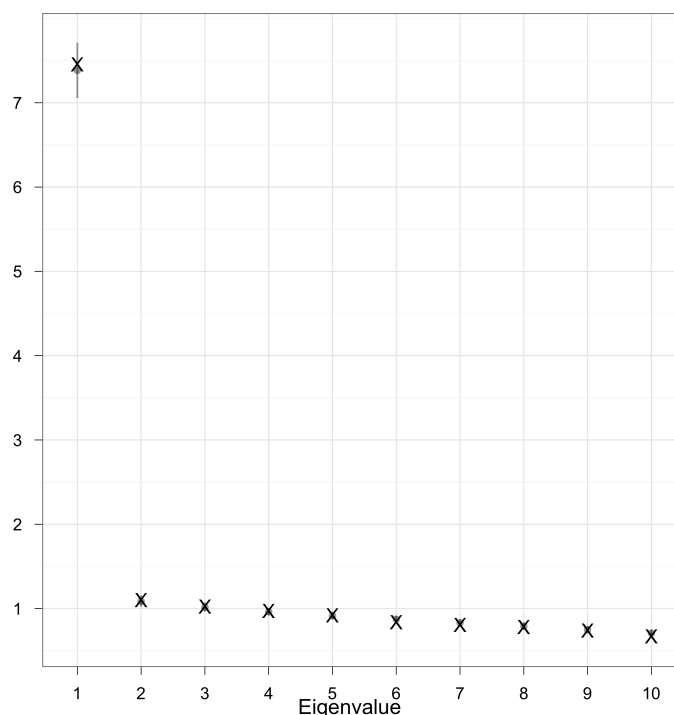
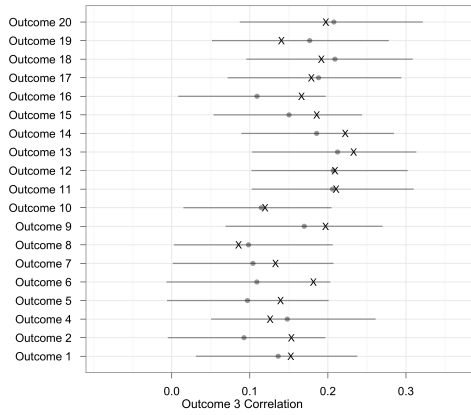


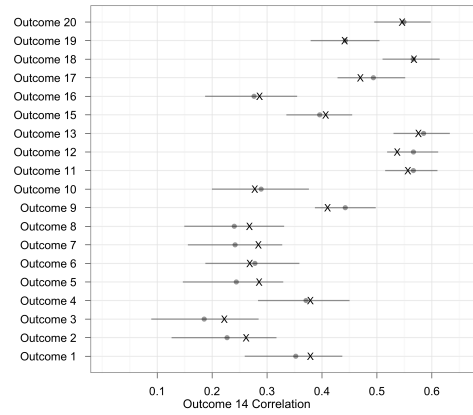
Figure 4.3: Plot of ten largest eigenvalues depicting the mean prediction (grey point) and 95% prediction intervals (grey line segment) of the eigenvalues calculated using replicated data. Eigenvalues computed from the observed data are denoted by a black “X”.

right-censored outcome is warranted.

The Mattis Dementia Rating Scale V asks a participant to examine pictures and identify similarities and differences. The closely related Mattis Dementia Rating Scale W asks a participant to compare words and identify similarities. Figure 4.5(b) depicts a histogram of the scores for Mattis Dementia Rating Scale W. For both outcomes, we see evidence of ceiling effects. Although each item’s description does not suggest its treatment as a right-censored outcome, we might nonetheless treat it as such rather than treat it as an ordinal categorical item. As discussed previously, the modeling of the outcome as a right-censored count will reduce the number of parameters to estimate and does not significantly alter the conceptual description of the item. In the case of Mattis Dementia Rating Scale V with

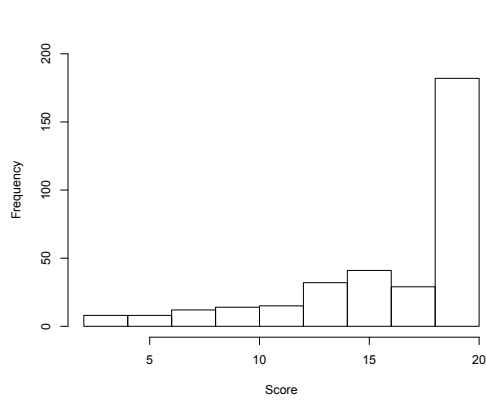


(a) Item 3

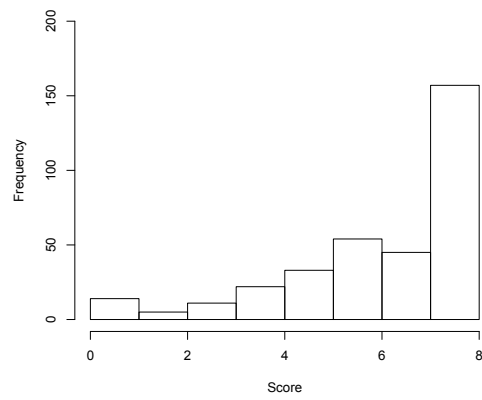


(b) Item 14

Figure 4.4: Pairwise correlation plots for a single outcome against the remaining outcomes depicting the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) for Kendall's τ values calculated using replicated data. Kendall's τ values computed from the observed data are denoted by a black "X".



(a) MDRS E



(b) MDRS W

Figure 4.5: Histograms of scores for Mattis Dementia Rating Scale E and W.

Table 4.2: Summary statistics for baseline responses of 341 participants to 19 SIVD executive functioning outcomes as well as outcome type assignment. ‘RC Count’ denotes a right-censored count outcome.

	Range	Mean	Median	Outcome Type
Digit Span Forward	3-12	7.69	8	Count
Digit Span Backwards	1-12	5.97	6	Count
Visual Span Forward	0-13	7.15	7	Count
Visual Span Backwards	0-12	6.18	6	Count
Verbal Fluency Letter F	1-26	11.8	12	Count
Verbal Fluency Letter A	0-40	10.2	10	Count
Verbal Fluency Letter S	0-50	12.4	12	Count
MDRS E	2-20	16.64	19	RC Count
MDRS G	0-1	0.96	1	Binary
MDRS H	0-1	0.98	1	Binary
MDRS I	0-1	0.95	1	Binary
MDRS J	0-1	0.97	1	Binary
MDRS K	0-1	0.98	1	Binary
MDRS L	0-1	0.79	1	Binary
MDRS O	0-1	0.94	1	Binary
MDRS V	9-16	14.9	16	RC Count
MDRS W	0-8	6.44	7	Ordered Cat.
MDRS X	0-3	2.66	3	Ordered Cat.
MDRS Y	0-3	2.93	3	Ordered Cat.

possible scores from 0 to 16 (although the lowest observed in this study is 9), we model the outcome as a right-censored count. In the case of Mattis Dementia Rating Scale W with a maximum score of only 8, we model the outcome as ordered categorical.

Having established the different conditional distributions for SIVD executive functioning items that we specify in the IRT model for mixed outcomes, we now address the relation of our estimate of the latent ability parameter to the variable of interest, volume of frontal white matter hyperintensities, as well as covariates for which we would like to adjust. These include age, sex, education and total brain volume. This leads to the regression equation:

$$E[\theta_i] = \beta_1 \text{Sex}_i + \beta_2 \text{Educ}_i + \beta_3 \text{Age}_i + \beta_4 \text{Vol}_i + \beta_5 \text{WMH}_i. \quad (4.32)$$

where Sex_i is the participant's sex (Female=1, Male=0), Educ_i is the participant's number of years of education attained, Age_i is the age of the participant at the time of examination, Vol_i is the total brain volume of the participant, and WMH_i is the frontal white matter hyperintensity volume. We used standardized versions of the continuous predictor variables.

Implementation of the MH within Gibbs sampler resulted in high autocorrelations within the item parameters as evidenced by low effective sample sizes and large number of iterations recommended by the Raftery-Lewis diagnostic (Raftery and Lewis, 1995). Ultimately, we used 100,000 iterations for the generalized latent trait model. We discarded the first 20,000 draws as burn-in. We employed the Geweke (Geweke, 1992) and Gelman-Rubin (Gelman and Rubin, 1992) diagnostics to check convergence.

Table 4.3 displays estimation results for the regression parameters in the model. Both point estimates and 95% credible intervals are presented. Based on the IRT model for mixed outcomes, we expect a 1 SD increase in frontal white matter hyperintensity volume to be associated with a -0.511 decrease in the latent variable with 95% posterior probability that the parameter occurs in the range (-0.654, -0.376). Note that this range does not include 0. Thus the posterior distribution for the coefficient for frontal white matter hyperintensity volume suggests a negative association between executive functioning ability and frontal white matter hyperintensity volume.

We now consider the fit of the model to the data using the posterior predictive model checking methods introduced in section 2.4. We first examined the marginal distributions of the responses. Figure 4.6 presents the histograms for the observed responses to Verbal Fluency A and Mattis Dementia Rating Scale E along with posterior predictive summaries. We specified Poisson and right-censored Poisson distributions as the conditional distribu-

Table 4.3: Posterior summaries for regression coefficients estimated using the IRT model for mixed outcomes.

Coefficient	Mean	Median	95% CI
Sex	0.655	0.657	(0.273, 1.021)
Education	0.512	0.513	(0.381, 0.644)
Age	-0.169	-0.169	(-0.307, -0.037)
Total Brain Vol.	0.147	0.148	(-0.022, 0.314)
Frontal WMH Vol.	-0.511	-0.510	(-0.654, -0.376)

tions for the respective outcomes. Overall, the model does a decent job of replicating the marginal distribution for the Verbal Fluency item. The fit of the model is not as assured for Mattis Dementia Rating Scale item E where the model generally under-predicts the number of participants to receive a score of 20.

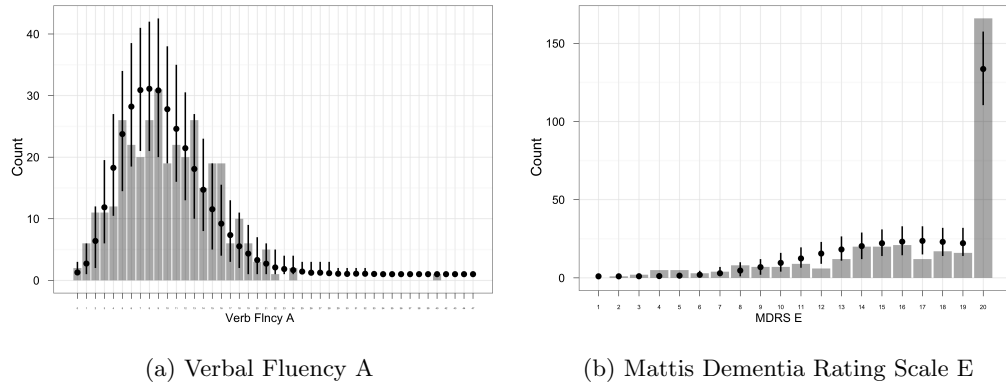


Figure 4.6: Histograms of the observed scores for the Verbal Fluency A and Mattis Dementia Rating Scale E. The black points indicate the mean count across replicated datasets for each score. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.

Having seen some mixed results in the model fit of the marginal distributions, we turn our attention to the dependence structure. On a global level, we compare $\{d(C^{obs}, C^{rep,m})\}$ and $\{d(C^{rep,m}, C^{rep,m'})\}$. The left-hand plot in Figure 4.7 displays these sets of points, $\{d(C^{obs}, C^{rep,m})\}$ and $\{d(C^{rep,m}, C^{rep,m'})\}$, against one another. We see that the cloud of points lies well to the left of the 45 degree line, sitting in the upper left of the plot. This suggests poor model fit of the dependence structure among the outcomes. The right-hand plot in Figure 4.7 further shows that the model does not replicate the eigenvalues of the observed rank correlation matrix very well. The large second eigenvalue further suggests that we may need to incorporate more than one latent variable to describe dependencies successfully.

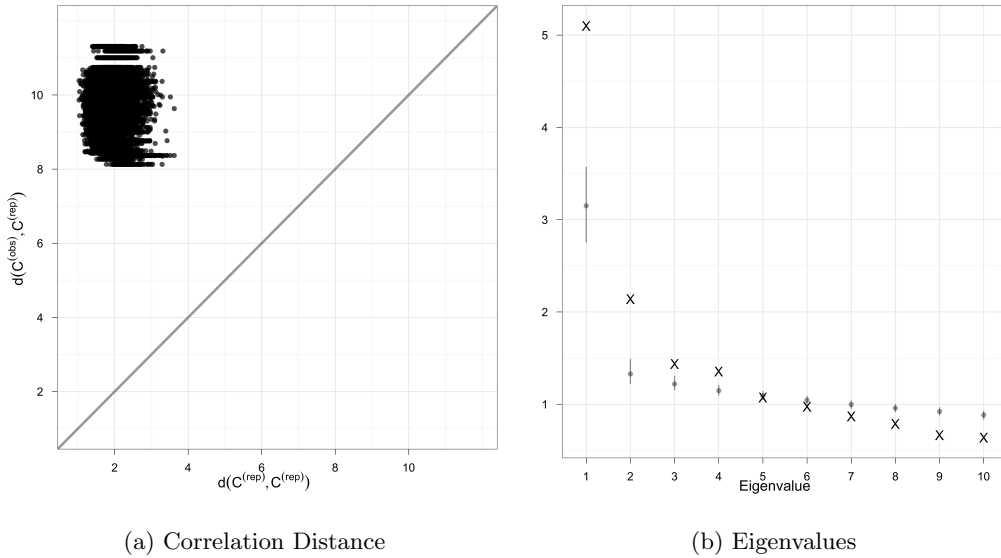


Figure 4.7: The left plot presents a scatterplot of $d_{sld}(C^{obs}, C^{rep,m})$ versus $d_{sld}(C^{rep,m}, C^{rep,m'})$ for all replicated datasets. The grey line represents the 45 degree line. On the right, a plot of the top ten eigenvalues depicting the mean prediction (grey point) and 95% prediction intervals (grey line segment) of the eigenvalues calculated using replicated data. Eigenvalues computed from the observed data are denoted by a red “X”.

Figure 4.8 displays pairwise rank correlations for Mattis Dementia Rating Scale J and

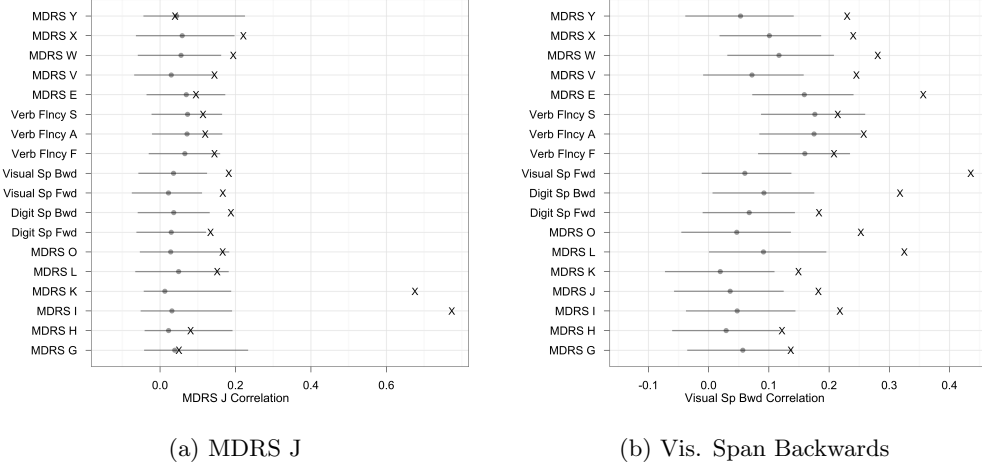


Figure 4.8: Pairwise correlation plots depicting the mean prediction (grey point) and 95% prediction intervals (grey line segment) for Kendall’s τ values calculated using replicated data. Kendall’s τ values computed from the observed data are denoted by a black “X”.

Visual Span Backwards. As one can see, there are some notable discrepancies between the model and observed data. In Figure 4.8(a), the pairwise rank correlations of Mattis Dementia Rating Scale J and the other items are partially well represented with some significant deviations for Mattis Dementia Rating Scale I and K; all three of these items ask participants to repeat alternating movements of some type. In Figure 4.8(b), the model does not fully capture a number of the correlations, with the discrepancy in fitting the correlation between Visual Span Forwards and Visual Span Backwards being the most noticeable. These shortcomings suggests that we may need some means in our model to account for residual correlation among these related items.

4.6 Discussion

In this chapter, we modified traditional item response theory models to accommodate mixed outcomes by adapting and extending generalized linear latent trait models. Although IRT models have traditionally been applied to binary and ordered categorical outcomes, they are increasingly being applied to mixed outcomes. Motivated by a mixed outcome neu-

ropsychological testing battery, we applied the methods proposed by Sammel et al. (1997), Moustaki and Knott (2000), and Dunson (2003) in the context of IRT models. Moreover, we extended these methods to account for outcome types seen in cognitive function testing, such as right-censored count outcomes and duration outcomes. We described Bayesian estimation of the IRT model for mixed outcomes and demonstrated the model on simulated data.

In addition, we extended the model hierarchically to relate the latent trait, or ability parameter in IRT terms, to covariates of interest. This allowed us to study the relationship between indicators of executive functioning and the volume of frontal lobe white matter hyperintensities in the SIVD study. Our analysis found a negative relationship between the latent variable and frontal lobe white matter hyperintensities.

We used the posterior predictive model checking methods presented in Section 2.4 to evaluate the fit of the model to the data. While the model fit some of the marginal distributions fairly well, others proved to be more problematic to fit. Moreover, the use of a single latent variable in the model proved to be insufficient for replicating the rank correlations of the observed SIVD responses. As a result, we may want to consider multiple latent variables in our model to account for residual correlations among the items. Multidimensional item response theory (Reckase, 2009) has been increasingly applied to handle such situations. Reise, Morizot, and Hays (2007) have promoted the use of bifactor models to account for residual correlation. Testlet models (Bradlow, Wainer, and Wang, 1999; Li, Bolt, and Fu, 2006) are another means by which IRT practitioners have sought to account for residual correlation.

Although the generalized linear latent variable framework allows for a variety of distributions to accommodate outcomes of different types, the specification of different distributions conditional on the latent variable requires additional implementation for each distribution type and can be time-consuming in this regard. Moreover, there is no guarantee that we have specified the correct conditional distribution for each outcome. Some initial simulation studies performed by the author have indicated that misspecification of the conditional outcome distributions may affect the accuracy with which the latent variables are estimated but these simulation studies have been small in scope. However, further work is required in

this area before making any definitive proclamations about the impact of misspecification on estimation. Nonetheless, the IRT model for mixed outcomes provides a flexible framework by which we may accommodate a variety of different outcome types encountered in cognitive testing.

Chapter 5

SEMIPARAMETRIC LATENT VARIABLE MODEL FOR MIXED OUTCOMES

5.1 Introduction

In Chapter 4, we discussed a parametric approach towards handling mixed outcomes in IRT models. While we discussed a wide range of outcomes that could be accommodated, we obviously were not exhaustive in the types of outcomes discussed. One could continue to further extend our model to accommodate many more types of outcomes that one may encounter. However, specification of the appropriate conditional distributions for the mixed outcomes is not our primary interest. Specification of a diverse set of distributions $F_1, \dots, F_J$ to model the J mixed outcomes is merely a means to relating the mixed outcomes to covariates.

In this chapter, we develop a semiparametric approach to mixed outcome latent variable models that avoids specification of outcome conditional distributions given the latent variables. Following on the extended rank likelihood approach of Hoff (2007), we start by assuming the existence of continuous latent responses underlying each observed outcome. We then make the assumption that the ordering of the underlying latent responses is consistent with the ordering of the observed outcomes. This approach is similar to that of Shi and Lee (1998) but does not require estimating unknown thresholds. When the data are continuous, our approach is analogous to the use of a rank likelihood (Pettitt, 1982). When the data are discrete, our approach relies on the assumption that the ordering of the latent responses is consistent with the partial ordering of the observed outcomes. Hoff (2007) introduced this general approach for estimating parameters of a semiparametric Gaussian copula model with arbitrary marginal distributions and designated the resulting likelihood as the *extended rank likelihood*.

Pettitt (1982) proposed the rank likelihood for estimation of semiparametric regression

models for (univariate) continuous outcomes. In addition to the semiparametric Gaussian copula model for mixed outcomes, Hoff (2008, 2009) developed semiparametric regression models for continuous or discrete outcomes using the extended rank likelihood. Bickel and Ritov (1997) studied the theoretical properties of the rank likelihood in a regression context and Hoff (2007) and Hoff, Niu, and Wellner (2011) provide theoretical results for the rank likelihood applied to copula estimation. Dobra and Lenkoski (2011) applied the extended rank likelihood method to the estimation of graphical models for multivariate mixed outcomes. The original idea for the following work on the semiparametric latent variable model was presented earlier by Gruhl, Erosheva, and Crane (2010, 2011). Murray, Dunson, Carin, and Lucas (2011) recently proposed a closely related factor analytic model for mixed data.

This chapter is organized as follows. We review the semiparametric copula model and introduce the semiparametric latent variable model in Section 5.2. We discuss Bayesian estimation of this model in Section 5.3 and consider a parameter expansion version of the semiparametric latent variable model. We extend the model hierarchically to analyze the relationship of the primary factor to covariates of interest in Section 5.4. Section 5.5 demonstrates the performance of the standard and parameter expansion models using simulated data. Finally, we use the semiparametric latent variable model to analyze the relationship between executive functioning and frontal white matter hyperintensities in Section 5.6.

5.2 *Semiparametric Latent Variable Model*

5.2.1 *Model Formulation*

As before, let $i = 1, \dots, I$ denote the i th participant, and let $j = 1, \dots, J$ denote the j th outcome. Let y_{ij} denote the observed response of participant i on outcome j with marginal distribution F_j . Moreover, let $F_j^{-1}(u) = \inf\{y : F_j(y) \geq u\}$ be the corresponding pseudo-inverse of F_j , then y_{ij} can be represented as $y_{ij} = F_j^{-1}(u_{ij})$ where u_{ij} is a uniform (0,1) random variable. An equivalent representation is $y_{ij} = F_j^{-1}[\Phi(z_{ij})]$ where $\Phi(\cdot)$ denotes the normal CDF and z_{ij} is distributed standard normal. The unobserved variables z_{ij} are latent responses underlying each observed response y_{ij} . Assuming that the set of correlations

between z_{ij} and $z_{ij'}$, $1 \leq j < j' \leq J$, is specified by the $J \times J$ correlation matrix $\mathbf{C}$, the Gaussian copula model is

$$\mathbf{z}_1, \dots, \mathbf{z}_n | \mathbf{C} \sim \text{i.i.d. } N(\mathbf{0}, \mathbf{C}) \quad (5.1)$$

$$y_{i,j} = F_j^{-1} [\Phi(z_{ij})]. \quad (5.2)$$

Here, $\mathbf{z}_i$ is the J -length vector of latent responses z_{ij} for participant i .

In some analyses, the primary focus is on the estimation of the correlation matrix $\mathbf{C}$ and not the estimation of the marginal distributions $F_1, \dots, F_J$. If the latent responses z_{ij} were known, estimation of $\mathbf{C}$ could proceed using standard methods. Although the latent responses are unknown, Hoff (2007) noted that we do have rank information about the latent responses through the observed responses because $y_{ij} < y_{ij'}$ implies $z_{ij} < z_{ij'}$ as the observed responses are determined by the latent responses through a monotone function, $y_{ij} = F_j^{-1} [\Phi(z_{ij})]$. Denote the full set of latent responses by $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_I)^T$ and the full set of observed responses by $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_I)^T$. Then $\mathbf{Z} \in D(\mathbf{Y})$ where

$$D(\mathbf{Y}) = \{\mathbf{Z} \in \mathbb{R}^{I \times J} : \max_k \{z_{kj} : y_{kj} < y_{ij}\} < z_{ij} < \min_k \{z_{kj} : y_{ij} < y_{kj}\}\}. \quad (5.3)$$

One can then construct a likelihood for $\mathbf{C}$ that does not depend on the specification of the marginal distributions $F_1, \dots, F_J$:

$$\begin{aligned} \Pr(\mathbf{Z} \in D(\mathbf{Y}) | \mathbf{C}, F_1, \dots, F_J) &= \int_{D(\mathbf{Y})} p(\mathbf{Z} | \mathbf{C}) d\mathbf{Z} \\ &= \Pr(\mathbf{Z} \in D(\mathbf{Y}) | \mathbf{C}). \end{aligned} \quad (5.4)$$

Equation (5.3) enables the following decomposition of the density of $\mathbf{Y}$:

$$\begin{aligned} p(\mathbf{Y} | \mathbf{C}, F_1, \dots, F_J) &= p(\mathbf{Y}, \mathbf{Z} \in D(\mathbf{Y}) | \mathbf{C}, F_1, \dots, F_J) \\ &= \Pr(\mathbf{Z} \in D(\mathbf{Y}) | \mathbf{C}, F_1, \dots, F_J) \times p(\mathbf{Y} | \mathbf{Z} \in D(\mathbf{Y}), \mathbf{C}, F_1, \dots, F_J) \\ &= \Pr(\mathbf{Z} \in D(\mathbf{Y}) | \mathbf{C}) \times p(\mathbf{Y} | \mathbf{Z} \in D(\mathbf{Y}), \mathbf{C}, F_1, \dots, F_J). \end{aligned}$$

This decomposition uses the fact that the probability of the event $\mathbf{Z} \in D(\mathbf{Y})$ does not depend on the marginal distributions $F_1, \dots, F_J$ and that the event $\mathbf{Z} \in D(\mathbf{Y})$ occurs whenever $\mathbf{Y}$ is observed. Thus, the dependence structure of $\mathbf{Y}$ can be estimated through $\mathbf{C}$

without any knowledge or assumptions about the marginal distributions. More details on the semiparametric Gaussian copula model can be found in Hoff (2007, 2009).

In the context of latent variable modeling, the main interest is typically not in directly estimating the correlations, $\mathbf{C}$, among observed variables but in characterizing the interdependencies in multivariate observed responses through unobserved variables. Latent variable models place structural constraints on the matrix of correlations among the observed responses and seek a more parsimonious description of the dependence structure. Moreover, as in the case of the IRT models presented in Chapter 4, estimates of the latent variables (conditional on an assumed latent structure) may be of interest to researchers as a means of characterizing individuals using sets of responses. Factor analysis is the most common type of latent variable model with continuous latent variables and continuous outcomes.

To develop a semiparametric approach for factor analysis with mixed outcomes, assume Q factors where Q is generally less than J , let $\boldsymbol{\eta}_i$ be a vector of factor scores for individual i and $\mathbf{H} = (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_I)^T$ be the $I \times Q$ factor matrix. Let Λ denote the $J \times Q$ matrix of factor loadings. We define our semiparametric latent variable model as

$$\boldsymbol{\eta}_i \sim N(\mathbf{0}, \mathbf{I}_Q) \quad (5.5)$$

$$\mathbf{z}_i | \Lambda, \boldsymbol{\eta}_i \sim N(\Lambda \boldsymbol{\eta}_i, \mathbf{I}_J) \quad (5.6)$$

$$\mathbf{z}_1, \dots, \mathbf{z}_n | \Lambda \sim \text{i.i.d. } N(\mathbf{0}, \mathbf{I}_J + \Lambda \Lambda^T) \quad (5.7)$$

$$y_{ij} = g_j(z_{ij}), \quad (5.8)$$

where equation (5.7) is the result of marginalizing out the factors, $\boldsymbol{\eta}_i$, in equation (5.6). Here, we define $g_j(z_{ij}) = F_j^{-1} \left(\Phi \left(z_{ij} / \sqrt{1 + \boldsymbol{\lambda}_j^T \boldsymbol{\lambda}_j} \right) \right)$, where $\boldsymbol{\lambda}_j$ denotes the j -th row of Λ and the marginal distribution F_j remains unspecified. Note that the functions $g_j(\cdot)$ are nondecreasing and preserve the orderings. The model given by equations (5.5)-(5.8) does not rely on the unrestricted correlation matrix $\mathbf{C}$ as does the Gaussian copula model. Assuming that a factor analytic model is appropriate for the data, it constrains the dependencies among the elements of $\mathbf{z}_i$ to be consistent with the functional form of $\mathbf{I}_J + \Lambda \Lambda^T$. As a result, the proposed semiparametric latent variable model is a structured case of the semiparametric Gaussian copula model and can be viewed as a semiparametric form of copula structure analysis (Kl  ppelberg and Kuhn, 2009).

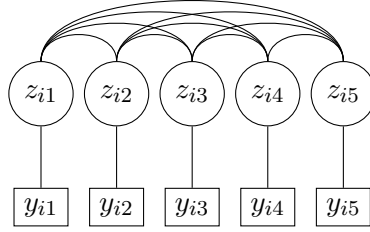


Figure 5.1: Extended Rank Likelihood Applied To Gaussian Copula

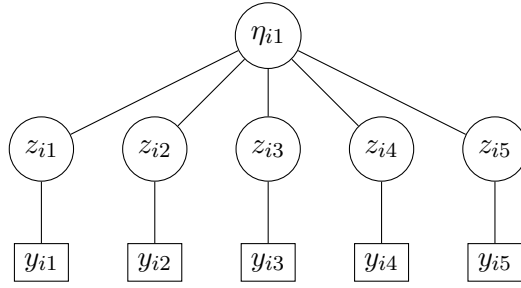


Figure 5.2: Extended Rank Likelihood Applied To Factor Analysis

Figure 5.1 depicts the semiparametric Gaussian copula model. In the unstructured case, we directly estimate the correlations between the latent responses z_{ij} with no restrictions; this general correlation structure is parameterized through $\mathbf{C}$ in the Gaussian copula model. In the structured case of our semiparametric latent variable model, we restrict the correlations of the latent responses z_{ij} after having marginalized out the latent factors to the functional form $\text{diag}(\mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{I}_J)^{-1/2} (\mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{I}_J) \text{diag}(\mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{I}_J)^{-1/2}$. Figure 5.2 depicts the semiparametric latent variable model conditional on a structure with a single latent factor. In this representation, the correlations among the latent responses z_{ij} are induced through their relationship to the single latent factor η_i . The general framework of the semiparametric latent variable model given by equations (5.5)-(5.8) can be used for any special cases of factor analysis.

Rather than limit ourselves strictly to a single factor as we did in the case of the IRT

model for mixed outcomes in Chapter 4, we specify a bifactor latent structure (Holzinger and Swineford, 1937) for the semiparametric latent variable model. We define bifactor models as having a specific structure on the loading matrix, $\mathbf{\Lambda}$, where each observed item loads on the primary factor and may (or may not) load on one or more of the secondary factors (Reise et al., 2007). Most commonly, bifactor models are applied such that an outcome loads on at most one secondary factor. The bifactor model is a useful tool for modeling the neuropsychological battery used in the SIVD study, as it retains a single underlying executive functioning factor while accounting for local dependencies among groups of related items.

5.2.2 Model Identification

The lack of identifiability of factor analysis models that is due to rotational invariance is well known (Anderson, 2003). If we define new factor loadings and new factor scores by $\tilde{\mathbf{\Lambda}} = \mathbf{\Lambda}\mathbf{T}$ and $\tilde{\boldsymbol{\eta}}_i = \mathbf{T}^{-1}\boldsymbol{\eta}_i$, where $\mathbf{T}$ is an orthonormal $Q \times Q$ matrix, then the model

$$\mathbf{z}_i | \mathbf{\Lambda}, \boldsymbol{\eta}_i \sim N(\tilde{\mathbf{\Lambda}}\tilde{\boldsymbol{\eta}}_i, \mathbf{I}_J) \quad (5.9)$$

is indistinguishable from the model in equation (5.6). In the case where the covariance of $\boldsymbol{\eta}_i$ is not restricted to the identity matrix, any nonsingular $Q \times Q$ matrix $\mathbf{T}$ results in the same indeterminacy. In this more general case, we must place Q^2 constraints to prevent this rotational invariance. When we restrict the covariance of $\boldsymbol{\eta}_i$ to the identity matrix, this restriction places $\frac{1}{2}Q(Q+1)$ constraints on the model. We are then left with $\frac{1}{2}Q(Q-1)$ additional constraints to place on the model. We may satisfy this requirement by specifying a bifactor structure with at least $\frac{1}{2}Q(Q-1)$ zeros in the matrix of loadings $\mathbf{\Lambda}$ (Anderson, 2003; Millsap, 2001).

While these restrictions may resolve rotational invariance, the issue of reflection invariance typically remains. Reflection invariance results from the fact that the signs of the loadings in any column in $\mathbf{\Lambda}$ may be switched. Thus, if $\mathbf{D}$ is a diagonal matrix of 1's and -1's precipitating the sign changes, $\mathbf{H}\mathbf{\Lambda}^T = \mathbf{H}\mathbf{D}\mathbf{D}\mathbf{\Lambda}^T = \tilde{\mathbf{H}}\tilde{\mathbf{\Lambda}}^T$. To resolve the issue of reflection invariance, we could specify a prior that places additional constraints on the signs of some of the loadings, e.g., as suggested by Congdon (2003) and Congdon (2006). However, because different choices of constraint placement could potentially have an im-

pact on model fit in complex factor models (Millsap, 2001; Erosheva and Curtis, 2011), we apply the relabeling algorithm proposed by Erosheva and Curtis (2011). This algorithm relies on a decision-theoretic approach and resolves reflection invariance associated with the sign-switching behavior in Bayesian factor analysis in a similar fashion to the relabeling algorithm introduced to address the label-switching problem in mixture models (Stephens, 2000).

In the semiparametric latent variable model, unlike the standard factor analysis model, specific means and variances are not identifiable. Let

$$\tilde{z}_{ij} = \mu_j + \sigma_j z_{ij} \quad (5.10)$$

where μ_j and σ_j are the specific mean and variance for item j . Moreover, if $\tilde{\mathbf{Z}} \in \mathbb{R}^{I \times J}$ denotes the matrix of elements $\tilde{z}_{ij}$ and

$$\tilde{D}(\mathbf{Y}) = \{\tilde{\mathbf{Z}} : \max\{\tilde{z}_{kj} : y_{kj} < y_{ij}\} < \tilde{z}_{ij} < \min\{\tilde{z}_{kj} : y_{ij} < y_{kj}\}\}, \quad (5.11)$$

then

$$\Pr(\tilde{\mathbf{Z}} \in \tilde{D}(\mathbf{Y}) | \mathbf{\Lambda}, \mathbf{H}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \Pr(\mathbf{Z} \in D(\mathbf{Y}) | \mathbf{\Lambda}, \mathbf{H}). \quad (5.12)$$

Thus, shifts in location and scale of the latent responses will not alter the probability of belonging to the set of feasible latent response values implied by orderings of the observed responses. As such, we set the specific means at $\boldsymbol{\mu} = \mathbf{0}$ and the specific variances at $\boldsymbol{\Sigma} = \mathbf{I}_J$.

5.3 Estimation

We take a Bayesian approach to estimate the semiparametric factor analysis model for mixed outcomes given by equations (5.5)-(5.8). We consider three different prior distributions for the factor loadings that are not restricted to zero in the model: a normal distribution, a t-distribution and a scale mixture of normals with a compound gamma mixing density. The normal distribution prior for the unrestricted factor loadings is specified when we employ a Gibbs sampling algorithm with standard semi-conjugate priors for factor analysis (Shi and Lee, 1998; Ghosh and Dunson, 2009). The latter two priors facilitate (and are induced by) a parameter expansion approach to estimation that builds upon the work of Ghosh

and Dunson (2009) on efficient computation for Bayesian factor analysis. In this section, we first employ the standard approach with some modification before proceeding to the development of the parameter expansion approach.

5.3.1 Standard Gibbs Sampling Approach

Due to the restrictions on the outcome specific means and variances discussed in Section 5.2, the loading matrix $\mathbf{\Lambda}$ is the only parameter for which we must specify a prior in the semi-parametric latent variable model. For the loadings, let $\boldsymbol{\lambda}'_j$ denote the K_j non-zero elements of the j -th row of $\mathbf{\Lambda}$. We then specify the following semi-conjugate prior:

$$\boldsymbol{\lambda}'_j \sim N(\mathbf{m}_{\lambda'_j}, \mathbf{S}_{\lambda'_j}). \quad (5.13)$$

Here, $\mathbf{m}_{\lambda'_j}$, a K_j -length vector, and $\mathbf{S}_{\lambda'_j}$, a $K_j \times K_j$ covariance matrix, are hyperparameters for the prior on $\boldsymbol{\lambda}'_j$.

The structural zero elements in the matrix of loadings $\mathbf{\Lambda}$ are specified in accordance with our substantive understanding of the research problem at hand. However, we must have enough zeros so that the model can be identified since we rely on the placement of these structural zeros to resolve rotational invariance. Formally, we specify the prior for these structural zero elements of $\mathbf{\Lambda}$ as

$$\lambda_{jq} \sim \delta_0, \quad (5.14)$$

where δ_0 is a distribution with its point mass concentrated at 0. We estimate loadings with no additional constraints on their signs. We then deal with potential multiple modes of the posterior that are due to reflection invariance of the factor model by applying the relabeling algorithm proposed by Erosheva and Curtis (2011).

We now develop the Gibbs algorithm for sampling factors $\mathbf{H}$ and loadings $\mathbf{\Lambda}$. Because the extended rank likelihood $\Pr(\mathbf{Z} \in D(\mathbf{Y}) | \mathbf{\Lambda}, \mathbf{H})$ involves a complicated integral, any expressions involving the extended rank likelihood will be difficult to compute. We avoid having to compute this integral by drawing from the joint posterior of $(\mathbf{Z}, \mathbf{H}, \mathbf{\Lambda})$ via Gibbs sampling. Given $\mathbf{Z} = \mathbf{z}$ and $\mathbf{Z} \in D(\mathbf{Y})$, the full conditional density of $\mathbf{\Lambda}$ can be written as

$$p(\mathbf{\Lambda} | \mathbf{H}, \mathbf{Z} = \mathbf{z}, \mathbf{Z} \in D(\mathbf{Y})) = p(\mathbf{\Lambda} | \mathbf{H}, \mathbf{Z} = \mathbf{z})$$

because the current draw values $\mathbf{Z} = \mathbf{z}$ imply $\mathbf{Z} \in D(\mathbf{Y})$. A similar simplification may be made with the full conditional density of $\mathbf{H}$. Given $\mathbf{\Lambda}, \mathbf{H}, \mathbf{Z} \in D(\mathbf{Y})$ and $\mathbf{Z}_{(-i)(-j)}$, the full conditional density of z_{ij} is proportional to a normal density with mean $\boldsymbol{\lambda}_j^T \boldsymbol{\eta}_i$ and variance 1 that is restricted to the region specified by $D(\mathbf{Y})$.

To sample from the approximate joint posterior including the latent responses, we proceed as in Hoff (2007) and Hoff (2009).

1. **Draw underlying latent responses $\mathbf{Z}$.** For each i and j , sample z_{ij} from $p(z_{ij}|\mathbf{\Lambda}, \mathbf{H}, \mathbf{Z}_{(-i)(-j)}, \mathbf{Z} \in D(\mathbf{Y}))$. More specifically,

$$z_{ij}^{(m)} \sim \text{TN}_{(z_l, z_u)}(\boldsymbol{\lambda}_j^T \boldsymbol{\eta}_i, 1) \quad (5.15)$$

where TN denotes truncated normal and z_l, z_u define the lower and upper truncation points,

$$z_l = \max\{z_{kj} : y_{kj} < y_{ij}\} \quad (5.16)$$

$$z_u = \min\{z_{kj} : y_{kj} > y_{ij}\}. \quad (5.17)$$

2. **Draw latent factors $\mathbf{H}$.** For each i , we can draw directly from the full conditional distribution for $\boldsymbol{\eta}_i$ as follows,

$$\boldsymbol{\eta}_i \sim \text{N}\left((\mathbf{I}_Q + \mathbf{\Lambda}^T \mathbf{\Lambda})^{-1} \mathbf{\Lambda}^T \mathbf{z}_i, (\mathbf{I}_Q + \mathbf{\Lambda}^T \mathbf{\Lambda})^{-1}\right). \quad (5.18)$$

3. **Draw loadings $\mathbf{\Lambda}$.** For each j , we can draw directly from the full conditional distribution for $\boldsymbol{\lambda}'_j$, the vector of non-zero loadings for outcome j .

$$\boldsymbol{\lambda}'_j \sim \text{N}\left(\left(S_{\lambda'_j}^{-1} + (\mathbf{H}'_j)^T \mathbf{H}'_j\right)^{-1} \left(S_{\lambda'_j}^{-1} m_{\lambda'_j} + (\mathbf{H}'_j)^T \mathbf{z}_j\right), \left(S_{\lambda'_j}^{-1} + (\mathbf{H}'_j)^T \mathbf{H}'_j\right)^{-1}\right), \quad (5.19)$$

where $\mathbf{H}'_j$ is a matrix comprised of the columns of $\mathbf{H}$ for which there are non-zero loadings in $\boldsymbol{\lambda}_j$.

5.3.2 Parameter Expansion Approach

Ghosh and Dunson (2009) observed that the choice of semi-conjugate priors for factor analysis models often results in poorly behaved Gibbs samplers. Specifically, as with hierarchical models, the use of proper but diffuse priors often results in slow mixing due to the high dependence among parameters. Ghosh and Dunson (2009) propose a parameter expansion approach to remedy this problem.

Liu, Rubin, and Wu (1998) proposed a parameter expanded algorithm to accelerate the EM algorithm and Liu and Wu (1999) applied parameter expansion to Gibbs sampling. Among notable applications of the parameter expansion technique, Gelman (2006) applied this approach to propose a set of prior distributions for variance parameters in hierarchical models as alternatives to the commonly adopted inverse gamma prior distribution. Ghosh and Dunson's (2009) application of parameter expansion to factor analysis models induces t or folded- t prior distributions on the factor loadings. Moreover, they demonstrate that the parameter expansion approach leads to greater efficiency in sampling than the standard Gibbs approach for a number of cases. Below, we derive an MCMC estimation algorithm for our semiparametric latent variable model using the parameter expansion approach. We compare the performance of this algorithm against the above proposed Gibbs sampler in a simulation study in Section 5.5.

The central idea behind the parameter expansion approach, using the terminology of Ghosh and Dunson (2009), is to start with a working model that is an overparameterized version of the initial inferential model. After proceeding through MCMC sampling, a transformation can be used to relate the draws from the working model to draws from the inferential model. For our application, the overparameterized model is formulated in terms of the latent responses,

$$\mathbf{z}_i^* \sim N(\mathbf{\Lambda}^* \boldsymbol{\eta}_i^*, \boldsymbol{\Sigma}), \quad (5.20)$$

$$\boldsymbol{\eta}_i^* \sim N(\mathbf{0}, \boldsymbol{\Psi}), \quad (5.21)$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ are diagonal matrices that are no longer restricted to the identity matrix. The latent responses $\mathbf{z}_i^*$, the latent variables $\boldsymbol{\eta}_i^*$ and the loadings $\mathbf{\Lambda}^*$ are unidentified in this

working model. The transformations from the working model to the inferential model are then specified as

$$\begin{aligned}\boldsymbol{\eta}_i &= \boldsymbol{\Psi}^{-1/2} \boldsymbol{\eta}_i^*, \\ \mathbf{z}_i &= \boldsymbol{\Sigma}^{-1/2} \mathbf{z}_i^*, \\ \boldsymbol{\Lambda} &= \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Lambda}^* \boldsymbol{\Psi}^{1/2}.\end{aligned}\tag{5.22}$$

To sample from the working model, we must specify priors for the diagonal elements of $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ as well as for $\boldsymbol{\Lambda}^*$. We specify these priors in terms of the precisions ψ_q^{-2} and σ_j^{-2} . In addition, we denote by $\boldsymbol{\lambda}_j^{*'}$ the non-zero elements of the j -th row of $\boldsymbol{\Lambda}^*$. The prior on $\boldsymbol{\lambda}_j'$ is then induced through the priors on ψ_q^{-2} , σ_j^{-2} and $\boldsymbol{\lambda}_j^{*'}$, rather than being specified directly. Our priors are

$$\begin{aligned}\psi_q^{-2} &\sim \text{Gamma}(\nu_\psi, \phi_\psi), \\ \sigma_j^{-2} &\sim \text{Gamma}(\nu_\sigma, \phi_\sigma), \\ \boldsymbol{\lambda}_j^{*'} &\sim \text{N}(\mathbf{m}_{\lambda_j^{*'}}, \mathbf{S}_{\lambda_j^{*'}}).\end{aligned}\tag{5.23}$$

We also consider the slightly simpler case where $\boldsymbol{\Sigma}$ is restricted to the identity matrix but the (diagonal) elements of $\boldsymbol{\Psi}$ are unrestricted in the working model.

Our Gibbs sampling procedure for the working model proceeds according to the following steps.

1. **Draw latent responses $\mathbf{Z}^*$.** For each i and j , sample z_{ij}^* from a truncated normal distribution according to

$$z_{ij}^* \sim \text{TN}_{(z_l^*, z_u^*)} \left((\boldsymbol{\lambda}_j^*)^T \boldsymbol{\eta}_i^*, \sigma_j^2 \right),\tag{5.24}$$

where as before z_l^*, z_u^* define the lower and upper truncation points,

$$z_l^* = \max\{z_{kj}^* : y_{kj} < y_{ij}\}\tag{5.25}$$

$$z_u^* = \min\{z_{kj}^* : y_{kj} > y_{ij}\}.\tag{5.26}$$

2. **Draw latent variables $\mathbf{H}^*$.** For each i , we can draw directly from the full conditional distribution for η_i^* as follows,

$$\eta_i^* \sim \text{N} \left(\left(\Psi^{-1} + (\Lambda^*)^T \Sigma^{-1} \Lambda^* \right)^{-1} (\Lambda^*)^T \Sigma^{-1} \mathbf{z}_i^*, \left(\Psi^{-1} + (\Lambda^*)^T \Sigma^{-1} \Lambda^* \right)^{-1} \right) \quad (5.27)$$

3. **Draw loadings Λ^* .** For each j , we can draw directly from the full conditional distribution for $\lambda_j^{*'}$, the vector of non-zero loadings for outcome j .

$$\lambda_j^{*'} \sim \text{N} \left(\left(S_{\lambda_j'}^{-1} + \sigma_j^{-2} (\mathbf{H}_j^{*'})^T \mathbf{H}_j^{*'} \right)^{-1} \left(S_{\lambda_j'}^{-1} m_{\lambda_j'} + \sigma_j^{-2} (\mathbf{H}_j^{*'})^T \mathbf{z}_j^* \right), \left(S_{\lambda_j'}^{-1} + \sigma_j^{-2} (\mathbf{H}_j^{*'})^T \mathbf{H}_j^{*'} \right)^{-1} \right). \quad (5.28)$$

4. **Draw inverse covariance matrix Ψ^{-1} .** For each q , we draw the diagonal element ψ_q^{-2} of Ψ^{-1} from the full conditional distribution as follows,

$$\psi_q^{-2} \sim \text{Gamma} \left(\phi_\psi + I/2, \nu_\psi + \frac{1}{2} \sum_i \eta_{iq}^2 \right). \quad (5.29)$$

5. **Draw inverse covariance matrix Σ^{-1} .** For each j , we draw the diagonal element σ_j^{-2} of Σ^{-1} from the full conditional distribution according to

$$\sigma_j^{-2} \sim \text{Gamma} \left(\phi_\sigma + I/2, \nu_\sigma + \frac{1}{2} (\mathbf{z}_j - \mathbf{H} \lambda_j)^T (\mathbf{z}_j - \mathbf{H} \lambda_j) \right). \quad (5.30)$$

After proceeding through steps 1-5 for a number of iterations, we discard a number of initial draws as burn-in and transform the remaining draws using equations (5.22) as part of a postprocessing step. Thus, we have posterior draws from our inferential model. The only remaining steps are to apply the relabeling algorithm of Erosheva and Curtis (2011), assess convergence and calculate posterior summaries for the parameters in the inferential model. In the slightly simpler case, where Σ is restricted to the identity matrix but the (diagonal) elements of Ψ are unrestricted in the working model, we skip step 5, the Gibbs sampling step for Σ , in the above algorithm.

We noted earlier that Ghosh and Dunson's (2009) application of parameter expansion to factor analysis models induces t or folded- t prior distributions on the factor loadings. If

the prior covariance matrix on λ_j' is diagonal, the prior induced on λ_{jk_j}' by the parameter expansion is the product of the normal distribution prior on $\lambda_{jk_j}^{*t}$ and the square root of a ratio of gamma distribution priors on σ_j^{-2} and $\psi_{k_j}^{-2}$. The ratio of gamma distributed random variables has a compound gamma distribution which is a form of the generalized beta prime distribution with the shape parameter fixed to 1. If we integrate out this ratio, the prior for λ_{jk_j} is a scale mixture of normals (West, 1987) with a compound gamma mixing density. Interestingly, this prior or related priors arise in research on shrinkage priors for high-dimensional regression problems (see Armagan, Dunson, and Clyde, 2011 and Polson and Scott, 2010). If the covariance of the latent responses, Σ , is fixed to the identity matrix, then the priors induced on the factor loadings are t -distributions as in Ghosh and Dunson (2009). Thus, we note that the developed parameter expansion approach results in different priors for the factor loadings Λ than the standard priors presented in Section 5.3.1. Derivations of the induced priors are presented in Appendix A.

In this chapter, we discussed different approaches to Bayesian estimation of the semi-parametric latent variable model specified by equations (5.5)-(5.8). We considered three different priors for the unrestricted factor loadings and these priors were associated with different estimation approaches. We summarize these priors and the corresponding estimation approaches in Table 5.1.

Table 5.1: Summary of prior distributions and estimation approaches.

Prior Distribution for λ_{jq}	Estimation Approach	Parameter Expansion Additional Parameters
Normal	Standard Gibbs	None
t	Parameter Expansion	Ψ
Scale Mixture of Normals w/ Compound Gamma Mixing Density	Parameter Expansion	Ψ, Σ

5.3.3 Generating Replicated Data For Posterior Predictive Model Checks

To employ the posterior predictive model checking methods discussed in Section 2.4, we now describe two methods for generating replicated data from the posterior predictive distribution for the semiparametric latent variable model. The more straightforward of the two approaches generates I vectors of predicted latent responses according to

$$\mathbf{z}_i^{(\text{rep}, m)} \sim \text{N} \left(\mathbf{0}, \mathbf{I}_J + \left(\boldsymbol{\Lambda}^{(m)} \right)^T \boldsymbol{\Lambda}^{(m)} \right) \quad (5.31)$$

where superscript (rep, m) denotes the m -th replicated draw and superscript (m) denotes the m -th posterior draw of the parameter. To generate predicted responses on the original data scale rather than the latent scale, we rely on the inverse empirical marginal CDF for each outcome, $\hat{F}_j^{-1}(\cdot)$, to transform the latent responses. That is,

$$y_{ij}^{(\text{rep}, m)} = \hat{F}_j^{-1} \left(\Phi \left(z_{ij}^{(\text{rep}, m)} / \sqrt{1 + \left(\boldsymbol{\lambda}_j^{(m)} \right)^T \boldsymbol{\lambda}_j^{(m)}} \right) \right). \quad (5.32)$$

Hoff (2007) notes that such an approach ignores the uncertainty in the estimation of F_j for the prediction of $\mathbf{y}_{ij}^{(\text{rep}, m)}$. Nonetheless, this method provides a posterior predictive distribution for which the univariate marginal distributions match those observed in the data. One might consider this approach a conditional posterior predictive distribution as it conditions on the empirical marginal distributions.

Hoff (2007) describes another procedure for calculating posterior predictive distributions that does incorporate uncertainty in estimation of the univariate marginal distributions for predicting $\mathbf{y}_{ij}^{(\text{rep}, m)}$. We generate a new vector of latent responses, $z_{I+1}^{(\text{rep}, m)}$, in addition to I sets drawn as part of the Gibbs sampling algorithm. If $z_{(I+1)j}$ falls between two latent responses, z_{ij} and $z_{i'j}$, that share the same value on the original data scale (i.e., $y_{ij} = y_{i'j}$) then $y_{(I+1)j}$ must also take this value as $g_j(\cdot)$ is monotonic. If $z_{(I+1)j}$ falls between two latent responses, z_{ij} and $z_{i'j}$, that do not share the same value on the original data scale, then we use linear interpolation to obtain a value for $y_{(I+1)j}$. Hoff (2007) points out that the two approaches are largely equivalent for large I . In the rest of the paper, we rely on the second approach to generate replicated data for our evaluations of model fit.

5.4 Hierarchical Formulation and Estimation

To relate covariates of interest to the primary factor in the semiparametric latent variable model and investigate their general association with the outcomes, we extend the model hierarchically. In Chapter 4, we extended the IRT model for mixed outcomes hierarchically to include covariates by specifying the mean of the latent variable as a function of covariates. In doing so, we made the assumption that the model follows structure of the multiple indicator multiple cause (MIMIC) model (Goldberger and Hauser, 1971; Jöreskog and Goldberger, 1975) and the multilevel IRT model described by Fox and Glas (2001) where covariates are related to outcomes through the latent variable or ability parameter only, rather than also relating covariates directly to outcomes as in other formulations (Dunson, 2003). We rely on the same assumption here. For implications of this assumption, please refer back to Chapter 4.

As before, we assume that

$$\boldsymbol{\eta}_i \sim N(\mathbf{m}_{\eta_i}, \boldsymbol{\Psi}), \quad (5.33)$$

where $\mathbf{m}_{\eta_i} = \mathbf{0}$, $\boldsymbol{\Psi} = \mathbf{I}_Q$. However, we replace the first element of $\mathbf{m}_{\eta_i}$ with a function of the covariates denoted by the P -length vector $\mathbf{x}_i$:

$$\mathbf{m}_{\eta_i} = (\mathbf{x}_i^T \boldsymbol{\beta}, 0, \dots, 0)^T. \quad (5.34)$$

The complete hierarchical semiparametric latent variable model is therefore

$$y_{ij} = g_j(z_{ij}) \quad (5.35)$$

$$\mathbf{z}_i | \Lambda, \boldsymbol{\eta}_i \sim N(\Lambda \boldsymbol{\eta}_i, \mathbf{I}_J) \quad (5.36)$$

$$\boldsymbol{\eta}_i \sim N(\mathbf{m}_{\eta_i}, \mathbf{I}_Q) \quad (5.37)$$

$$\mathbf{m}_{\eta_i} = (\mathbf{x}_i^T \boldsymbol{\beta}, 0, \dots, 0)^T. \quad (5.38)$$

When we employ a parameter expansion approach for estimation of the model in equations (5.35)-(5.38),

$$\boldsymbol{\eta}_i^* \sim N(\mathbf{m}_{\eta_i}, \boldsymbol{\Psi}), \quad (5.39)$$

$$\mathbf{m}_{\eta_i} = (x_i^T \boldsymbol{\beta}^*, 0, \dots, 0)^T, \quad (5.40)$$

and the diagonal elements of Ψ are no longer restricted during MCMC. For β^* , we specify the semi-conjugate prior:

$$\beta^* \sim N(\mathbf{m}_\beta, \mathbf{S}_\beta). \quad (5.41)$$

Moreover, to further facilitate efficient computation, we add an additional working parameter, α , as suggested by Ghosh and Dunson (2009), so that

$$\mathbf{m}_{\eta_i} = (\alpha + \mathbf{x}_i^T \beta^*, 0, \dots, 0)^T. \quad (5.42)$$

Relaxing the restriction on the mean of the latent variable promotes better mixing of the regression coefficients. For α , we use the semi-conjugate prior:

$$\alpha \sim N(m_\alpha, s_\alpha^2). \quad (5.43)$$

To estimate the hierarchical model (equations (5.35)-(5.38)), we modify the steps for drawing η_i^* and Ψ in the sampling algorithm from Section 5.3 to account for the inclusion of covariates and the additional working parameter, α , in $\mathbf{m}_{\eta_i}$. We sample β^* and α according to their full conditionals:

$$\beta^* \sim N\left(\left(\psi_1^{-2} \mathbf{X}^T \mathbf{X} + \mathbf{S}_\beta^{-1}\right)^{-1} \left(\psi_1^{-2} \mathbf{X}^T (\eta_{q=1}^* - \mathbf{1}_I \alpha) + \mathbf{S}_\beta^{-1} \mathbf{m}_\beta\right), \left(\psi_1^{-2} \mathbf{X}^T \mathbf{X} + \mathbf{S}_\beta^{-1}\right)^{-1}\right), \quad (5.44)$$

$$\alpha \sim N\left(\left(\psi_1^{-2} I + s_\alpha^{-2}\right)^{-1} \left(\psi_1^{-2} \mathbf{1}_I^T (\eta_{q=1}^* - \mathbf{X} \beta^*) + s_\alpha^{-2} m_\alpha\right), \left(\psi_1^{-2} I + s_\alpha^{-2}\right)^{-1}\right). \quad (5.45)$$

where $\mathbf{X}$ is an $I \times P$ matrix of covariates and $\eta_{q=1}^*$ is a vector of the primary factor scores, the first column of $\mathbf{H}^*$. In the post-processing stage for the parameter expansion approach, we make the transformations:

$$\eta_i = \Psi^{-1/2} (\eta_i^* - \alpha), \quad (5.46)$$

$$\beta = \beta^* \psi_1^{-1}, \quad (5.47)$$

where $\alpha = (\alpha, 0, \dots, 0)^T$. In the parameter expansion approach where the diagonal of Ψ is unrestricted, given diagonal S_β , the induced prior on β_p is a scale mixture of normals with a gamma mixing density. This normal-gamma prior was proposed by Griffin and Brown (2010) as a prior for regression coefficients that generalizes the Bayesian Lasso (Park and Casella, 2008).

5.5 A Simulated Data Example

To test the semiparametric latent variable model, we generated simulated data and then examined the model's ability to recover the data generating parameters. We compared three approaches to estimation: the standard Gibbs sampling approach (denoted henceforth as Standard), the parameter expansion approach where the diagonal elements of both Σ and Ψ are unrestricted during estimation (denoted as PX1), and the parameter expansion approach where the diagonal of only Ψ is unrestricted during estimation (denoted as PX2). Recall that the parameter expansion approaches induce different priors for the factor loadings, Λ .

We generated data for $I = 500$ hypothetical individuals on $J = 15$ different outcomes. The data generating process employed a bifactor model with two secondary factors in addition to the primary factor. The loading matrix used to simulate the data is

$$\Lambda = \begin{pmatrix} 0.613 & 2.900 & 0.000 \\ 1.667 & 0.000 & 0.000 \\ 1.498 & 0.000 & 1.325 \\ 2.084 & 0.000 & -1.037 \\ 0.030 & 0.000 & 0.000 \\ 1.268 & 0.000 & 0.994 \\ -1.271 & 0.000 & 0.000 \\ -2.100 & 0.000 & 2.356 \\ 1.131 & 0.000 & 0.000 \\ 1.470 & 0.000 & 0.000 \\ -0.335 & 0.000 & 0.000 \\ 2.235 & 0.000 & 0.000 \\ -1.267 & 2.972 & 0.000 \\ 2.499 & -0.420 & 0.000 \\ 1.937 & -0.020 & 0.000 \end{pmatrix}. \quad (5.48)$$

Here, all outcomes load on the general factor; outcomes 1, 13, 14 and 15 load on one secondary factor; and outcomes 3, 4, 6 and 8 load on the other secondary factor. For each individual, we simulated $\boldsymbol{\eta}_i \sim N(\mathbf{0}, \mathbf{I}_Q)$. We subsequently generated a matrix of latent

responses, $\mathbf{Z}$, with mean $\mathbf{H}^T \mathbf{\Lambda}$ and variance 1. Finally, we randomly drew cutoffs for each outcome in order to produce discretized “observed” responses from the continuous latent responses so that the number of unique values for each outcome ranged from 2 (outcome 1) to 30 (outcome 9).

To fit the model to the simulated data, we employed each estimation approach to generate 50,000 MCMC draws, the first 10,000 of which we discarded as burn-in. In addition, we thinned the posterior samples, keeping only every 10th draw. Thus, we used 4,000 draws from each approximate posterior distribution to make our inferences.

Table 5.2 displays a summary of the posterior estimates of the loadings. For each estimation approach, all 95% posterior credible intervals for the loadings contained the data generating parameter. Similarly, the estimation approaches do a good job of recovering the true factor values, $\mathbf{H}$, with roughly 96% of the 95% posterior credible intervals containing the data generating values in each of the three estimation approaches (detailed results not shown).

Table 5.3 compares the efficiency of each estimation method by examining summary statistics of the effective sample sizes¹ of the posterior samples for the $\mathbf{\Lambda}$ parameters. Considering both the mean value and the different quantiles, the PX1 approach outperformed the others. Examining traceplots for $\lambda_{14,1}$ and $\lambda_{4,3}$ in Figure 5.3, the chains from the PX1 approach displayed better mixing and less autocorrelation. The trace plots for $\lambda_{14,1}$ are typical of the trace plots for other non-zero elements of $\mathbf{\Lambda}$ while the trace plots for $\lambda_{4,3}$ are typical of cases where there is extreme disparity between the effective sample sizes of the MCMC draws using each approach. In examining the trace plots for other non-zero elements of $\mathbf{\Lambda}$ (not shown), the PX1 approach tended to outperform the other estimation methods in similar fashion. Due to greater efficiency, we employed PX1 estimation for the rest of the paper.

Although the model given by equations (5.5)-(5.8) is identified provided that we have enough fixed zeros in $\mathbf{\Lambda}$, for evaluating convergence we have found it useful to additionally

¹Effective sample sizes were calculated using the `coda` package in R.

Table 5.2: Posterior means of the factor loadings $\mathbf{\Lambda}$ from the three different approaches to semiparametric latent variable model estimation.

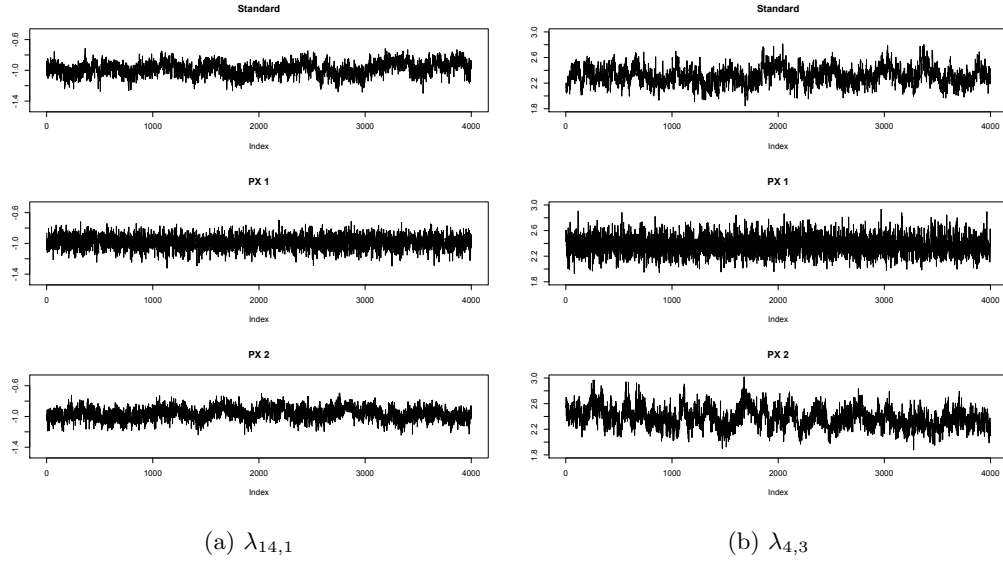
Loading	Truth	Standard Post. Mean	PX1 Post. Mean	PX2 Post. Mean
$\lambda_{1,1}$	0.61	0.67	0.70	0.73
$\lambda_{2,1}$	1.67	1.78	1.83	1.83
$\lambda_{3,1}$	1.50	1.52	1.55	1.54
$\lambda_{4,1}$	2.08	2.05	2.09	2.06
$\lambda_{5,1}$	0.03	0.01	0.01	0.01
$\lambda_{6,1}$	1.27	1.42	1.46	1.51
$\lambda_{7,1}$	-1.27	-1.41	-1.45	-1.45
$\lambda_{8,1}$	-2.10	-1.95	-2.14	-2.22
$\lambda_{9,1}$	1.13	1.16	1.18	1.19
$\lambda_{10,1}$	1.47	1.39	1.43	1.43
$\lambda_{11,1}$	-0.34	-0.33	-0.34	-0.34
$\lambda_{12,1}$	2.23	2.06	2.13	2.12
$\lambda_{13,1}$	-1.27	-1.14	-1.29	-1.19
$\lambda_{14,1}$	2.50	2.31	2.37	2.38
$\lambda_{15,1}$	1.94	1.78	1.82	1.82
$\lambda_{1,2}$	2.90	2.75	2.91	3.03
$\lambda_{13,2}$	2.97	2.87	3.19	2.92
$\lambda_{14,2}$	-0.42	-0.50	-0.50	-0.51
$\lambda_{15,2}$	-0.02	0.04	0.04	0.04
$\lambda_{3,3}$	1.32	1.28	1.28	1.27
$\lambda_{4,3}$	-1.04	-0.98	-0.98	-0.97
$\lambda_{6,3}$	0.99	1.14	1.14	1.19
$\lambda_{8,3}$	2.36	2.16	2.35	2.46

consider the scaled loadings, $\check{\mathbf{\Lambda}}$ where

$$\check{\lambda}_{jq} = \frac{\lambda_{jq}}{\sqrt{1 + \boldsymbol{\lambda}_j^T \boldsymbol{\lambda}_j}}, \quad (5.49)$$

Table 5.3: Summary of effective sample sizes for Λ parameters by estimation approach.

Estimation Approach	Mean	Quantiles				
		0%	25%	50%	75%	100%
Standard	605	11	101	149	376	4231
PX1	2611	329	1622	3169	3818	4000
PX2	540	10	63	106	402	4000

Figure 5.3: Trace plots for $\lambda_{14,1}$ and $\lambda_{4,3}$ for the three different approaches.

so that

$$C = \text{Cor}(\mathbf{Z}) \quad (5.50)$$

$$= \text{diag}(\Lambda\Lambda^T + \mathbf{I}_J)^{-1/2} (\Lambda\Lambda^T + \mathbf{I}_J) \text{diag}(\Lambda\Lambda^T + \mathbf{I}_J)^{-1/2} \quad (5.51)$$

$$= \check{\Lambda}\check{\Lambda}^T + \check{\Sigma}. \quad (5.52)$$

Here, $\check{\Sigma}$ is a $J \times J$ diagonal matrix with diagonal entry $\sigma_j^2 = (1 + \lambda_j^T \lambda_j)^{-1}$. In our experi-

ence, trace plots of the scaled loadings often appear more stable than those of the unscaled loadings. Likewise, we observed that separate chains for estimation of $\mathbf{\Lambda}$ that have yet to converge to similar values often appear to have converged when rescaled, providing an indication that the separate chains are estimating similar dependence structures. These scaled loadings may also be useful for inferential purposes as they place the latent responses on the same scale marginally with $\text{Var}(z_{ij}) = 1$ for all outcomes $j = 1, \dots, J$.

Posterior predictive model checks further suggest that the model fits the simulated data well. In the following, we present posterior predictive checks using the PX1 model replicated data. Figure 5.4 demonstrates that the model appears to successfully replicate the marginal distributions of the outcomes 6 and 14. We observed similarly good results for other outcomes.

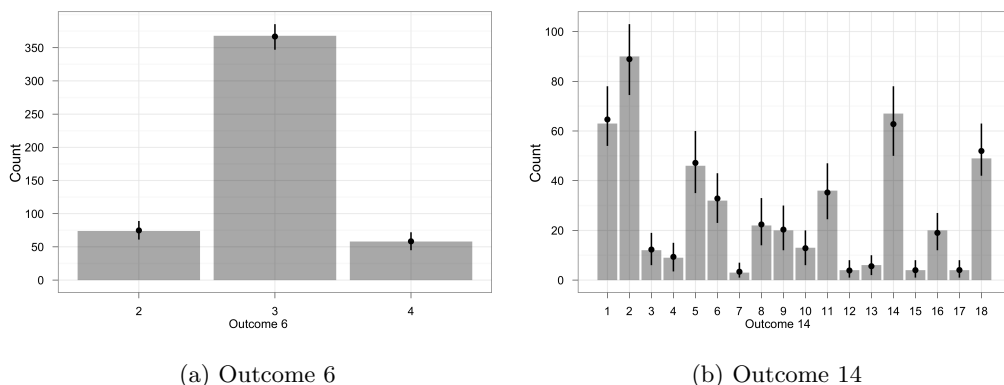


Figure 5.4: Histograms of the observed outcome scores in the simulated dataset. The black dots indicate the mean count across replicated datasets for each grade. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.

Figure 5.5 presents the pairwise rank correlations between outcome 14 and the other outcomes. The plot displays the observed and mean predicted rank correlations together with their 95% posterior predictive intervals. For each of the 14 pairwise rank correlations in the plot, the posterior credible interval covers the observed value, suggesting an adequate model fit for this outcome. Though not shown, similar results are obtained for the other

outcomes.

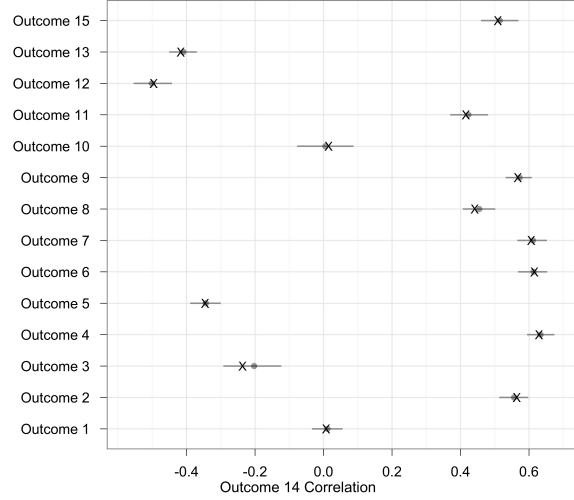


Figure 5.5: The mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) for outcome 5 Kendall's τ values calculated using replicated data. Kendall's τ values computed from the observed data are denoted by a black "X".

Figure 5.6 presents correlation distance and eigenvalue posterior predictive model check plots. Both figures detect no discrepancy between the model-predicted and observed rank correlation matrices.

In the context of simulated data, we demonstrated that the 3 different estimation approaches can ably recover data-generating values. However, the parameter expansion approach, PX1, where both Σ and Ψ are unrestricted in the working model during the MCMC sampling, outperformed the other approaches in that it displayed better mixing properties during MCMC estimation. In our simulation example, the posterior predictive model checks indicated that marginal distributions for observed outcomes, correlations among observed outcomes, and eigenvalues of the observed correlation matrix were replicated well by the model.

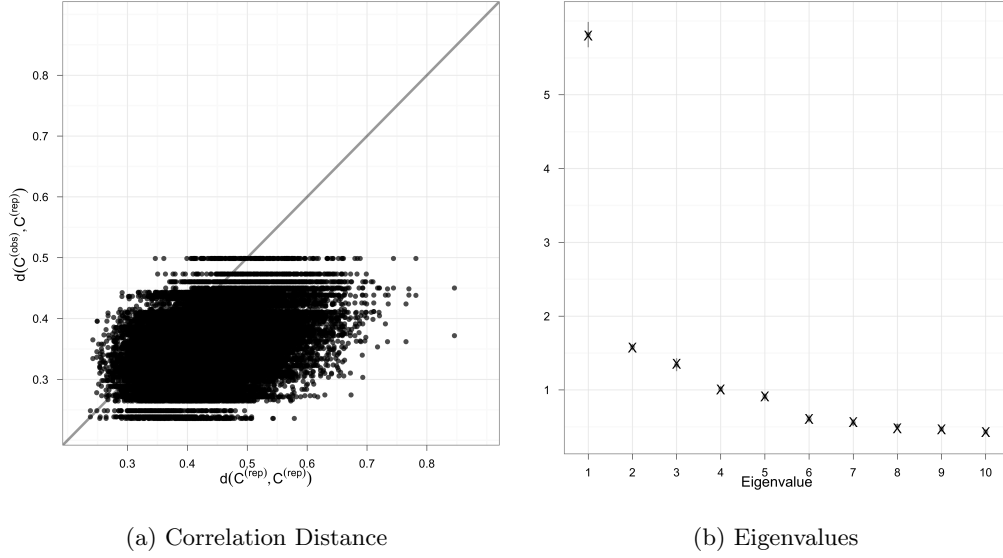


Figure 5.6: The left plot presents scatterplots of $d_{sld}(C^{obs}, C^{rep,m})$ versus $d_{sld}(C^{rep,m}, C^{rep,m'})$ for all replicated datasets. The grey line represents the 45 degree line. The right plot displays the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) of the largest ten eigenvalues calculated using replicated data. Eigenvalues computed from the observed data are denoted by a black “X”.

5.6 Application to the SIVD Study

As in section 4.5, we are interested in relating the individual’s level of executive functioning to the white matter hyperintensity volume located in the frontal lobe at the first study visit. We present again Table 5.4 which displays basic information for the 19 outcomes as well as some summary statistics observed in the data for $I = 341$ participants. For many of the binary outcomes as well as the Mattis Dementia Rating Scale outcomes E and V, the mean and median scores are very close to the largest possible score.

To illustrate the challenges of modeling cognitive outcomes from the SIVD study parametrically, recall the cases of Mattis Dementia Rating Scale E and W. For Mattis Dementia

Table 5.4: Summary statistics for $I = 341$ responses to 19 SIVD executive functioning outcomes as well as outcome type assignment. ‘RC Count’ denotes a right-censored count outcome.

	Range	Mean	Median	Outcome Type
Digit Span Forward	3-12	7.69	8	Count
Digit Span Backwards	1-12	5.97	6	Count
Visual Span Forward	0-13	7.15	7	Count
Visual Span Backwards	0-12	6.18	6	Count
Verbal Fluency Letter F	1-26	11.8	12	Count
Verbal Fluency Letter A	0-40	10.2	10	Count
Verbal Fluency Letter S	0-50	12.4	12	Count
MDRS E	2-20	16.64	19	RC Count
MDRS G	0-1	0.96	1	Binary
MDRS H	0-1	0.98	1	Binary
MDRS I	0-1	0.95	1	Binary
MDRS J	0-1	0.97	1	Binary
MDRS K	0-1	0.98	1	Binary
MDRS L	0-1	0.79	1	Binary
MDRS O	0-1	0.94	1	Binary
MDRS V	9-16	14.9	16	RC Count
MDRS W	0-8	6.44	7	Ordered Cat.
MDRS X	0-3	2.66	3	Ordered Cat.
MDRS Y	0-3	2.93	3	Ordered Cat.

Rating Scale outcome E, participants are given one minute and are asked to name as many items found in supermarkets as they can. The participant’s score is the number of valid items named, censored at 20. A histogram of observed scores for this outcome in Fig-

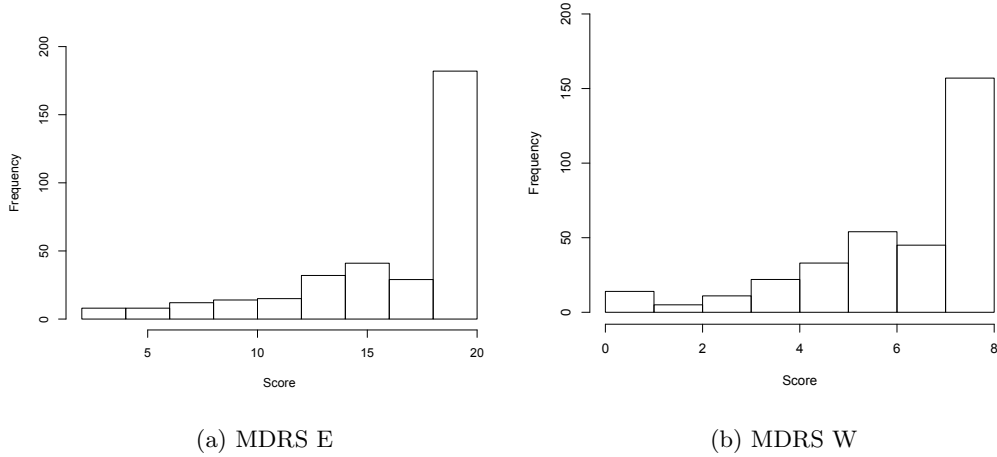


Figure 5.7: Histograms of scores for MDRS E and W items.

Figure 5.7(a) shows some evidence of a ceiling effect for this item. Similarly, Figure 5.7(b) depicts a histogram of observed scores for Mattis Dementia Rating Scale W that asks a participant to compare words and identify similarities. Although the description in this case does not suggest right-censoring, there is also some evidence of a ceiling effect in the histogram. We might treat Mattis Dementia Rating Scale outcome W as right-censored rather than an ordered categorical outcome in a parametric approach. These are just two examples that illustrate the uncertainty and challenge in specifying appropriate parametric distributions for each cognitive outcome in the SIVD study. To bypass this specification, yet still model the interdependencies among test items, we used the hierarchical semiparametric latent variable model.

We are interested in modeling the relationship between the primary factor and the volume of white matter hyperintensities located in the frontal lobe of the brain. Controlling for other covariates, we specified the mean of the primary factor for participant i as

$$E[\eta_{i1}] = \beta_1 \text{Sex}_i + \beta_2 \text{Educ}_i + \beta_3 \text{Age}_i + \beta_4 \text{Vol}_i + \beta_5 \text{WMH}_i, \quad (5.53)$$

where Sex_i is the participant's sex (Female=1, Male=0), Educ_i is the number of years of

education, Vol_i is the total brain volume of the participant, and WMH_i is the frontal white matter hyperintensity volume. We used standardized versions of the continuous predictor variables.

One-factor semiparametric model We started our analysis by examining the $Q = 1$ model with a single latent factor explaining interdependencies among the test items. To estimate the model, we utilized the parameter-expanded Gibbs sampling algorithm. Even though we found this approach to be more efficient than the standard Gibbs sampler, we still observed high autocorrelation within the chains for factor loadings. We drew 50,000 MCMC samples and discarded the first half as burn-in. We used trace plots and the Geweke (Geweke, 1992) and Raftery-Lewis (Raftery and Lewis, 1995) diagnostic tests to assess convergence.

Table 5.6 displays posterior summaries for the regression coefficients, β . We observed a negative relationship between the primary factor and frontal white matter hyperintensity volume. The accompanying 95% posterior credible interval (-0.466, -0.205) did not contain zero, suggesting a negative association between frontal white matter hyperintensity volume and the primary factor. That is, participants with a greater volume of white matter hyperintensities accumulated in the frontal lobe tended to perform worse on indicators of executive functioning.

We evaluated model fit using posterior predictive model checks. We began by examining the fit of the marginal distributions. Figure 5.8 displays the histograms of observed responses for Verbal Fluency A and Mattis Dementia Rating Scale E along with posterior predictive summaries. In each case, the model appeared to do a satisfactory job of approximating the data. We found similarly good approximations of the marginal distributions in the observed data for the other outcomes as well.

We assessed the model's ability to replicate the observed dependence structure in the data at a global level by examining the correlation distance and the eigenvalues posterior predictive model checking plots (Figure 5.9). The cloud of correlation matrix distances in Figure 5.9(a) sits in the upper left of the plot, well removed from the 45 degree line. In Figure 5.9(c), the first eigenvalue was well approximated by the model but the subsequent

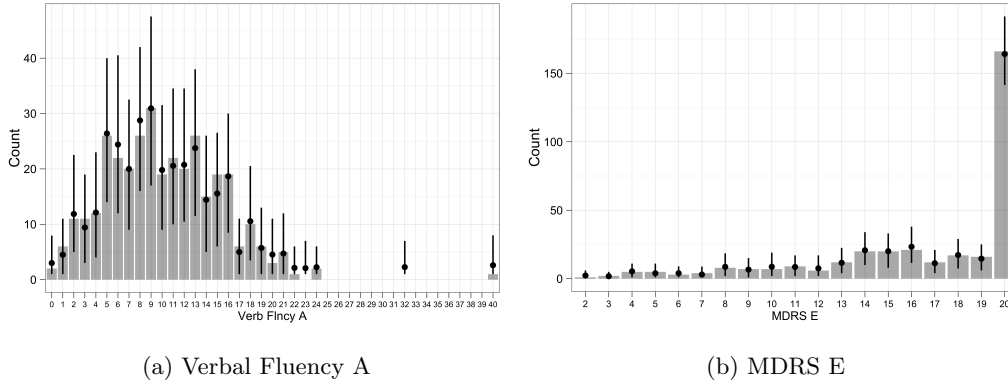


Figure 5.8: Histograms of the observed scores for the Verbal Fluency A and MDRS E. The black points indicate the mean count across replicated datasets for each score. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.

eigenvalues indicated model misfit, suggesting that additional factors may be necessary to more accurately represent the dependence structure in the data.

We reviewed the pairwise rank correlations to better understand the shortcomings of the single factor model and direct the next steps in our model building process. Figures 5.10(a) and 5.10(b) display the pairwise correlation plots for the Mattis Dementia Rating Scale J and Visual Span Backwards outcomes for the single factor model. In both cases, the model fit the majority of the pairwise correlations well. However, in each case, there were a few outcomes with poorly fitted correlations. For Mattis Dementia Rating Scale J, the model did not appear to fully capture the correlation with the conceptually-related Mattis Dementia Rating Scale I and K; all three of these outcomes ask participants to repeat alternating movements of some type. Likewise, for Visual Span Backwards, the correlation with Visual Span Forwards was not accurately approximated by the single factor model. In addition, the correlations between Visual Span Backwards and the Mattis Dementia Rating Scale outcomes L and O were not well approximated. Mattis Dementia Rating Scale outcomes L and O involve copying drawings and, in this sense, also incorporate a visual component

that may be the source of the residual correlation between the outcomes. Given that the lack of fit is observed among outcomes that are related conceptually (e.g., that are parts of a subtest or a subscale), our next step was to consider a bifactor model.

Examining the pairwise correlations for the other items in this manner, we identified possible secondary factors to account for residual correlation. We applied an iterative process where we specified an additional secondary factor, refit the model and then checked the fit of this new model. Ultimately, we specified a bifactor model with one general cognitive ability factor and 3 secondary factors (for a total of $Q = 4$) as listed in Table 5.5. It is important to note that, although we identified these secondary factors using the posterior predictive model checks, they nevertheless have substantive interpretations as they link conceptually related outcomes. The second factor loads on Mattis Dementia Rating Scale outcomes I, J and K, test items that all involve repetition of alternating movements. The third factor loads on the Visual Span outcomes and Mattis Dementia Rating Scale outcomes L, O and V. These test items all include visual or drawing components. The fourth factor links three Mattis Dementia Rating Scale outcomes that ask participants to identify similarities and dissimilarities.

Bifactor semiparametric model For the semiparametric bifactor model with $Q = 4$, we drew 500,000 MCMC samples and discarded the first 50,000 as burn-in. We kept every 50th draw, leaving us with 9,000 posterior draws. As with the single factor model, we checked convergence using trace plots and the Geweke (Geweke, 1992) and Raftery-Lewis (Raftery and Lewis, 1995) diagnostic tests. Convergence was satisfactory but, compared to the single factor model, the mixing was considerably slower for a few of the secondary factors that exhibited high levels of autocorrelation. We should also note that the speed of convergence was influenced by the choice of hyperparameters for Σ and Ψ in the parameter expanded model. The bifactor model represented the dependence structure of the observed responses better. In Figure 5.9(b), we no longer see a discrepancy among the correlation distances suggesting model misfit. Figure 5.9(d) shows that the bifactor model provided a good fit to the observed eigenvalues well beyond the first eigenvalue. As can be seen in Figures 5.10(c) and 5.10(d), the bifactor model did a better job of replicating the pairwise

Table 5.5: Proposed factor structure for SIVD executive functioning outcomes. * indicates a non-zero factor loading to be estimated.

	Factor			
	1	2	3	4
MDRS G	*	0	0	0
MDRS H	*	0	0	0
MDRS I	*	*	0	0
MDRS J	*	*	0	0
MDRS K	*	*	0	0
MDRS L	*	0	*	0
MDRS O	*	0	*	0
Digit Sp Fwd	*	0	0	0
Digit Sp Bwd	*	0	0	0
Visual Sp Fwd	*	0	*	0
Visual Sp Bwd	*	0	*	0
Verb Flncy F	*	0	0	0
Verb Flncy A	*	0	0	0
Verb Flncy S	*	0	0	0
MDRS E	*	0	0	0
MDRS V	*	0	*	0
MDRS W	*	0	0	*
MDRS X	*	0	0	*
MDRS Y	*	0	0	*

rank correlations compared to the single factor model.

Table 5.6 displays posterior summaries for the regression parameters. We saw little change in our estimate for the parameter of interest, β_5 , the coefficient for frontal WMH

when adding additional factors. Thus, our substantive conclusion regarding the association between an individual's executive functioning and the volume of white matter hyperintensities in the frontal region of the brain remains the same whether we use the one-factor model or the better fitting bifactor model. Based on our semiparametric latent variable model, we expect a 1SD increase in frontal white matter hyperintensity volume to be associated with a 0.335SD decrease in the primary factor. In examining the other coefficients, we see that none of the 95% posterior credible intervals have shifted to the extent that we would alter our posterior belief about whether zero is a plausible value for the parameter. However, the coefficients for sex, age and total brain volume did decrease by 30-40% in magnitude.

Table 5.6: Posterior summaries for regression coefficients for single factor, $Q = 1$, and bifactor, $Q = 4$, models.

Coefficient	$Q = 1$			$Q = 4$		
	Mean	Median	95% CI	Mean	Median	95% CI
Sex	0.234	0.233	(-0.061, 0.516)	0.155	0.152	(-0.134, 0.440)
Education	0.354	0.354	(0.232, 0.479)	0.325	0.325	(0.206, 0.446)
Age	-0.126	-0.126	(-0.246, 0.004)	-0.078	-0.078	(-0.201, 0.044)
Total Brain Vol.	0.069	0.069	(-0.080, 0.215)	0.046	0.045	(-0.096, 0.194)
Frontal WMH Vol.	-0.330	-0.328	(-0.466, -0.205)	-0.335	-0.336	(-0.464, -0.208)

5.7 Discussion

In this chapter, we have developed a semiparametric latent variable model for multivariate mixed outcome data. This model, unlike common parametric latent variable modeling approaches for mixed outcome data (Sammel et al., 1997; Moustaki and Knott, 2000; Dunson, 2003; Shi and Lee, 1998), does not require the specification of conditional distributions for each outcome given the latent variables. In Chapter 4 when applying the generalized linear latent variable model framework to item response theory models, we found it challenging

to pick appropriate conditional distributions for each outcome encountered in real data, to extend the parametric models to account for all cases of distributions, and time-consuming to extend estimation methods appropriately. Moreover, the specification of outcome conditional distributions given the latent variables may be of little interest by itself in any research setting where the main question is in the relationship between a common factor (or factors) and a covariate of interest. Our proposed semiparametric latent variable framework allows one to model interdependencies among observed mixed outcome variables by specifying an appropriate latent variable model while, at the same time, avoiding the specification of outcome distributions conditional on the common latent variables. We have demonstrated this approach for the single-factor and bifactor models, incorporating a covariate effect on the general factor.

The extended rank likelihood can readily be employed with other latent variable models, including item response theory models (Van der Linden and Hambleton, 1997) and structural equation models (Bollen, 1989). In structural equation models, the focus is often on characterizing the relationship between latent variables and/or between latent variables and fixed covariates as in the case of our hierarchical model. In such cases where the focus is not on the loadings or outcome-related parameters, the proposed semiparametric approach would be quite useful in dealing with mixed outcome data. However, the extended rank likelihood may not be as useful in cases where outcome-specific parameters on the scale of the observed outcomes are of interest. For example, in item response theory models, one is often interested in examining the item difficulty and discrimination parameters to better understand the characteristics of individual test questions. When using the extended rank likelihood, the difficulty parameter, the analogue to the specific mean in the factor model, is not directly identifiable. Nonetheless, one could still carry out posterior inference by relying on the relationship between the difficulty parameter and the latent trait. In a two parameter item response theory model for binary outcomes, the probability of positive response when the factor score is set to zero is a one-to-one function of the difficulty parameter. Such an alternative, however, may render the semiparametric approach less convenient for a practitioner who is primarily interested in parameters characterizing the properties of individual outcomes.

We employed the semiparametric latent variable model to study the association between the volume of white matter hyperintensities in the frontal lobe and cognitive testing outcomes related to executive functioning from the Subcortical Ischemic Vascular Dementia (SIVD) study. The semiparametric latent variable model allowed us to analyze the mixed cognitive testing outcomes without requiring the specification of parametric distributions for the outcomes conditional on the latent variables. It has been hypothesized that a greater volume of frontal lobe white matter hyperintensities will be associated with more impaired executive functioning. Consistent with this hypothesis, we found a negative association between the primary factor in our model and the volume of white matter hyperintensities.

We started our model-building process by fitting the one-factor semiparametric model. We relied on posterior predictive model checks to evaluate whether the model well approximated the dependence structure among observed responses. We successfully used posterior predictive model checks to diagnose model misfit and to guide us in identifying a secondary factor structure for the bifactor model. Our posterior predictive checks approach can therefore be thought of as a method of exploratory bifactor analysis (Jennrich and Bentler, 2011) when the secondary factor structure is not known in advance. It also provides a mechanism by which statistical methodologists can work together with substantive experts to develop models that are theoretically justified and that are consistent with the data.

While we did not focus on formal model selection in this paper, one could explore the use of model fit criteria and other model selection methods to determine the factor structure for our semiparametric model. For example, one could use the methods of Knowles and Ghahramani (2011) and Rai and Daumé III (2009) to incorporate the Indian Buffet Process prior to simultaneously estimate the loadings, the loadings structure and the number of factors. Within the bifactor model framework, Jennrich and Bentler (2011) recently proposed using a rotation criterion to explore the secondary factor structure. Dunson (2006) presented a Bayesian model averaging approach that accounts for the uncertainty in the number of factors.

In our work with the cognitive testing data, we found that the semiparametric model was more elegant and much easier in implementation than the standard parametric approaches for mixed outcome data. Nonetheless, a formal comparison of the two methods is needed

to fully understand the differences and impact on conclusions. In addition, an alternative approach to handling mixed outcomes in practice is to use mainstream latent variable modeling software and treat all outcomes as ordered categorical. In cases where the number of categories exceeds the number that can be accommodated by existing software, the outcomes are typically consolidated into fewer categories to satisfy the software's constraints. It would similarly be useful to understand how the semiparametric model performs relative to this software-constrained approach.

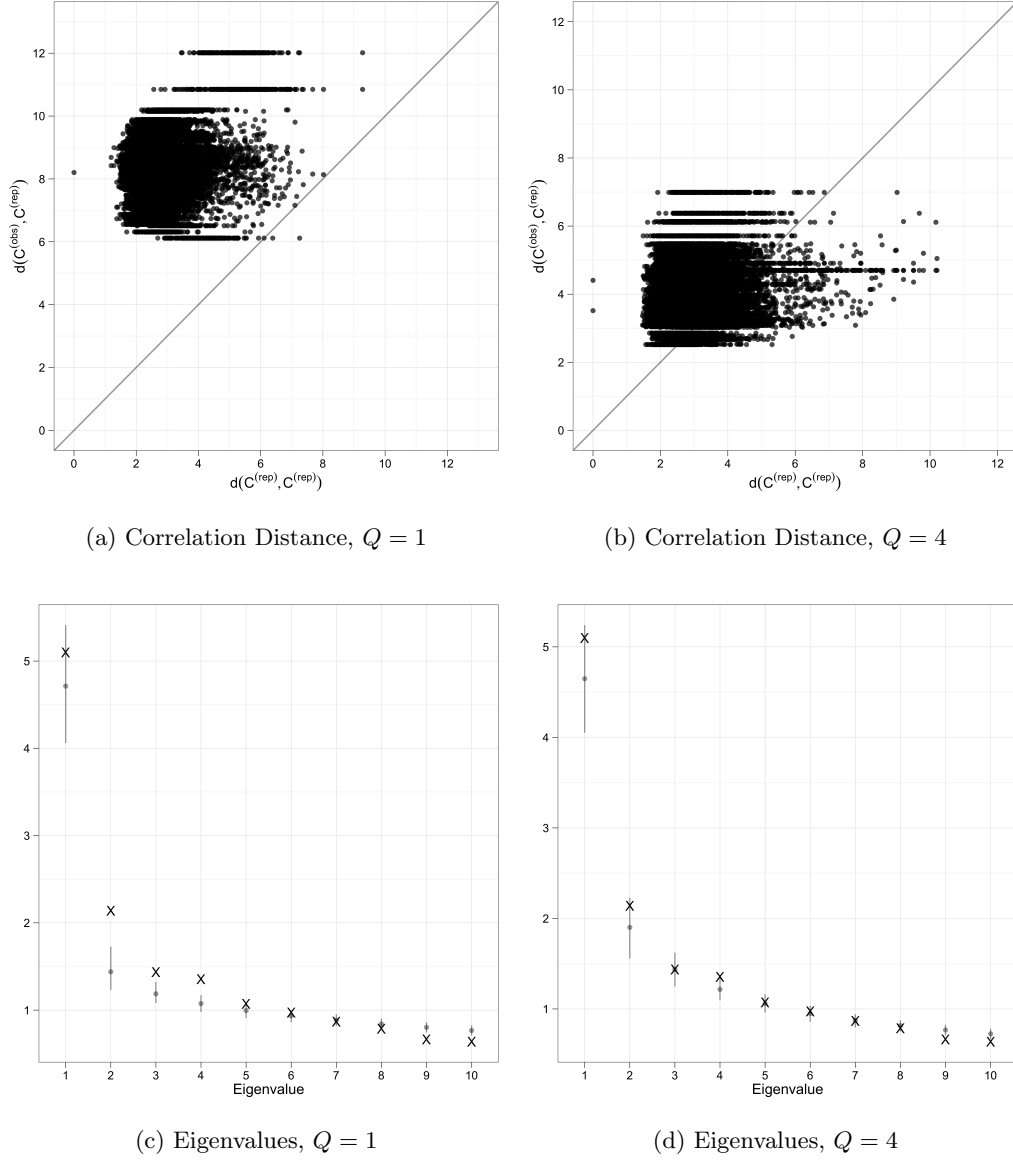


Figure 5.9: Correlation distance and eigenvalue plots for the $Q = 1$ and $Q = 4$ models. The left plot presents scatterplots of $d_{slid}(C^{obs}, C^{rep,m})$ versus $d_{slid}(C^{rep,m}, C^{rep,m'})$ for all replicated datasets. The grey line represents the 45 degree line. The right plot displays the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) of the largest ten eigenvalues calculated using replicated data. Eigenvalues computed from the observed data are denoted by a black “X”.

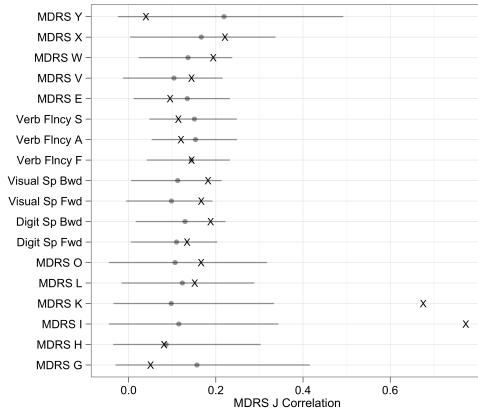
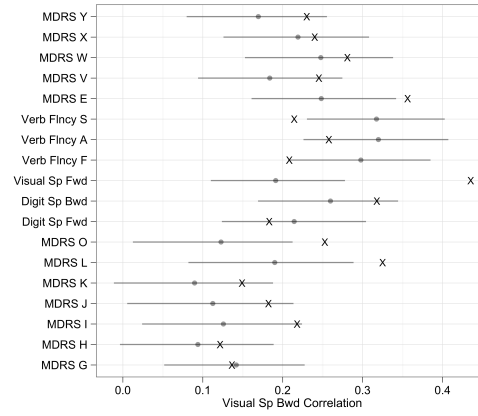
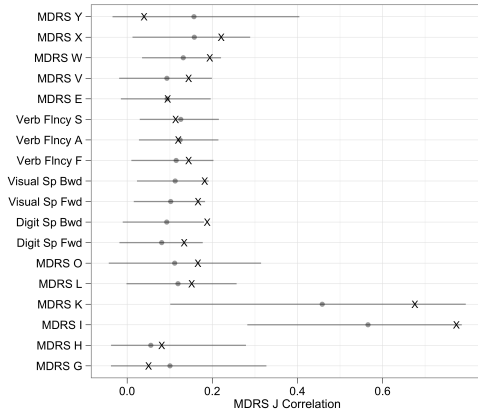
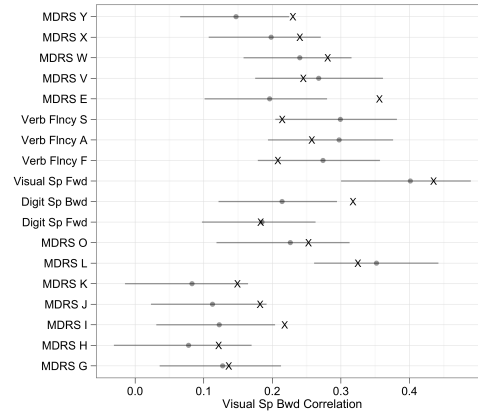
(a) MDRS J, $Q = 1$ (b) Visual Span Backwards, $Q = 1$ (c) MDRS J, $Q = 4$ (d) Visual Span Backwards, $Q = 4$

Figure 5.10: Pairwise correlation plots for the single factor ($Q = 1$) and bifactor models ($Q = 4$). Each pairwise correlation plot depicts the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) for Kendall's τ values calculated using replicated data. Kendall's τ values computed from the observed data are denoted by a black "X".

Chapter 6

SEMIPARAMETRIC MULTIVARIATE REGRESSION FOR MIXED OUTCOMES

6.1 Introduction

In Chapters 4 and 5, we investigated the association between multivariate mixed outcomes and a set of covariates using a hierarchical Bayesian latent variable model. As we saw, one of the challenges in applying these models is the specification of a latent structure that approximates the dependence structure in the data well. For binary and ordered categorical outcomes, an alternative method for investigating the associations between multivariate responses and covariates is the multivariate (ordered) probit regression model. In this chapter, we propose a semiparametric multivariate regression model for mixed data that is distinguished first by its ability to handle data of arbitrary type. This flexibility is achieved by using the extended rank likelihood method proposed by Hoff (2007). Second, the model readily accounts for residual correlations whereas, in the hierarchical Bayesian latent variable model, we must hypothesize or search for a latent structure to accommodate residual correlations. Finally, we take a reduced rank regression approach and factor the coefficient matrix into two vectors. As we demonstrate later, the semiparametric multivariate regression model with reduced rank form can be viewed as a more general formulation of the hierarchical semiparametric bifactor model with the latent variables marginalized out.

Ashford and Sowden (1970) originally proposed the multivariate probit model as a means of generalizing the standard probit model for correlated binary responses. However, until computational developments in the 1990s, use of the model was limited as analysis of the likelihood function was difficult without simplifying restrictions such as those in Ochi and Prentice (1984). Starting with Chib and Greenberg (1998), Bayesian MCMC and Monte Carlo EM (MCEM) methods were employed successfully for the estimation of these models. Recent areas of application of the multivariate probit model and variants include marketing

(Rossi and Allenby, 2003), medicine (Li and Schafer, 2008) and political science (Alvarez, Katz, Llamosa, and Martinez, 2009).

Reduced rank regression developed by Anderson (1951) and Izenman (1975) decomposes the coefficient matrix in a multivariate regression model into two lower rank (i.e., less than full rank) matrices and results in a more parsimonious, as well as perhaps a more interpretable, representation of the coefficient matrix. In our case, we assume a rank-one structure for the multivariate regression coefficient matrix. Our rank-one structure for the coefficient matrix may be viewed as consisting of row and column effects that are analogous to the population regression coefficients and factor loadings in the hierarchical latent variable models of Chapters 4 and 5. Schmidli (1996) notes that reduced rank regression models are well suited for situations where correlated outcomes may be represented by a factor analysis model where the factors are functions of covariates.

Most articles presenting the methods and theory of multivariate reduced rank regression assume the data are normally distributed. Moving beyond the Gaussian assumption, Variyam, Blaylock, and Smallwood (1998) employed reduced rank regression in a multivariate probit context. Yee and Hastie (2003) proposed reduced rank vector generalized linear models for different data types, focusing on categorical data. Heinen and Rengifo (2008) presented reduced rank multivariate dispersion models that include Gamma, inverse Gaussian, Poisson, binomial and negative binomial distributions among others. Finally, although not presented in the context of multivariate reduced rank regression, Hung and Wang (2011) proposed using the reduced rank parameterization of coefficients for matrix-variate logistic regression as a means of relating a matrix of covariates to a univariate binary response while extracting row and column information from the matrix of covariates.

The remainder of this chapter is organized as follows. In Section 6.2, we introduce the semiparametric multivariate reduced rank regression model. We additionally discuss identifiability and the relationship between the semiparametric multivariate regression model and the hierarchical semiparametric latent variable model. MCMC methods for estimation of the model are presented in Section 6.3. We demonstrate the model on a simulated data example in Section 6.4. Finally, we apply the model to the SIVD dataset that we previously analyzed with the IRT and semiparametric latent variable models in Sections 4.5 and 5.6.

6.2 Model

6.2.1 Model Formulation

Let $\mathbf{Y}$ be an $I \times J$ matrix of ordinal outcomes where I denotes the number of individuals and J the total number of outcomes (items) for each individual. Denote by $\mathbf{Z}$ an $I \times J$ matrix of latent responses such that

$$y_{ij} = g_j(z_{ij}) \quad (6.1)$$

where $i = 1, \dots, I$, $j = 1, \dots, J$ and g_j is a monotone transformation particular to outcome j . Let $g = (g_1, \dots, g_J)$ represent the vector of monotone transformations. Then,

$$\mathbf{y}_i = g(\mathbf{z}_i) \quad (6.2)$$

where $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ})^T$ and $\mathbf{z}_i = (z_{i1}, \dots, z_{iJ})^T$ are transposed row vectors from their respective matrices, $\mathbf{Y}$ and $\mathbf{Z}$.

Let $\mathbf{X}$ represent the $I \times P$ matrix of covariates with P covariates for each individual. Let $\mathbf{x}_i$ be a transposed row vector from this matrix. Then, under the semiparametric multivariate regression model, the vector of latent responses, $\mathbf{z}_i$, is distributed

$$\mathbf{z}_i \sim N(\mathbf{B}^T \mathbf{x}_i, \mathbf{C}) \quad (6.3)$$

where $\mathbf{B}$ is a $P \times J$ matrix of coefficients and $\mathbf{C}$ is a $J \times J$ covariance matrix. If each g_j is a step function and we estimate the thresholds at which each jump takes place, then we have the multivariate ordered probit model. In multivariate probit regression, the matrix $\mathbf{B}$ of coefficients typically contains all unique elements where each coefficient describes the relationship between a particular covariate and a particular outcome (in the presence of the other covariates). In other instances, the columns of the matrix $\mathbf{B}$ may be restricted to be equal so that the relationship between a covariate and an outcome is the same for all outcomes (for instance if the data are longitudinal as in Chib and Greenberg (1998) for the Six Cities example).

We propose using the parameterization $\mathbf{B} = \boldsymbol{\beta} \boldsymbol{\lambda}^T$ where $\boldsymbol{\beta}$ is a P -length vector and $\boldsymbol{\lambda}$ is a J -length vector. Equation (6.3) then becomes

$$\mathbf{z}_i \sim N(\boldsymbol{\lambda} \boldsymbol{\beta}^T \mathbf{x}_i, \mathbf{C}). \quad (6.4)$$

In this restricted formulation, the coefficient matrix $\mathbf{B} = \boldsymbol{\beta}\boldsymbol{\lambda}^T$ is now comprised of $P + J$ parameters rather than $P \times J$ parameters and is being represented by a rank one matrix. Employing the matrix normal distribution, the equivalent formulation for the matrix of latent responses is

$$\mathbf{Z} \sim \text{MN}(\mathbf{X}\boldsymbol{\beta}\boldsymbol{\lambda}^T, \mathbf{I}_I, \mathbf{C}). \quad (6.5)$$

The parameters $\boldsymbol{\beta}$ can be viewed as describing the relationships between the covariates and a “typical” outcome in the set of J outcomes. We may in turn think of the quantity $\boldsymbol{\beta}^T \mathbf{x}_i$ as the expected latent response and the parameters $\boldsymbol{\lambda}$ as scaling parameters with each vector element λ_j specific to the corresponding outcome j . The parameters $\boldsymbol{\lambda}$ serve a similar role as factor loadings in the semiparametric latent variable model. We explore this connection further in Section 6.2.3. We will rely on the extended rank likelihood for estimating the parameters of interest $\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{C}$ and therefore are not required to specify or estimate the transformations g_j .

6.2.2 Identifiability

The proposed model in equation (6.4) is not identifiable and there are two primary areas of concern. First, the respective signs and scales of $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ cannot be determined as $\mathbf{B} = \boldsymbol{\beta}\boldsymbol{\lambda}^T = (c\boldsymbol{\beta})(\boldsymbol{\lambda}/c)$ for any constant c . As a result, it is necessary to fix a value of c in order to identify these parameters. To do so, we restrict one (at the model user’s discretion) $\lambda_j > 0$ and set $\prod_j |\lambda_j| = 1$.

Secondly, let $\mathbf{d} = (d_1^{-1/2}, \dots, d_J^{-1/2})$ be a J -length vector of arbitrary constants and let $\mathbf{D}$ be a $J \times J$ matrix with the vector $\mathbf{d}$ on the diagonal and 0 elsewhere. Then, as noted by Chib and Greenberg (1998), $p(\mathbf{z}_i | \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{C}) = p(\mathbf{D}\mathbf{z}_i | \boldsymbol{\beta}, \mathbf{D}\boldsymbol{\lambda}, \mathbf{D}\mathbf{C}\mathbf{D})$. This equivalence will affect our estimation procedure as the scale of the latent response matrix $\mathbf{Z}$ is not fixed. To address the latter identifiability issue, we follow the convention in the multivariate probit analysis literature and restrict the covariance matrix $\mathbf{C}$ to be a correlation matrix. As will be discussed in Section 6.3, this restriction complicates the estimation process as sampling correlation matrices in MCMC can be challenging.

6.2.3 Relation To Hierarchical Semiparametric Latent Variable Model

In Chapter 5, we proposed a hierarchical semiparametric latent variable model that was constructed to relate multivariate mixed outcomes to covariates of interest. Recall that we formulated the model as follows,

$$y_{ij} = g_j(z_{ij}) \quad (6.6)$$

$$z_{ij} | \boldsymbol{\eta}_i \sim N(\boldsymbol{\lambda}_j^T \boldsymbol{\eta}_i, 1) \quad (6.7)$$

$$\boldsymbol{\eta}_i \sim N(\mathbf{m}_{\boldsymbol{\eta}_i}, \mathbf{I}_Q) \quad (6.8)$$

where $\boldsymbol{\lambda}_j$ and $\boldsymbol{\eta}_i$ are Q -length vectors. Moreover, $\mathbf{m}_{\boldsymbol{\eta}_i} = (\boldsymbol{\beta}^T \mathbf{x}_i, 0, \dots, 0)^T$ with Q elements. The vector $\mathbf{z}_i = (z_{i1}, \dots, z_{iJ})$ is conditionally distributed

$$\mathbf{z}_i | \boldsymbol{\eta}_i \sim N(\boldsymbol{\Lambda} \boldsymbol{\eta}_i, \mathbf{I}_J), \quad (6.9)$$

where $\boldsymbol{\Lambda}$ is the $J \times Q$ matrix of loadings.

If we integrate out $\boldsymbol{\eta}_i$,

$$p(\mathbf{z}_i | \boldsymbol{\Lambda}, \mathbf{m}_{\boldsymbol{\eta}_i}) = \int (2\pi)^{-J/2} \exp\left(-\frac{1}{2} (\mathbf{z}_i - \boldsymbol{\Lambda} \boldsymbol{\eta}_i)^T (\mathbf{z}_i - \boldsymbol{\Lambda} \boldsymbol{\eta}_i)\right) (2\pi)^{-Q/2} \exp\left(-\frac{1}{2} (\boldsymbol{\eta}_i - \mathbf{m}_{\boldsymbol{\eta}_i})^T (\boldsymbol{\eta}_i - \mathbf{m}_{\boldsymbol{\eta}_i})\right) d\boldsymbol{\eta}_i \quad (6.10)$$

$$= (2\pi)^{-J/2} \exp\left(-\frac{1}{2} \mathbf{z}_i^T \mathbf{z}_i - \frac{1}{2} \mathbf{m}_{\boldsymbol{\eta}_i}^T \mathbf{m}_{\boldsymbol{\eta}_i}\right) \int (2\pi)^{-Q/2} \exp\left[-\frac{1}{2} \boldsymbol{\eta}_i^T (\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q) \boldsymbol{\eta}_i + \frac{1}{2} \boldsymbol{\eta}_i^T (\boldsymbol{\Lambda} \mathbf{z}_i + \mathbf{m}_{\boldsymbol{\eta}_i}) + \frac{1}{2} (\boldsymbol{\Lambda} \mathbf{z}_i + \mathbf{m}_{\boldsymbol{\eta}_i})^T \boldsymbol{\eta}_i\right] d\boldsymbol{\eta}_i \quad (6.11)$$

$$= (2\pi)^{-J/2} |\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{z}_i^T \mathbf{z}_i - \frac{1}{2} \mathbf{m}_{\boldsymbol{\eta}_i}^T \mathbf{m}_{\boldsymbol{\eta}_i} + \frac{1}{2} (\boldsymbol{\Lambda} \mathbf{z}_i + \mathbf{m}_{\boldsymbol{\eta}_i})^T (\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q)^{-1} (\boldsymbol{\Lambda} \mathbf{z}_i + \mathbf{m}_{\boldsymbol{\eta}_i})\right] \cdot \int (2\pi)^{-Q/2} \left|(\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q)^{-1}\right|^{-1/2} \exp\left[\left(\boldsymbol{\eta}_i - (\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q)^{-1} (\boldsymbol{\Lambda} \mathbf{z}_i + \mathbf{m}_{\boldsymbol{\eta}_i})\right)^T (\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q) \left(\boldsymbol{\eta}_i - (\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q)^{-1} (\boldsymbol{\Lambda} \mathbf{z}_i + \mathbf{m}_{\boldsymbol{\eta}_i})\right)\right] d\boldsymbol{\eta}_i \quad (6.12)$$

$$= (2\pi)^{-J/2} |\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{z}_i^T \mathbf{z}_i - \frac{1}{2} \mathbf{m}_{\boldsymbol{\eta}_i}^T \mathbf{m}_{\boldsymbol{\eta}_i} + \frac{1}{2} (\boldsymbol{\Lambda} \mathbf{z}_i + \mathbf{m}_{\boldsymbol{\eta}_i})^T (\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q)^{-1} (\boldsymbol{\Lambda} \mathbf{z}_i + \mathbf{m}_{\boldsymbol{\eta}_i})\right] \quad (6.13)$$

$$= (2\pi)^{-J/2} |\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q|^{-1/2} \exp\left[-\frac{1}{2} \mathbf{z}_i^T \left(\mathbf{I}_J - \boldsymbol{\Lambda} (\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q)^{-1} \boldsymbol{\Lambda}^T\right) \mathbf{z}_i - \frac{1}{2} \mathbf{m}_{\boldsymbol{\eta}_i}^T \left(\mathbf{I}_Q - (\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q)^{-1}\right) \mathbf{m}_{\boldsymbol{\eta}_i} + \frac{1}{2} \mathbf{z}_i^T \boldsymbol{\Lambda} (\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q)^{-1} \mathbf{m}_{\boldsymbol{\eta}_i} + \frac{1}{2} \mathbf{m}_{\boldsymbol{\eta}_i} (\boldsymbol{\Lambda}^T \boldsymbol{\Lambda} + \mathbf{I}_Q)^{-1} \boldsymbol{\Lambda}^T \mathbf{z}_i\right]. \quad (6.14)$$

By applying Sylvester's Determinant Theorem and the Woodbury Matrix Identity, we have

$$|\mathbf{I}_Q + \mathbf{\Lambda}^T \mathbf{\Lambda}| = |\mathbf{I}_J + \mathbf{\Lambda} \mathbf{\Lambda}^T| \quad (6.15)$$

$$(\mathbf{\Lambda}^T \mathbf{\Lambda} + \mathbf{I}_Q)^{-1} = \mathbf{I}_Q - \mathbf{\Lambda}^T (\mathbf{I}_J + \mathbf{\Lambda} \mathbf{\Lambda}^T)^{-1} \mathbf{\Lambda} \quad (6.16)$$

$$\mathbf{I}_J - \mathbf{\Lambda} (\mathbf{\Lambda}^T \mathbf{\Lambda} + \mathbf{I}_Q)^{-1} \mathbf{\Lambda}^T = (\mathbf{I}_J + \mathbf{\Lambda} \mathbf{\Lambda}^T)^{-1} \quad (6.17)$$

$$\mathbf{\Lambda} (\mathbf{\Lambda}^T \mathbf{\Lambda} + \mathbf{I}_Q)^{-1} = (\mathbf{I}_J + \mathbf{\Lambda} \mathbf{\Lambda}^T)^{-1} \mathbf{\Lambda}. \quad (6.18)$$

If we denote $\mathbf{\Sigma}_{\mathbf{\Lambda}} = \mathbf{I}_J + \mathbf{\Lambda} \mathbf{\Lambda}^T$,

$$p(\mathbf{z}_i | \mathbf{\Lambda}, \mathbf{m}_{\eta_i}) = (2\pi)^{-1/2} |\mathbf{\Sigma}_{\mathbf{\Lambda}}|^{-1/2} \exp \left(-\frac{1}{2} \mathbf{z}_i^T \mathbf{\Sigma}_{\mathbf{\Lambda}}^{-1} \mathbf{z}_i - \frac{1}{2} \mathbf{m}_{\eta_i}^T \mathbf{\Lambda}^T \mathbf{\Sigma}_{\mathbf{\Lambda}}^{-1} \mathbf{\Lambda} \mathbf{m}_{\eta_i} \right. \quad (6.19)$$

$$\left. + \frac{1}{2} \mathbf{z}_i^T \mathbf{\Sigma}_{\mathbf{\Lambda}}^{-1} \mathbf{\Lambda} \mathbf{m}_{\eta_i} + \frac{1}{2} \mathbf{m}_{\eta_i}^T \mathbf{\Lambda}^T \mathbf{\Sigma}_{\mathbf{\Lambda}}^{-1} \mathbf{z}_i \right) \\ = (2\pi)^{-1/2} |\mathbf{\Sigma}_{\mathbf{\Lambda}}|^{-1/2} \exp \left(-\frac{1}{2} (\mathbf{z}_i - \mathbf{\Lambda} \mathbf{m}_{\eta_i})^T \mathbf{\Sigma}_{\mathbf{\Lambda}}^{-1} (\mathbf{z}_i - \mathbf{\Lambda} \mathbf{m}_{\eta_i}) \right). \quad (6.20)$$

Recall that $\mathbf{m}_{\eta_i} = (\boldsymbol{\beta}^T \mathbf{x}_i, 0, \dots, 0)^T$ so,

$$\mathbf{z}_i | \mathbf{\Lambda}, \boldsymbol{\beta}, \mathbf{x}_i \sim N(\boldsymbol{\lambda}_{q=1} \boldsymbol{\beta}^T \mathbf{x}_i, \mathbf{I}_J + \mathbf{\Lambda} \mathbf{\Lambda}^T), \quad (6.21)$$

where $\boldsymbol{\lambda}_{q=1}$ is the first column of the matrix $\mathbf{\Lambda}$. This is essentially the MIMIC model of Jöreskog and Goldberger (1975) with secondary factors to account for residual correlation. This mean structure is the same as the mean structure in the semiparametric multivariate reduced-rank regression model. In the hierarchical semiparametric latent variable model, we interpret the coefficients $\boldsymbol{\beta}$ as measuring the association between covariates and the primary latent factor that is a common component of the responses for an individual. In the multivariate reduced rank regression setting, we analogously view the parameters $\boldsymbol{\beta}$ as describing the relationships between the covariates and a typical outcome in the set of J outcomes.

Meanwhile, we characterize the dependence structure of $\mathbf{z}_i | \mathbf{x}_i$ by $\mathbf{I}_J + \mathbf{\Lambda} \mathbf{\Lambda}^T$. As seen with the SIVD dataset, unless we have identified a suitable bifactor structure, the dependence structure of the bifactor model may be insufficient to replicate the dependence structure observed within the data. When the number of outcomes is not large (as in the case of the SIVD dataset), it may make sense to estimate $\mathbf{C}$ directly rather than use a restricted approximation of it so as to avoid the search to find a suitable latent structure. If the

number of outcomes is large so that computation and estimation become difficult or slow, then the parsimonious formulation of the dependence structure by the latent variable model may be advantageous.

6.3 Estimation

We discussed the sampling the latent responses $\mathbf{Z}$ in Chapter 5 and we use the same approach for estimation of the model in equation (6.4). The primary remaining challenge is handling the identifiability restrictions from Section 6.2.2. There have been numerous developments in the sampling of correlation matrices including work by Chib and Greenberg (1998), Barnard, McCulloch, and Meng (2000), Liu (2001), Edwards and Allenby (2003), Liu and Daniels (2006), Lawrence, Bingham, Liu, and Nair (2008). Tabet (2007) summarizes much of this work.

We follow the procedure of Edwards and Allenby (2003) who sample covariance matrices by ignoring the lack of identifiability during the MCMC sampling. To obtain identified correlation matrix estimates, they post-process the draws from the MCMC chains. This procedure can also be considered as a parameter expansion approach (Liu et al., 1998; Liu and Wu, 1999) to estimation. In this case, the parameter expansion approach is used to facilitate implementation rather than to improve efficiency. Whereas it may be difficult to sample correlation matrices, there exist straightforward means to sample covariance matrices using the inverse Wishart distribution. Edwards and Allenby (2003) draw covariance matrices using the inverse Wishart distribution and, in a post-processing step, transform these covariance matrices to correlation matrices. Convergence is monitored using the correlation matrix draws. Hoff (2007) uses this approach in estimating the semiparametric Gaussian copula model and Lawrence et al. (2008) takes a similar approach but makes the transformation after each draw rather than for all draws combined in a post-processing step.

Compared to the alternate approaches of Chib and Greenberg (1998) and Barnard et al. (2000), this approach is easily implemented. A shortcoming of this approach is its sensitivity to prior specification for the unidentified parameters. If the prior is too weak, then convergence and numerical stability issues may arise (Lawrence et al., 2008; Tabet, 2007). If the prior is too strong, it may determine the posterior.

We also handle the unidentifiability of $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ through post-processing. To identify these parameters, recall that we restrict one $\lambda_j > 0$ (at the user's discretion) and set $\prod_j |\lambda_j| = 1$. Alternatively we could restrict the scale of $\boldsymbol{\lambda}$ by choosing to set one of the λ_j equal to 1 as is sometimes done with the loadings in latent variable modeling (Fox, 2010).

In the parameter expansion terminology of Ghosh and Dunson (2009), equation (6.4) with the identifiability restrictions in Section 6.2.2 constitutes our inferential model. Our parameter expanded working model then is

$$\mathbf{z}_i^* \sim N\left(\boldsymbol{\lambda}^* (\boldsymbol{\beta}^*)^T \mathbf{x}_i, \mathbf{C}^*\right), \quad (6.22)$$

where $\boldsymbol{\lambda}^*$, $\boldsymbol{\beta}^*$, and $\mathbf{C}^*$ are the working parameters.

Let $\mathbf{D}$ be a $J \times J$ matrix with the square root of the diagonal elements of $\mathbf{C}^*$ on its diagonal and 0, elsewhere and let $\lambda_{j'}$ be the element chosen to be restricted positive. The following transformations link the working model to the inferential model,

$$\begin{aligned} \tilde{\boldsymbol{\lambda}} &= \mathbf{D}^{-1} \boldsymbol{\lambda}^*, \\ c_{\boldsymbol{\lambda}} &= \prod_j |\tilde{\lambda}_j|, \\ \mathbf{z}_i &= \mathbf{D}^{-1} \mathbf{z}_i^*, \\ \boldsymbol{\lambda} &= \text{sgn}\left(\tilde{\lambda}_{j'}\right) \tilde{\boldsymbol{\lambda}} / c_{\boldsymbol{\lambda}}, \\ \mathbf{C} &= \mathbf{D}^{-1} \mathbf{C}^* \mathbf{D}^{-1}, \\ \boldsymbol{\beta} &= \text{sgn}\left(\tilde{\lambda}_{j'}\right) c_{\boldsymbol{\lambda}} \boldsymbol{\beta}^*. \end{aligned} \quad (6.23)$$

Before discussing the Gibbs sampling steps of making draws from the working model, we complete the model by specifying prior distributions for $\boldsymbol{\lambda}^*$, $\boldsymbol{\beta}^*$, and $\mathbf{C}^*$ rather than directly for $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$ and $\mathbf{C}$. The priors are

$$\boldsymbol{\lambda}^* \sim N(\mathbf{m}_{\boldsymbol{\lambda}}, \mathbf{S}_{\boldsymbol{\lambda}}), \quad (6.24)$$

$$\boldsymbol{\beta}^* \sim N(\mathbf{m}_{\boldsymbol{\beta}}, \mathbf{S}_{\boldsymbol{\beta}}), \quad (6.25)$$

$$\mathbf{C}^* \sim \text{Inv. Wishart}(\nu_C, \phi_C). \quad (6.26)$$

We sample from the posterior distribution for $(\mathbf{Z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\beta}^*, \mathbf{C}^*)$ using Gibbs sampling with a hybrid (Hamiltonian) Monte Carlo (HMC) step to draw from the full conditional for $\boldsymbol{\lambda}^*$.

We employ the HMC step for $\boldsymbol{\lambda}^*$ to speed convergence as we found directly sampling from the full conditional for $\boldsymbol{\lambda}^*$ exhibited very slow convergence. Recall from Section 2.2 that HMC uses the derivative of the log of the full conditional distribution to produce directed moves.

The Gibbs sampling algorithm proceeds as follows.

1. **Draw latent responses $\mathbf{Z}^*$.** For each i and j , sample z_{ij}^* from $p(z_{ij}^* | \boldsymbol{\lambda}^*, \boldsymbol{\beta}^*, \mathbf{C}^*, \mathbf{Z}_{(-i)(-j)}^*, \mathbf{Z}^* \in D(\mathbf{Y}))$. More specifically, for each j and for each $y = \text{unique}\{y_{1j}, \dots, y_{nj}\}$,

$$z_{ij}^* \sim \text{TN}_{(z_l^*, z_u^*)}(\mu_{z_{ij}^*}, \sigma_{z_{ij}^*}^2) \quad (6.27)$$

where TN denotes truncated normal and z_l^*, z_u^* define the lower and upper truncation points,

$$z_l^* = \max\{z_{kj}^* : y_{kj} < y\} \quad (6.28)$$

$$z_u^* = \min\{z_{kj}^* : y_{kj} > y\}. \quad (6.29)$$

The conditional mean and variance are

$$\mu_{z_{ij}^*} = \mathbf{M}_{ij} + (z_{i,-j}^* - \mathbf{M}_{i,-j}) (\mathbf{C}_{-j,-j}^*)^{-1} \mathbf{C}_{-j,j}^* \quad (6.30)$$

and

$$\sigma_{z_{ij}^*}^2 = \mathbf{C}_{j,j}^* - \mathbf{C}_{j,-j}^* (\mathbf{C}_{-j,-j}^*)^{-1} \mathbf{C}_{-j,j}^* \quad (6.31)$$

where

$$M = \mathbf{X}\boldsymbol{\beta}^* (\boldsymbol{\lambda}^*)^T. \quad (6.32)$$

2. **Draw $\boldsymbol{\lambda}^*$.** We rely on HMC to sample $\boldsymbol{\lambda}^*$. In order to implement HMC, we calculated the derivative of the log joint density $\log p(\mathbf{Z}^*, \boldsymbol{\lambda}^* | \boldsymbol{\beta}^*, \mathbf{C}^*, \mathbf{X}) = \log p$ with respect to $\boldsymbol{\lambda}^*$,

$$\frac{d \log p}{d \boldsymbol{\lambda}^*} = (\mathbf{C}^*)^{-1} \left(\mathbf{Z}^* - \mathbf{X}\boldsymbol{\beta}^* (\boldsymbol{\lambda}^*)^T \right)^T \mathbf{X}\boldsymbol{\beta}^* - \mathbf{S}_{\boldsymbol{\lambda}}^{-1} (\boldsymbol{\lambda} - \mathbf{m}_{\boldsymbol{\lambda}}). \quad (6.33)$$

3. **Draw β^* .** We sample β^* according to its full conditional distribution,

$$\beta^* \sim N \left(\left(\left((\lambda^*)^T (\mathbf{C}^*)^{-1} \lambda^* \right) \mathbf{X}^T \mathbf{X} + \mathbf{S}_\beta^{-1} \right)^{-1} \left(\mathbf{X}^T \mathbf{Z}^* (\mathbf{C}^*)^{-1} \lambda^* + \mathbf{S}_\beta^{-1} \mathbf{m}_\beta \right), \right. \\ \left. \left(\left((\lambda^*)^T (\mathbf{C}^*)^{-1} \lambda^* \right) \mathbf{X}^T \mathbf{X} + \mathbf{S}_\beta^{-1} \right)^{-1} \right). \quad (6.34)$$

4. **Draw $\mathbf{C}^*$.** We sample $\mathbf{C}^*$ according to its full conditional distribution,

$$\mathbf{C}^* \sim \text{Inv. Wishart} \left(\nu_C + I, \left(\mathbf{Z}^* - \mathbf{X} \beta^* (\lambda^*)^T \right)^T \left(\mathbf{Z}^* - \mathbf{X} \beta^* (\lambda^*)^T \right) + \phi_C \right). \quad (6.35)$$

Once we have drawn a number of posterior samples using the above steps, we discard some number of initial draws as burn-in and we post-process the draws using the transformations (6.23) to obtain λ , β , and $\mathbf{C}$.

6.4 A Simulated Data Example

We tested the proposed model against simulated data to evaluate how well the model recovered the data generating parameters. We simulated responses to $J = 20$ outcomes for $I = 400$ participants. For each participant $i = 1, \dots, 400$, we generated $P = 5$ covariate values comprising $\mathbf{x}_i$. The covariate values along with randomly drawn parameter values for λ , β and $\mathbf{C}$ were used to simulate latent responses according to equation (6.4). We scaled the data generating values of λ so that their product was equal to 1. Finally, we drew a random number of thresholds to generate observed responses by discretizing the latent responses.

To estimate the data generating parameters, we drew 60,000 samples using the model, discarded the first 30,000 as burn-in and kept every 10th draw of the remaining samples. Tables 6.1 and 6.2 present posterior summaries for λ and β . In the tables we see that the model successfully recovered the data generating values for λ and β . The posterior mean estimates were generally close to the the data-generating values. 19 of 20 posterior 95% credible intervals for λ parameters contained the data generating truth and 5 of 5 for β contained the corresponding truth. 97.9% of the posterior 95% credible intervals for the $J(J-1)/2$ parameters in $\mathbf{C}$ contained the true parameter values.

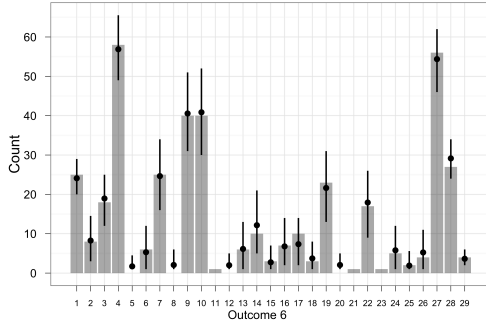
Table 6.1: Posterior summary of λ .

Outcome	Truth	Post Mean	Post. Median	Post. 95% CI
1	1.839	1.785	1.779	(1.587, 2.018)
2	0.079	0.058	0.058	(0.038, 0.078)
3	-2.311	-2.449	-2.438	(-2.800, -2.153)
4	0.313	0.327	0.325	(0.283, 0.378)
5	-3.034	-2.957	-2.942	(-3.430, -2.554)
6	-2.653	-2.450	-2.438	(-2.779, -2.177)
7	-1.832	-1.799	-1.792	(-2.015, -1.635)
8	1.412	1.391	1.384	(1.235, 1.581)
9	2.670	2.689	2.679	(2.388, 3.051)
10	0.847	0.843	0.841	(0.754, 0.946)
11	-0.797	-0.780	-0.778	(-0.876, -0.705)
12	3.136	3.130	3.122	(2.771, 3.547)
13	2.038	2.073	2.069	(1.848, 2.354)
14	-3.530	-3.421	-3.405	(-3.903, -3.051)
15	-0.015	-0.026	-0.026	(-0.046, -0.005)
16	0.520	0.509	0.507	(0.457, 0.570)
17	3.618	3.082	3.062	(2.516, 3.762)
18	-0.074	-0.088	-0.088	(-0.111, -0.065)
19	-2.766	-2.907	-2.893	(-3.430, -2.460)
20	1.977	1.918	1.909	(1.656, 2.222)

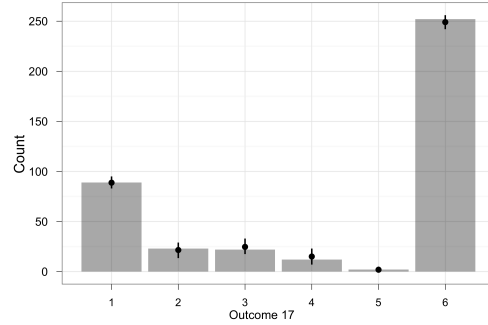
The posterior predictive model checks introduced in Section 2.4 similarly show that the model does a good job of approximating the simulated data. Figure 6.1 demonstrates that the model is able to replicate the marginal distributions. Similarly, Figure 6.2 indicates that the model approximates the observed rank correlations well.

Table 6.2: Posterior summary of β .

Outcome	Truth	Post Mean	Post. Median	Post. 95% CI
1	3.341	3.476	3.490	(3.156, 3.682)
2	3.699	3.855	3.872	(3.508, 4.077)
3	0.401	0.412	0.413	(0.374, 0.440)
4	-1.006	-1.053	-1.057	(-1.117, -0.957)
5	-0.693	-0.720	-0.723	(-0.765, -0.655)



(a) Outcome 6



(b) Outcome 17

Figure 6.1: Histograms of the observed outcome scores in the simulated dataset. The black dots indicate the mean count across replicated datasets for each grade. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.

6.5 Application to the SIVD Study

In Chapters 4 and 5, we analyzed the SIVD dataset using the parametric IRT and semiparametric latent variable models. Here, we analyze this dataset by using the semiparametric multivariate reduced rank regression model. Recall that the goal of our analysis is to investigate the relationship between executive functioning and the volume of white matter hyperintensities located in the frontal lobe of the brain. To explore this relationship, we

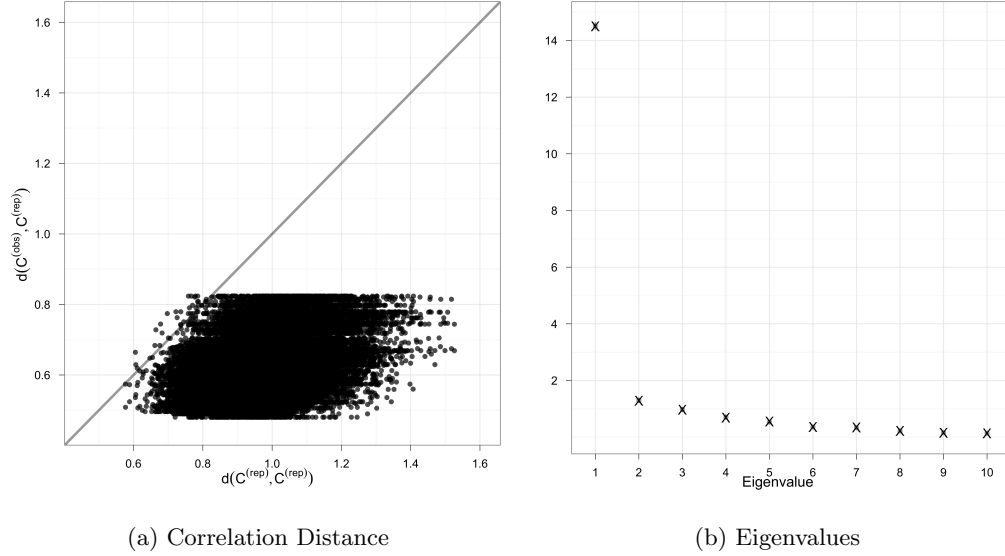


Figure 6.2: The left plot presents scatterplots of $d_{sld}(C^{obs}, C^{rep,m})$ versus $d_{sld}(C^{rep,m}, C^{rep,m'})$ for all replicated datasets. The grey line represents the 45 degree line. The right plot displays the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) of the largest ten eigenvalues calculated using replicated data. Eigenvalues computed from the observed data are denoted by a black “X”.¹

used the executive functioning outcomes from the SIVD neuropsychological battery and MRI-measured brain volumes. There are $J = 19$ different outcomes for $I = 341$ study participants. Controlling for age, sex, education, and total brain volume, we have

$$\beta^T \mathbf{x}_i = \beta_1 \text{Sex}_i + \beta_2 \text{Educ}_i + \beta_3 \text{Age}_i + \beta_4 \text{Vol}_i + \beta_5 \text{WMH}_i, \quad (6.36)$$

where Sex_i is the participant’s sex (Female=1, Male=0), Educ_i is the number of years of education, Vol_i is the total brain volume of the participant, and WMH_i is the frontal white matter hyperintensity volume. We used standardized versions of the continuous predictor variables.

To estimate the model, we employed the Gibbs sampling algorithm with the HMC step detailed in Section 6.3. We drew 200,000 samples and discarded the first half as burn-in,

keeping every 25th draw. The large number of draws was required to ensure convergence among a small subset of the parameters, particularly the regression coefficient for the Sex covariate. We used trace plots and the Geweke (Geweke, 1992) and Raftery-Lewis (Raftery and Lewis, 1995) diagnostic tests to assess convergence. Despite restricting only λ_1 to be positive, all of the posterior means for $\boldsymbol{\lambda}$ were positive suggesting that all of an individual's responses are positively correlated as we would expect based on intuition and the analyses in the previous chapters.

Table 6.3 presents posterior summaries for the $\boldsymbol{\beta}$ parameters. Recall that we standardized the continuous covariates (education, age, total brain volume, frontal white matter hyperintensity volume) to have mean 0 and standard deviation 1. We immediately notice that the coefficient for our covariate of interest, frontal white matter hyperintensity volume, is negative and that the 95% credible interval (-0.201, -0.052) does not contain 0 as the corresponding intervals did not in previous analyses.

Table 6.3: Posterior summaries for regression coefficients, $\boldsymbol{\beta}$.

Coefficient	Mean	Median	95% CI
Gender	0.192	0.194	(0.046, 0.351)
Education	0.129	0.132	(0.044, 0.212)
Age	-0.073	-0.073	(-0.144, -0.004)
Total Brain Vol.	0.038	0.037	(-0.030, 0.110)
Frontal WMH Vol.	-0.125	-0.127	(-0.201, -0.052)

We then evaluated the fit of the model to the data using the posterior predictive model checking methods discussed in Section 2.4. Figure 6.3 displays the histograms for the observed responses to the Verbal Fluency A and Mattis Dementia Rating Scale E items along with posterior predictive summaries. In each case, the model appeared to do a satisfactory job of approximating the data as did the semiparametric bifactor model.

Content with the model's ability to approximate the marginal distributions of the ob-

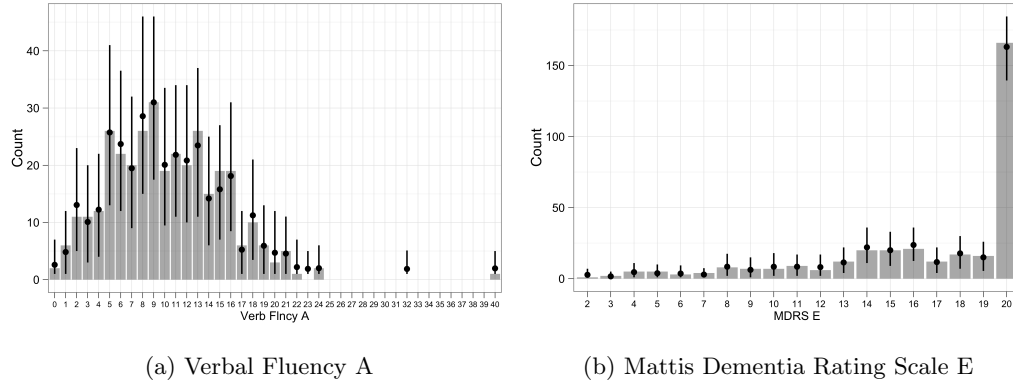


Figure 6.3: Histograms of the observed scores for the Verbal Fluency A and Mattis Dementia Rating Scale E. The black points indicate the mean count across replicated datasets for each grade. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.

served responses well, we examined the model's ability to replicated the dependence structure among the observed responses. Figure 6.4 presents the model's ability to replicate the rank correlations observed in the data at a global level. Both the correlation distance and eigenvalue plots suggested that the semiparametric multivariate reduced-rank regression model adequately represents the dependence structure in the data. Although we achieved similarly good fit with the semiparametric latent variable model, we had to identify a suitable bifactor structure in order to achieve this type of fit. No such model search was necessary in this case.

At the individual item level, we further checked the pairwise rank correlations. Figure 6.5 displays the pairwise correlation plots for Mattis Dementia Rating Scale J and Visual Span Backwards. Not surprisingly, the model also did a good job of approximated the observed pairwise rank correlations at the individual item level.

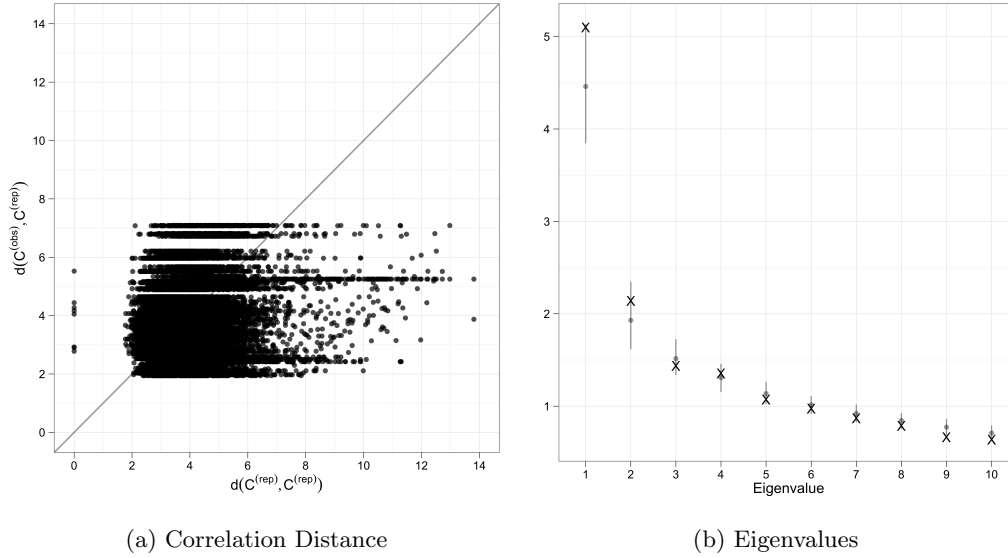


Figure 6.4: The left plot presents a scatterplot of $d_{sld}(C^{obs}, C^{rep,m})$ versus $d_{sld}(C^{rep,m}, C^{rep,m'})$ for all replicated datasets. The grey line represents the 45 degree line. On the right, a plot of top ten eigenvalues depicting the mean prediction (grey point) and 95% prediction intervals (grey line segment) of the eigenvalues calculated using replicated data. Eigenvalues computed from the observed data are denoted by a black “X”.

6.6 Discussion

In this chapter, we introduced an alternative to latent variable models for investigating the association between multivariate mixed outcomes and covariates of interest. Our semiparametric multivariate regression model directly relates covariates to multivariate outcomes using a reduced rank representation of the coefficient matrix. In particular, we used a rank-one form that is the product of two vectors; one of these vectors may be viewed as a set of regression coefficients between the covariates and a typical outcome and the other as a vector that scales the typical response to each particular outcome. This simple structure facilitates interpretation and retains the same mean structure as the hierarchical semiparametric latent variable model. As with the semiparametric latent variable model in Chapter 5, our

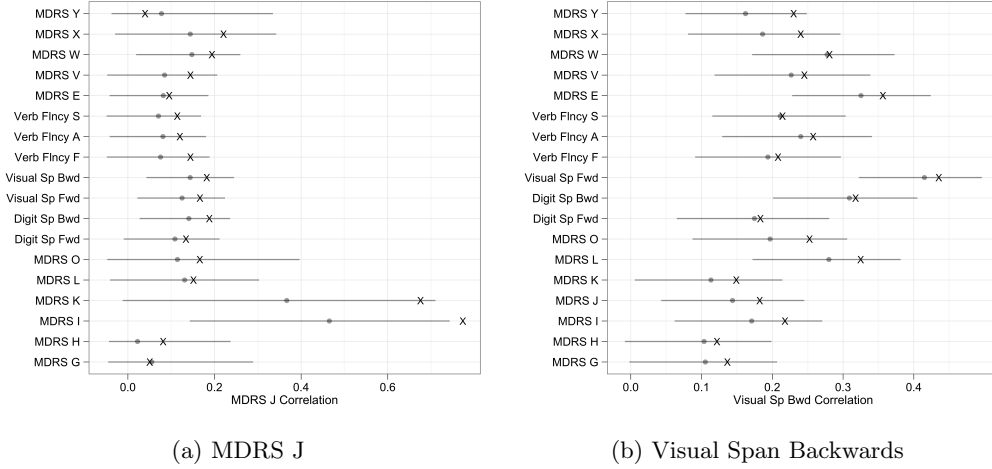


Figure 6.5: Pairwise correlation plots for Mattis Dementia Rating Scale J and Visual Span Backwards. Each pairwise correlation plot depicts the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) for Kendall’s τ values calculated using replicated data. Kendall’s τ values computed from the observed data are denoted by a black “X”.

reduced rank regression model for mixed outcomes uses the extended rank likelihood so that we may accommodate outcomes of arbitrary type.

A critical difference between the semiparametric latent variable model and the semiparametric multivariate regression model is that correlations of the latent responses underlying the observed variables are directly estimated in the semiparametric multivariate regression model estimation. In the semiparametric bifactor model, we must identify and impose a secondary factor structure. However, the semiparametric bifactor model allows us to obtain estimates of the primary factor scores for each individual. To the degree that practitioners may be interested in these estimates, the latent variable estimates may be an important feature of the semiparametric latent variable model.

We applied the semiparametric multivariate regression model to the SIVD study data. We found that the model did fit the data well according to the posterior predictive model checking methods. As in the previous chapters, we estimated a negative association be-

tween the executive functioning indicators and the volume of frontal lobe white matter hyperintensities.

A key assumption of the proposed model is a rank-one structure for coefficient matrix. It is of course possible that a higher rank structure might be necessary to describe the relationship between the outcomes and the covariates. In general, it would be valuable to explore the effects of model misspecification in semiparametric multivariate regression and gain a better understanding of how model misspecification may be detected.

Finally, we noticed in the case of the SIVD analysis that the model was particularly slow to converge for all of the parameters. In the case where the regression coefficients are slow to converge, it may be worth introducing an additional working parameter, an unidentified intercept, in the parameter expanded estimation scheme to help reduce autocorrelation. In the hierarchical semiparametric latent variable model in Section 5.4, we obtained better mixing of the regression coefficient chains with the introduction of an unidentified intercept.

6.7 *Postscript: Model Review*

Beginning with Chapter 4 and continuing through this chapter, we considered three different models for addressing the same question concerning the association between covariates and multivariate outcomes of mixed type. Each of the three proposed models offers a means to make inference on the relationship between covariates and a multivariate set of outcomes. Figures 6.6, 6.7 and 6.8 present diagrams depicting the models, allowing us to visually distinguish them. In each plot, the circles represent latent variables and the squares represent observed variables with solid borders corresponding to outcomes and dashed borders corresponding to covariates. The one-sided solid arrows represent regression associations while the two-sided dashed lines represent residual correlations.

With the IRT model for mixed outcomes in Chapter 4 (Figure 6.6), we take a parametric approach to modeling mixed outcomes whereas the subsequent models in Chapters 5 (Figure 6.7) and this chapter (Figure 6.8) adopt a semiparametric approach. In contrast to the parametric approach in the IRT model for mixed outcomes, the semiparametric models do not require the modeler to specify the appropriate type of distribution for each outcome conditional on the latent variable. Moreover, if one does encounter additional types

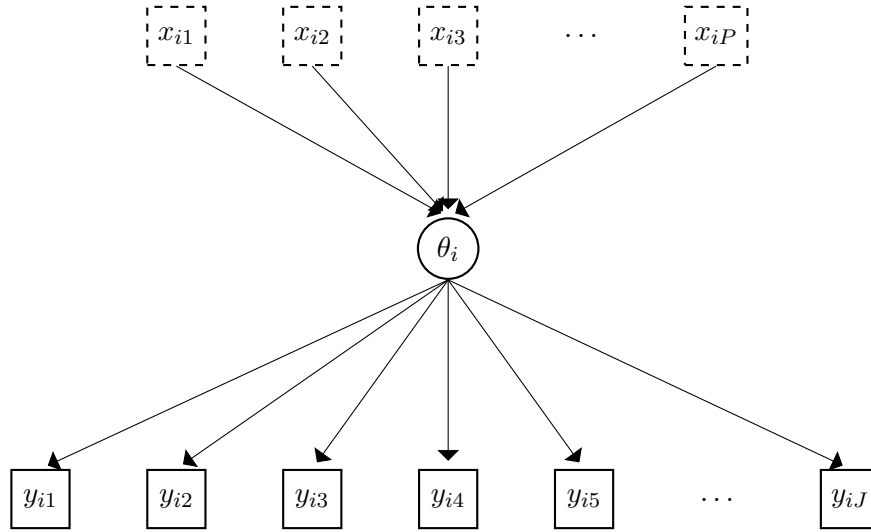


Figure 6.6: Diagram of the IRT Model for Mixed Outcomes. The circles represent latent variables and the squares represent observed variables with solid borders corresponding to outcomes and dashed borders corresponding to covariates. The one-sided solid arrows represent regression coefficients.

of outcomes that we have not covered here, the semiparametric approach does not require additional modeling and programming to accommodate these additional types of outcomes whereas the parametric approach would.

Of course, there are situations where the application of the proposed IRT model for mixed outcomes will have advantages. If we are interested in the parameters that describe the different items such as the difficulty parameter, a location parameter that has no equivalent in the semiparametric latent variable model, then the parametric approach may be preferred or at a minimum may be more direct. Also, in using the semiparametric approach, we are depending solely on the ranks of the data. As a result, we lose some information in that small differences in the orderings of the data may not reflect possibly large differences on the observed data scale. However, one advantage of this feature is that inference is invariant to monotonic transformations of the data.

The IRT model for mixed outcomes and the semiparametric latent variable model rely

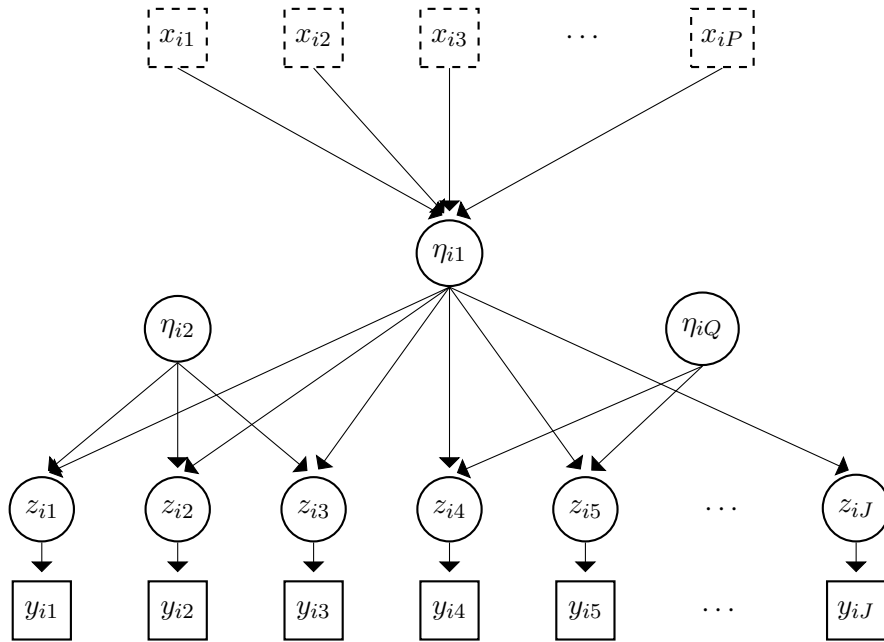


Figure 6.7: Diagram of the Semiparametric Latent Variable Model with Bifactor Structure for Mixed Outcomes. The circles represent latent variables and the squares represent observed variables with solid borders corresponding to outcomes and dashed borders corresponding to covariates. The one-sided solid arrows represent regression coefficients.

on the latent variables to induce dependence among the outcomes. As with the standard IRT model, we restrict the IRT model for mixed outcomes to a single latent variable model. This limits the type of dependence structures that may be approximated by the model. For the semiparametric latent variable model, we employ a bifactor structure to accommodate residual correlations. This structure expands the types of dependence structures that may be represented but also necessitates specification of the structure. The semiparametric multivariate regression model (with rank one structure) meanwhile does not require any search for a sufficient bifactor structure as it estimates the residual correlations directly.

In the semiparametric multivariate regression model, however, we must estimate every element of the correlation matrix rather than rely on the parsimonious description that a latent variable model affords. As the number of outcomes grows, the number of parameters

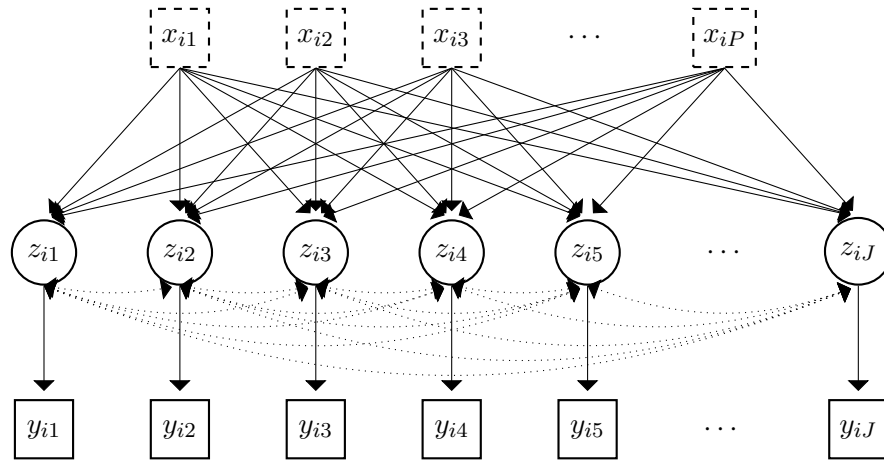


Figure 6.8: Diagram of the Semiparametric Multivariate Regression Model for Mixed Outcomes. The circles represent latent variables and the squares represent observed variables with solid borders corresponding to outcomes and dashed borders corresponding to covariates. The one-sided solid arrows represent regression coefficients while the two-sided dashed lines represent residual correlations.

to be estimated will greatly increase and the matrix inversions required in the estimation procedure will further slow down estimation. In situations with a large number of outcomes, the latent variable approach to multivariate modeling may be advantageous. Although we limited the IRT model for mixed outcomes to a single latent variable, multidimensional IRT models (Reckase, 2009) could be used to define a latent structure in a manner similar to our approach for the semiparametric latent variable model to attempt to give a better fit to the observed rank correlations. Finally, in some circumstances, the latent variable estimates may be of value to practitioners who are interested in individual level parameters summarizing a multivariate set of responses.

We demonstrated all three of the models on the SIVD study data, investigating the association between executive functioning indicators and the volume of white matter hyperintensities in the frontal lobe. The mixed types of outcomes (e.g., binary, count, ordered categorical, censored) in the data presented a challenge in applying more traditional methods to answer our substantive question. Figure 6.9 displays posterior predictive check

plots for each model of the marginal distributions. The top row of plots is for Verbal Fluency A and the bottom row is for Mattis Dementia Rating Scale E. Although the IRT model did not fair poorly, the semiparametric models did a better job of approximating the data.

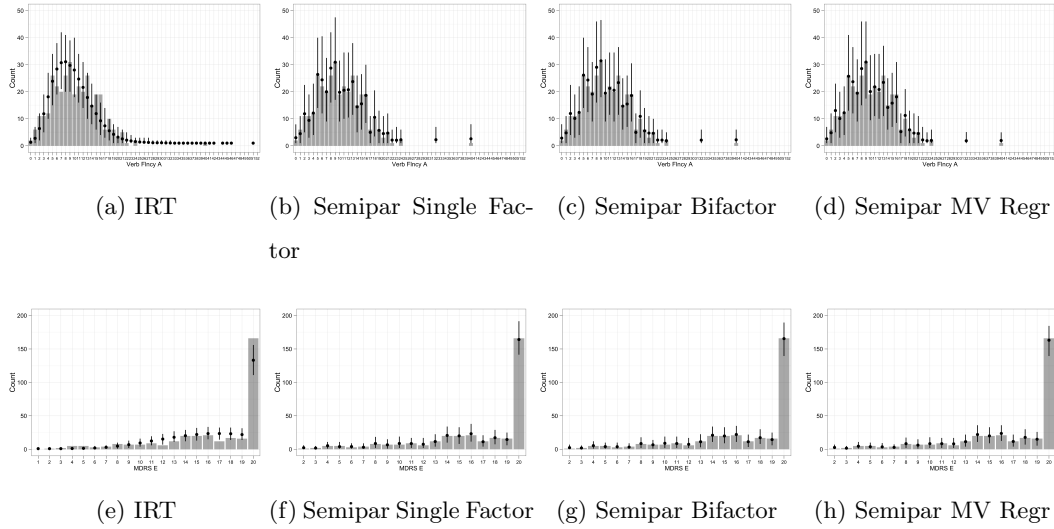


Figure 6.9: Histograms of the observed scores for the Verbal Fluency A (top row) and Mattis Dementia Rating Scale E (bottom row) by model. The black points indicate the mean count across replicated datasets for each score. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.

In assessing the fit of the different models, we also focused on each model's ability to replicate the rank correlations observed in the data. Figure 6.10 displays plots of the correlation distance and eigenvalue posterior predictive model checks for fit of all models to the SIVD data. Models with a single latent variable applied in Chapters 4 and 5 did a poor job of replicating the rank correlations observed in the data. The discrepancy in model fit disappeared with the semiparametric bifactor model used in Chapter 5 once we identified a suitable secondary factor structure. Meanwhile, the semiparametric multivariate regression model replicated the dependence structure of the data well on our initial attempt.

In analyzing data from the SIVD study, all three of the models came to the same conclusion, suggesting a negative relationship between the executive functioning indicators

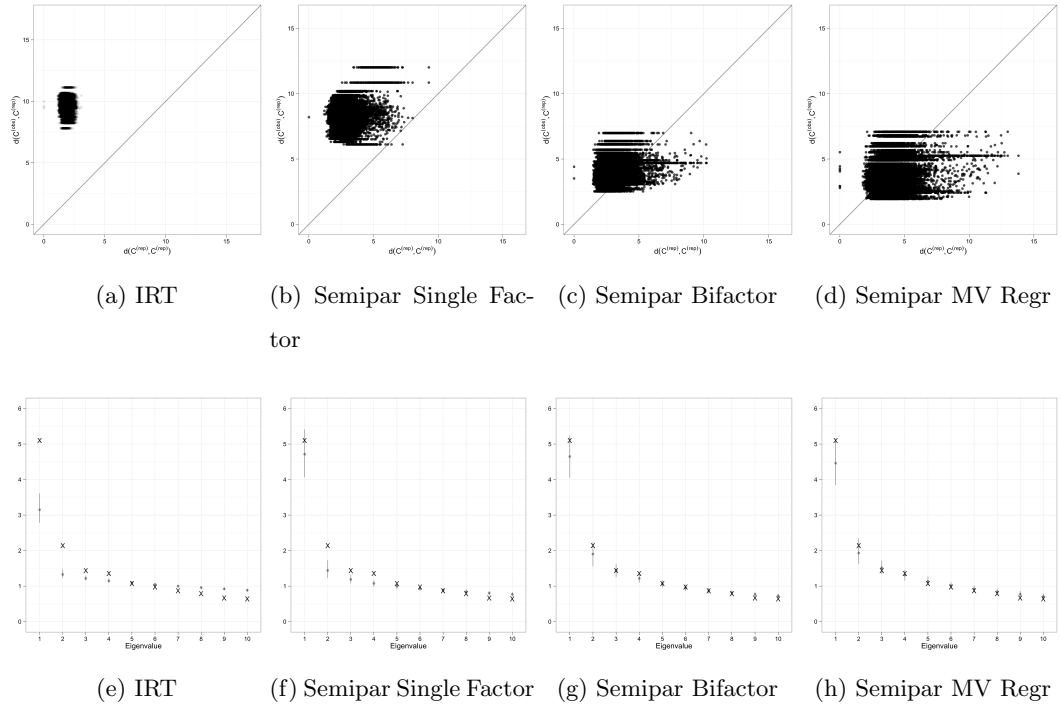


Figure 6.10: Correlation distance and eigenvalue plots for the IRT model, the single factor and bifactor semiparametric latent variable models, and the semiparametric multivariate regression model. The top row of plots present scatterplots of $d_{sld}(C^{obs}, C^{rep,m})$ versus $d_{sld}(C^{rep,m}, C^{rep,m'})$ for all replicated datasets. The grey line represents the 45 degree line. The bottom row of plots display the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) of the largest ten eigenvalues calculated using replicated data. Eigenvalues computed from the observed data are denoted by a black "X".

and the volume of frontal lobe white matter hyperintensities after controlling for age, sex, education and total brain volume. All three models, however, took different approaches to analyzing the data, resulting in different fits of the model to the data. As highlighted above, each approach has different strengths and weaknesses and, as is often the case when choosing an appropriate statistical method, the substantive context may dictate which method may

be most appropriate.

Chapter 7

SEMIPARAMETRIC CORRELATED BAYESIAN PARTIAL
MEMBERSHIP MODEL**7.1 Introduction**

Mixture models provide a model-based approach to clustering. Population-level mixture models describe a population as a collection of subpopulations where each observational unit belongs exclusively to one of the subpopulations. Individual-level mixture models, on the other hand, allow each observational unit to belong to multiple subpopulations at once, with varying degrees of membership among individuals. Individual-level mixture models can be viewed as a relaxation of population-level mixtures such as finite mixture or latent class models. It is important to note that degree of membership is conceptually different from uncertainty regarding membership.

The common example for an individual-level model is the genetic composition of individuals. Individuals may originate from a multitude of ancestries and their varying degree of membership to these ancestries may be reflected in their genetic traits. A common form of individual-level mixture model is the mixed membership (MM) model. MM models such as the Grade of Membership (GoM) and latent Dirichlet allocation (LDA) have been used to characterize subpatterns of disease and disability (Woodbury et al., 1978; Erosheva, Fienberg, and Joutard, 2007), classify documents as mixtures of topics (Blei et al., 2003), describe voters using Irish election data (Gormley and Murphy, 2009), determine the sector composition of stocks (Doyle and Elkan, 2009), and apply network analysis to social networks and protein interaction networks (Airoldi, Blei, Fienberg, and Xing, 2008).

Beyond MM models, there have been many related efforts in machine learning under the category of fuzzy or soft clustering. Fuzzy k-means is one popular soft clustering method (Bezdek, 1981). Within model-based clustering, the Bayesian partial membership (BPM) model, introduced by Heller et al. (2008), is another recently developed class of

individual-level mixture models. In this chapter, we focus on the BPM model. Heller et al. (2008) applied the model to binary data. In this chapter, we focus on the case of normally distributed data and extend the BPM model to accommodate more flexible correlation structures among class memberships and mixed outcomes.

The BPM model in its proposed form and MM models in general employ a Dirichlet distribution for the class memberships, assuming a near independence structure among class memberships. In cases where class memberships may be correlated, it may be more appropriate to explicitly model the correlations among classes. To allow for more flexible correlations among class or subpopulation memberships, we use the logistic normal distribution proposed by Blei and Lafferty (2007) for mixed membership topic models. The ability to model correlation among class memberships could be useful for cognitive testing data where increasing membership in a group typified by individuals with a deficit in one cognitive area could be associated with an increase in membership in another group typified by individuals with a deficit in a different cognitive area.

As in Chapters 5 and 6, we apply the extended rank likelihood method (Hoff, 2007) to accommodate mixed type data. The BPM and MM models can model mixed outcomes using a parametric approach that specifies different exponential family distributions appropriate for each outcome type. Similarly to other chapters in this dissertation, we however take the semiparametric approach so as to avoid specifying a distribution for each outcome. Although Heller et al. (2008) formulated the BPM model with the capability to handle mixed data, they only use binary data in their examples. Previously, Jorgensen and Hunt (1996) proposed finite mixture models for categorical and continuous data. Banerjee, Merugu, Dhillont, and Ghosh (2004) developed methods for soft clustering with mixed data by relying on Bregman divergences. The general MM framework described in Erosheva (2002) also allows for different types of outcomes.

In Section 7.2, we present the BPM model and, in Section 7.3, we compare it to the MM model. We detail our extensions to the BPM model in Section 7.4. Section 7.4 also describes the estimation procedure for our extended BPM model. We demonstrate the extended BPM model on simulated data in Section 7.5. Finally, we use the extended BPM model to analyze two datasets, NBA player statistics from the 2010-11 season (Section 7.6)

and the SIVD study data described earlier (Section 7.7).

7.2 The Bayesian Partial Membership Model

To best understand the BPM model, we follow the exposition of Heller et al. (2008) and first consider a standard finite mixture model. Let $\mathbf{y}_i$ be a vector of p outcomes for the i th observational unit. Let K denote the number of classes or mixture components, let $p_k(\cdot)$ specify the density particular to class k and let $\boldsymbol{\psi}_k$ specify the parameters characterizing $p_k(\cdot)$ for class k . The probability density for $\mathbf{y}_i$, given a collection of parameters $\boldsymbol{\Psi} = (\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_K)$ for all K classes and given the latent class membership indicator π_{ik} for class k and individual i , is

$$p(\mathbf{y}_i | \boldsymbol{\Psi}, \boldsymbol{\pi}_i) = \sum_k^K \pi_{ik} p_k(\mathbf{y}_i | \boldsymbol{\psi}_k) \quad (7.1)$$

where $\pi_{ik} \in \{0, 1\}$, $\sum_k \pi_{ik} = 1$. For the MM model, one replaces π_{ik} with a membership score g_{ik} . Instead of being restricted to $\{0, 1\}$, the membership score g_{ik} is allowed to range continuously between 0 and 1 but is subject to the constraint $\sum_k g_{ik} = 1$. Note that neither π_{ik} nor g_{ik} are observed. Furthermore MM models employ the additional assumption that, conditional on the membership vector $\mathbf{g}$, the observations are independent so that

$$p(\mathbf{y}_i | \boldsymbol{\Psi}, \mathbf{g}_i) = \prod_j^J \sum_k^K g_{ik} p_{jk}(y_{ij} | \boldsymbol{\psi}_{jk}). \quad (7.2)$$

MM models have most commonly been applied to binary and ordinal categorical data (Eroshova, 2002).

An alternative means of specifying (7.1) is through the product of the densities,

$$p(\mathbf{y}_i | \boldsymbol{\Psi}, \boldsymbol{\pi}) = \prod_k^K p_k(\mathbf{y}_i | \boldsymbol{\psi}_k)^{\pi_k}. \quad (7.3)$$

Heller et al. (2008) use this formulation to specify the BPM model. In their formulation,

$$p(\mathbf{y}_i | \boldsymbol{\Psi}, \mathbf{g}) = \frac{1}{c} \prod_k^K p_k(\mathbf{y}_i | \boldsymbol{\psi}_k)^{g_k} \quad (7.4)$$

where again $g_k \in [0, 1]$ and c is a normalizing constant.

Heller et al. (2008) focus on the case where p_k is an exponential family density (denoted $\text{Exp}(\cdot)$),

$$p_k(\mathbf{y}_i|\psi_k) = \text{Exp}(\psi_k). \quad (7.5)$$

Here, ψ_k denotes the natural parameters for the particular exponential family distribution. If one substitutes this expression into (7.4), one sees that

$$p(\mathbf{y}_i|\Psi, \mathbf{g}) = \text{Exp}\left(\sum_k g_k \psi_k\right). \quad (7.6)$$

Thus $\mathbf{y}_i$ is distributed according to the same exponential family distribution as the classes but with natural parameters that are a convex combination of the natural parameters of the different classes. The use of exponential family distributions allows one to model a variety of outcome types.

Beyond the difference in formulation, an additional difference between the MM and BPM models lies in the model assumptions. Whereas the MM model assumes the outcomes are conditionally independent given the class memberships, the BPM model makes no such assumption.

Heller et al. (2008) complete specification of the BPM model with the following hierarchical model. Let α be a K -dimensional vector of hyperparameters. Mixture weights are drawn from a Dirichlet distribution

$$\rho \sim \text{Dir}(\alpha). \quad (7.7)$$

Individual weights are then drawn from a Dirichlet distribution as well

$$\mathbf{g}_i \sim \text{Dir}(a\rho) \quad (7.8)$$

where a is a positive scale parameter for which they specify an exponential prior distribution. They specify a conjugate prior for the natural parameters for each class such that

$$\psi_k \sim \text{Conj}(\boldsymbol{\lambda}, \boldsymbol{\nu}). \quad (7.9)$$

Finally the observations are distributed according to (7.6).

Although this model may seem very different in aim from the latent variable models in the previous chapters, the similarities become more apparent if one considers the relationship between latent class models and continuous latent variable models (Heinen, 1996; Bartholomew et al., 2011). Similarly to Erosheva (2002), who compared the Grade of Membership model and factor analysis models, we consider a factor analysis model with $K - 1$ factors, $\boldsymbol{\eta}_i$, so that

$$y_{ij} \sim N(\mu_j + \boldsymbol{\lambda}_j^T \boldsymbol{\eta}_i, \sigma_j^2). \quad (7.10)$$

If the factors $\boldsymbol{\eta}_i$ are restricted so that $\sum_k \eta_{ik} \leq 1$, define the K -length vectors $\mathbf{g}_i = (\eta_{i1}, \dots, \eta_{i(K-1)}, 1 - \sum_k \eta_{ik})$ and $\mathbf{m}_j = (\lambda_{i1} + \mu_j, \dots, \lambda_{i(K-1)} + \mu_j, \mu_j)$. These definitions lead to the following restatement:

$$y_{ij} \sim N(\mathbf{m}_j^T \mathbf{g}_i, \sigma_j^2) \quad (7.11)$$

which is the BPM model for normally distributed data with class-specific means and a common variance. Thus, the BPM model may be thought of as a type of factor analysis where the factor scores are constrained.

7.3 Comparison of Partial Membership and Mixed Membership Structures

In this section, we contrast the BPM model with the MM model. Although the BPM model is very similar in spirit to the MM model, the differences in formulation and model assumptions can be observed in data generated by two individual-level mixture models. First, as in Heller et al. (2008), we compare the data generated by each model when the class densities are Gaussian. Following that, we study the model differences when the data are binary, comparing the probability of a success under each model and how the differences in the probability of a success vary for different class memberships.

Consider the simple scenario of two outcomes and three classes where the outcomes for each class are normally distributed. Initially, we fixed the variance to be common across all three classes, 4 for the first outcome and 25 for the second. We specified the means presented in Table 7.1 for each class. We then generated 1000 random membership vectors from a Dirichlet ($a\boldsymbol{\rho}$) distribution with $a = 1$ and $\boldsymbol{\rho} = (1/3, 1/3, 1/3)$. Using these membership

Table 7.1: Means by class for outcomes 1 and 2.

Outcome	Class		
	1	2	3
1	10	25	40
2	25	40	10

scores, we simulated 1000 bivariate outcomes. The results are depicted in Figure 7.1. The left plot shows the MM model and the center plot displays the corresponding BPM model with a diagonal covariance matrix (that is, local independence was assumed as in the case of the MM model), $\Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 25 \end{pmatrix}$. The right plot shows BPM model results with a full covariance matrix where the variances of the outcomes are the same as the previous two cases but the correlation between the outcomes was set to 0.4 so that $\Sigma = \begin{pmatrix} 4 & 4 \\ 4 & 25 \end{pmatrix}$. In Figure 7.1, the MM model under the parameter values appears to generate points in three columns. Looking more closely, each column can be divided horizontally into three parts corresponding to the means for each class for y_{i2} . Dividing the columns in this manner produces $K^2 = 9$ clusters of points, consistent with the equivalent latent class representation described by Erosheva (2002). The BPM model, in both the diagonal and full covariance matrix cases, generates points in a more cloud-like structure. One can see that the BPM model with the full covariance matrix generates a set of points that is “tilted”, albeit slightly, as compared to the set generated by the BPM model with a diagonal covariance matrix.

By varying the values of a , we can further compare the models. If we set $a = 10$, the membership scores will fluctuate more closely around $1/3$ than $a = 1$. Figure 7.2 presents 1000 generated datapoints with membership scores generated from a Dirichlet ($a\rho$) distribution with $a = 10$ and $\rho = (1/3, 1/3, 1/3)$. In the case of the MM model, the 3^2 clusters become slightly more apparent while the data generated by the BPM models reduce to single clusters with less variation. If we set $a = 1/10$, the membership scores tend to be

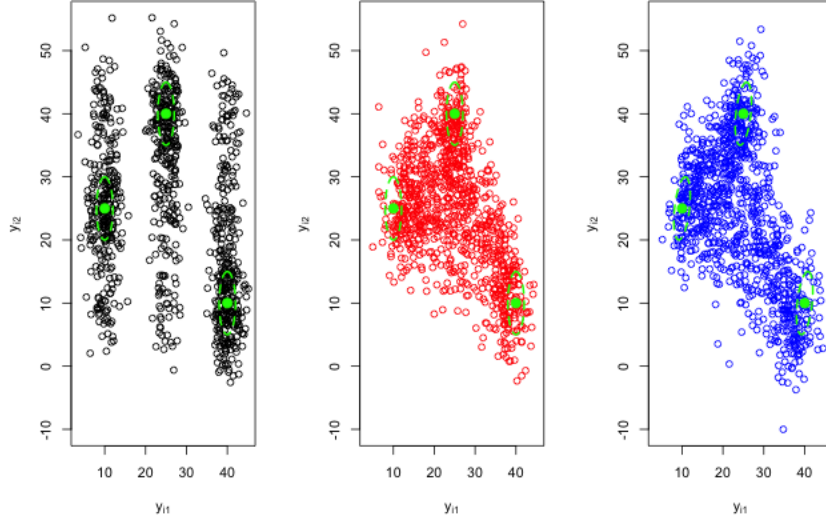


Figure 7.1: Simulated data according to different individual-level mixture model formulations. The green points represent the class centers and the green ellipse represents a 1SD contour. Left: MM. Center: BPM with diagonal covariance matrix. Right: BPM with full covariance matrix.

closer to the extremes 0 or 1. Figure 7.3 presents the simulated data from each model with this set of membership scores. The three plots now appear largely similar. The primary differences are that the set of points generated by the BPM model with full covariance matrix is “tilted” as compared to the other two and that the MM model appears to show greater variation in points on the periphery.

We subsequently set class-specific variances (and correlations in the case of the full covariance BPM) and again generated points from each individual-level mixture model. The class specific variances are listed in Table 7.2. For the BPM model with full covariance matrix, the correlations by class were set to 0.4, -0.4, and 0.7. Figure 7.4 presents the data generated by these models. The sets of points generated by the MM and BPM model with diagonal covariance appear rectangular in shape. The set of points from the BPM model with diagonal covariance is more densely populated in the center while one can faintly make

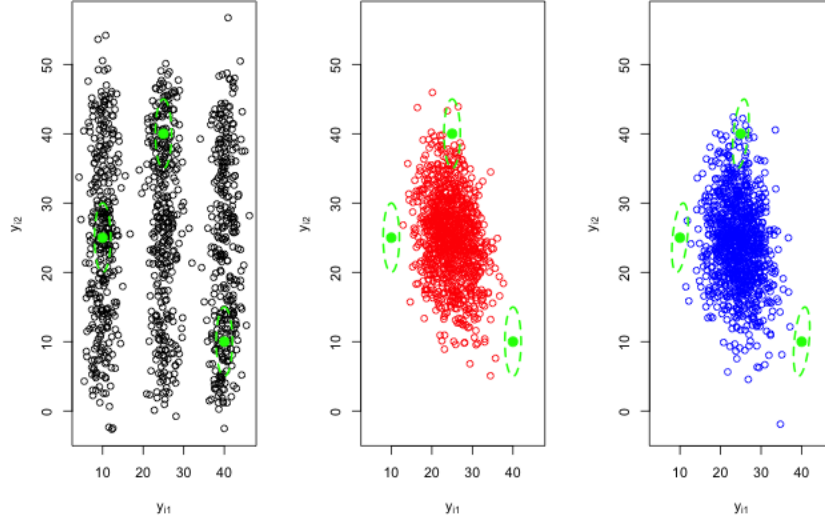


Figure 7.2: Simulated data according to different individual-level mixture model formulations with membership scores generated using scale parameter $a = 10$. Left: MM. Center: BPM with diagonal covariance matrix. Right: BPM with full covariance matrix.

Table 7.2: Variances by class for outcomes 1 and 2.

Outcome	Class		
	1	2	3
1	4	25	9
2	36	1	4

the clusters in the set of points generated by the MM model. The BPM model with full covariance matrices on the other hand is more triangular in structure.

Figures 7.5 and 7.6 provide the corresponding plots for membership vectors generated by $a = 10$ and $a = 1/10$ respectively. With $a = 10$, we again see the greater concentration of points into a single cluster for the BPM models while the different clusters become a

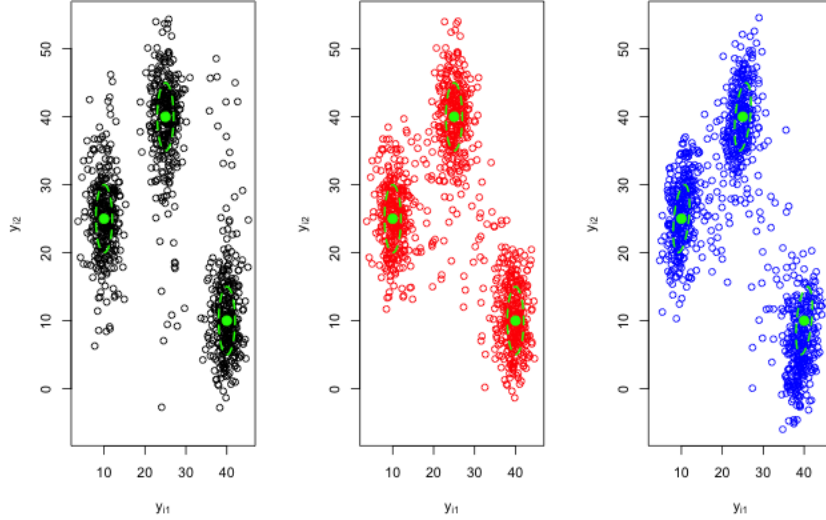


Figure 7.3: Simulated data according to different individual-level mixture model formulations with membership scores generated using scale parameter $a = 1/10$. Left: MM. Center: BPM with diagonal covariance matrix. Right: BPM with full covariance matrix.

little more apparent for the MM Model. In the case of $a = 1/10$, the MM and BPM with diagonal covariance models again appear very similar. The full covariance BPM model however displays a triangular boundary with an empty center.

To provide further explication of the similarities and the differences between the MM and BPM models, we consider a single binary outcome, $X \in \{0, 1\}$, for an arbitrary individual with membership vector $\mathbf{g} = (g_1, \dots, g_k, \dots, g_K)$ assuming K classes. Assume that each extreme profile outcome is generated by a Bernoulli distribution with the probability of success, p_k , different for each class. Under the MM model we have,

$$p(x|\mathbf{g}) = \sum_k g_k p_k^x (1 - p_k)^{1-x}, \quad (7.12)$$

so that

$$p(0|\mathbf{g}) = \sum_k g_k (1 - p_k) \quad (7.13)$$

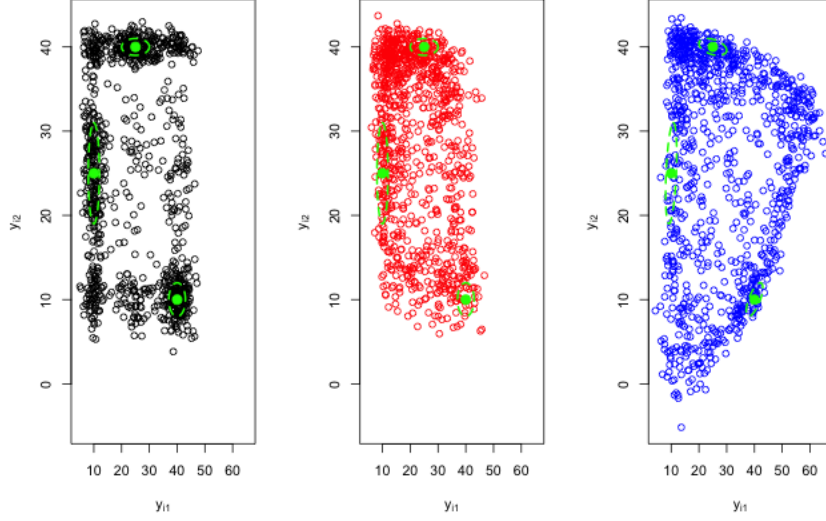


Figure 7.4: Simulated data according to different individual-level mixture model formulations. The green points represent the class centers and the green ellipse represents a 1SD contour. Left: MM. Center: BPM with diagonal covariance matrix. Right: BPM with full covariance matrix.

$$p(1|\mathbf{g}) = \sum_k g_k p_k. \quad (7.14)$$

Thus we see that the probability mass function for a binary outcome under the MM model can be expressed as that of a Bernoulli where the probability of success is a convex combination/weighted arithmetic mean of the class parameters,

$$p(x|\mathbf{g}) = \left(\sum_k g_k p_k \right)^x \left(\sum_k g_k (1 - p_k) \right)^{1-x}. \quad (7.15)$$

Under the partial membership model, we have

$$\begin{aligned} p(x|\mathbf{g}) &= \frac{1}{c} \prod_k [p_k^x (1 - p_k)^{1-x}]^{g_k} \\ &= \frac{1}{c} \prod_k [p_k^x]^{g_k} \prod_k [(1 - p_k)^{1-x}]^{g_k} \\ &= \frac{1}{c} \prod_k [p_k^{g_k}]^x \prod_k [(1 - p_k)^{g_k}]^{1-x} \end{aligned}$$

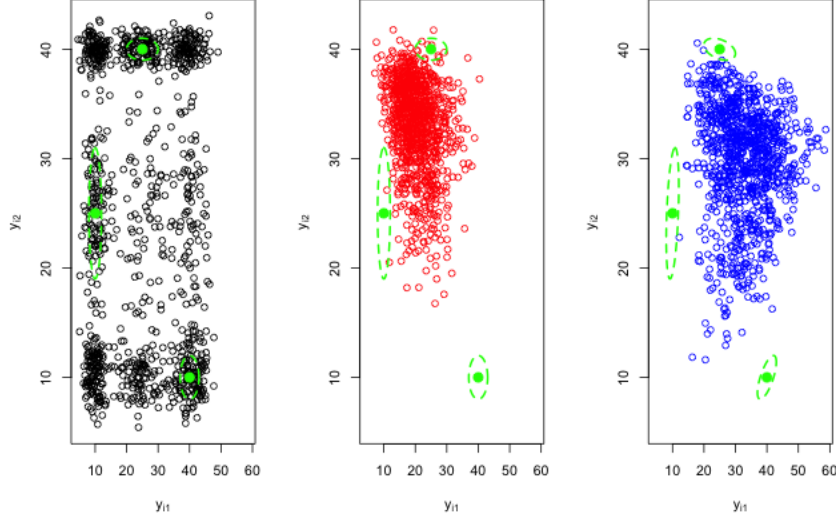


Figure 7.5: Simulated data according to different individual-level mixture model formulations with membership scores generated with scale parameter $a = 10$. Left: MM. Center: BPM with diagonal covariance matrix. Right: BPM with full covariance matrix.

$$= \frac{1}{c} \left(\prod_k p_k^{g_k} \right)^x \left(\prod_k (1 - p_k)^{g_k} \right)^{1-x}.$$

Notice that the first product term is essentially the weighted geometric mean of the class probabilities of success. To figure out the normalizing constant c ,

$$\begin{aligned} \sum_x p(x) &= p(0) + p(1) \\ &= \frac{1}{c} \prod_k (1 - p_k)^{g_k} + \frac{1}{c} \prod_k p_k^{g_k} = 1 \\ c &= \prod_k p_k^{g_k} + \prod_k (1 - p_k)^{g_k}. \end{aligned}$$

Thus

$$p(x|\mathbf{g}) = \left(\frac{\prod_k p_k^{g_k}}{\prod_k p_k^{g_k} + \prod_k (1 - p_k)^{g_k}} \right)^x \left(\frac{\prod_k (1 - p_k)^{g_k}}{\prod_k p_k^{g_k} + \prod_k (1 - p_k)^{g_k}} \right)^{1-x}. \quad (7.16)$$

Under the partial membership, $x|\mathbf{g}$ also has a Bernoulli distribution but where the probability of success is related to the weighted geometric mean of the class probabilities of success

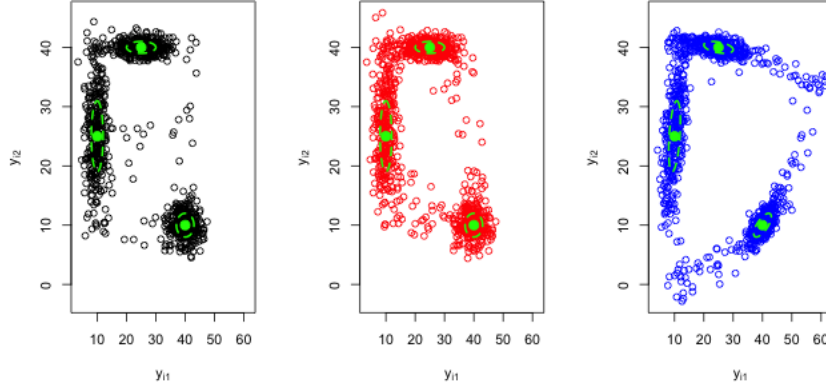


Figure 7.6: Simulated data according to different individual-level mixture model formulations with membership scores generated with scale parameter $a = 1/10$. Left: MM. Center: BPM with diagonal covariance matrix. Right: BPM with full covariance matrix.

as opposed to the weighted arithmetic mean.

We now look at various examples to understand how the probability of success for $x|\mathbf{g}$ differs between the two models. We set $K = 3$ classes. In Table 7.3, we see that when the class probabilities are relatively close together there is little difference between the two models.

However, in Table 7.4, when the class probabilities have much more spread, the differences between the two models are more noticeable.

Next, as in Erosheva (2005), consider a 2×2 table where X_1 and X_2 denote binary outcomes. The table has been normalized so that the sum of the cell entries $p_{lm} = P(X_1 = l, X_2 = m)$, $l, m = 1, 2$, add up to one. Let $\lambda_1 = p_{11} + p_{12} = P(X_1 = 1)$, $\lambda_2 = p_{11} + p_{21} = P(X_2 = 1)$ denote the corresponding marginal probabilities of positive responses. See Table 7.5. Further, set $K = 2$ classes with the grade of membership in the first of the two classes denoted by g and membership in the second quantified by $1 - g$. Let λ_1^1 and λ_1^2 denote the marginal probabilities for X_1 for classes 1 and 2 respectively. We use similar

Table 7.3: For $K = 3$ and $(p_1 = .4, p_2 = .5, p_3 = .6)$, we calculate the Bernoulli probability of a success for $x|\mathbf{g}$ under the mixed membership model (p_{mm}) and the partial membership model (p_{pm}) for varying $\mathbf{g}$.

	$\mathbf{g}$	p_{mm}	p_{pm}
1	(0.34,0.21,0.46)	0.5122	0.5124
2	(0.23,0.43,0.34)	0.5109	0.5111
3	(0.52,0.37,0.11)	0.4596	0.4591
4	(0.08,0.85,0.07)	0.4987	0.4987
5	(0.29,0.19,0.52)	0.5235	0.5238
6	(0.27,0.45,0.28)	0.5013	0.5013
7	(0.57,0.37,0.06)	0.4488	0.4483
8	(0.41,0.45,0.15)	0.4739	0.4736
9	(0.51,0.11,0.38)	0.4867	0.4865
10	(0.33,0.4,0.26)	0.4928	0.4928

notation for the marginals for X_2 . Under local independence, we have

$$p_{11} \cdot p_{22} = p_{12} \cdot p_{21}. \quad (7.17)$$

For the mixed membership model, we then have

$$\lambda_1(g) = g \cdot \lambda_1^1 + (1 - g) \cdot \lambda_1^2 \quad (7.18)$$

$$\lambda_2(g) = g \cdot \lambda_2^1 + (1 - g) \cdot \lambda_2^2 \quad (7.19)$$

for the marginal probabilities conditional on g and

$$p_{12}(g) = (g\lambda_1^1 + (1 - g)\lambda_1^2) \cdot (g(1 - \lambda_2^1) + (1 - g)(1 - \lambda_2^2)) \quad (7.20)$$

$$p_{21}(g) = (g(1 - \lambda_1^1) + (1 - g)(1 - \lambda_1^2)) \cdot (g\lambda_2^1 + (1 - g)\lambda_2^2) \quad (7.21)$$

$$p_{22}(g) = (g(1 - \lambda_1^1) + (1 - g)(1 - \lambda_1^2)) \cdot (g(1 - \lambda_2^1) + (1 - g)(1 - \lambda_2^2)) \quad (7.22)$$

Table 7.4: For $K = 3$ and $(p_1 = .02, p_2 = .15, p_3 = .90)$, we calculate the Bernoulli probability of a success for $x|\mathbf{g}$ under the mixed membership model (p_{mm}) and the partial membership model (p_{pm}) for varying $\mathbf{g}$.

	$\mathbf{g}$	p_{mm}	p_{pm}
1	(0.34,0.21,0.46)	0.4497	0.341
2	(0.23,0.43,0.34)	0.3761	0.2906
3	(0.52,0.37,0.11)	0.1673	0.0826
4	(0.08,0.85,0.07)	0.1908	0.1624
5	(0.29,0.19,0.52)	0.5028	0.4241
6	(0.27,0.45,0.28)	0.3249	0.2296
7	(0.57,0.37,0.06)	0.1192	0.0609
8	(0.41,0.45,0.15)	0.2062	0.1151
9	(0.51,0.11,0.38)	0.3681	0.2061
10	(0.33,0.4,0.26)	0.3039	0.1944

Table 7.5: Notation for cell and marginal probabilities for two binary variables, X_1 and X_2 .

	X_2		
X_1	p_{11}	p_{12}	λ_1
	p_{21}	p_{22}	$1 - \lambda_1$
	λ_2	$1 - \lambda_2$	1

for the cell probabilities conditional on g .

We now compare these results with the corresponding results from the partial membership models of Heller et al. (2008). The marginal probabilities conditional on g take the

form,

$$\lambda_1(g) = \frac{1}{c_1}(\lambda_1^1)^g(\lambda_1^2)^{(1-g)} \quad (7.23)$$

$$= \left(\frac{\lambda_1^1}{c_1}\right)^g \left(\frac{\lambda_1^2}{c_1}\right)^{(1-g)} \quad (7.24)$$

$$\lambda_2(g) = \frac{1}{c_2}(\lambda_2^1)^g(\lambda_2^2)^{(1-g)} \quad (7.25)$$

$$= \left(\frac{\lambda_2^1}{c_2}\right)^g \left(\frac{\lambda_2^2}{c_2}\right)^{(1-g)} \quad (7.26)$$

and the cell probabilities under the local independence assumption are

$$p_{12}(g) = \frac{1}{c_1}(\lambda_1^1)^g(\lambda_1^2)^{(1-g)} \frac{1}{c_2}(1 - \lambda_2^1)^g(1 - \lambda_2^2)^{(1-g)} \quad (7.27)$$

$$= \left(\frac{\lambda_1^1(1 - \lambda_2^1)}{c_1 c_2}\right)^g \left(\frac{\lambda_1^2(1 - \lambda_2^2)}{c_1 c_2}\right)^{(1-g)} \quad (7.28)$$

$$p_{21}(g) = \frac{1}{c_1}(1 - \lambda_1^1)^g(1 - \lambda_1^2)^{(1-g)} \frac{1}{c_2}(\lambda_2^1)^g(\lambda_2^2)^{(1-g)} \quad (7.29)$$

$$= \left(\frac{(1 - \lambda_1^1)\lambda_2^1}{c_1 c_2}\right)^g \left(\frac{(1 - \lambda_1^2)\lambda_2^2}{c_1 c_2}\right)^{(1-g)} \quad (7.30)$$

$$p_{22}(g) = \frac{1}{c_1}(1 - \lambda_1^1)^g(1 - \lambda_1^2)^{(1-g)} \frac{1}{c_2}(1 - \lambda_2^1)^g(1 - \lambda_2^2)^{(1-g)} \quad (7.31)$$

$$= \left(\frac{(1 - \lambda_1^1)(1 - \lambda_2^1)}{c_1 c_2}\right)^g \left(\frac{(1 - \lambda_1^2)(1 - \lambda_2^2)}{c_1 c_2}\right)^{(1-g)}. \quad (7.32)$$

We plotted the marginal probabilities for different λ while varying g from 0 to 1. We started with the $\lambda_1^1 = 0.1, \lambda_1^2 = 0.8, \lambda_2^1 = 0.3, \lambda_2^2 = 0.6$, the values used in Erosheva (2005), and reduced λ_1^1 to 0 and increased λ_2^2 to 1 while holding λ_1^2, λ_2^1 constant. Table 7.6 presents the λ values used to generate the marginal probability plots.

Figure 7.7 presents the marginal probability plots for these scenarios with Scenario 1 on the far left and Scenario 5 on the far right. The green points are the BPM values for λ_1 and λ_2 for varying values of g whereas the red points are the corresponding values for the MM model. For Scenario 1, we see that the BPM points produce a nonlinear path but, for these values of the λ 's, appears to reasonably approximate the MM model. As λ_{11} decreased and λ_{22} increased over the scenarios, the paths of points increasingly diverge. For scenario 5, the BPM model produces points that take only three pairs of values, sitting at corner values. At $g = 0$ and $g = 1$, the BPM produces λ values equivalent to the MM model. For values

Scenario	λ_1^1	λ_1^2	λ_2^1	λ_2^2
1	0.1	0.8	0.3	0.6
2	0.05	0.8	0.3	0.95
3	0.01	0.8	0.3	0.99
4	0.001	0.8	0.3	0.999
5	0	0.8	0.3	1

Table 7.6: λ values used to generate marginal probability plots in Figure 7.7.

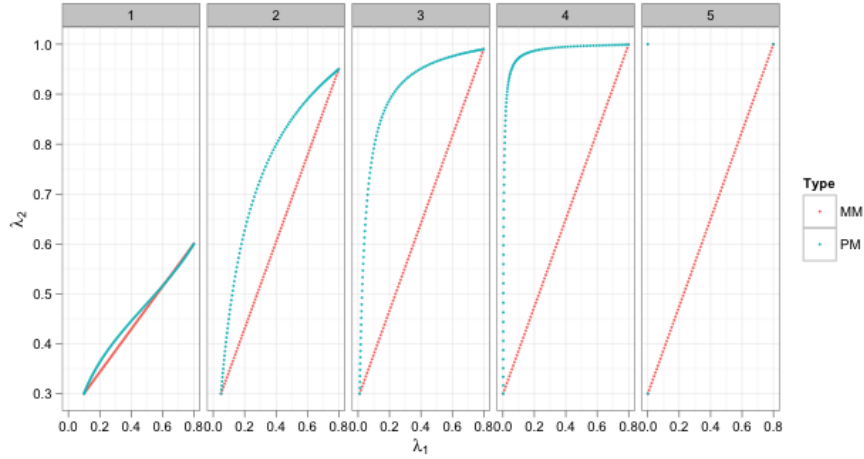


Figure 7.7: Marginal Probability Plots for Scenarios 1-5 in Table 7.6. Scenario 1 is the leftmost plot and Scenario 5 is the rightmost.

of g where $0 < g < 1$ in scenario 5, $\lambda_1 = 0$ and $\lambda_2 = 1$ under the BPM model. In scenarios where one of the class conditional response probabilities equals 1, any partial membership in the class implies that that individual's probability for that outcome must be 1. Similarly, when one of the class conditional response probabilities equals 0, any partial membership in that class implies that the probability for that outcome must be 0. We do not observe this property in the mixed membership model. Moreover, as one of the class probabilities

decreases to 0 or increases to 1, the marginal probabilities under the BPM and MM models increasingly diverge and shown in Figure 7.7.

In this section, we compared the behavior of the MM and BPM models for continuous and discrete distributions. In the case of the continuous distributions, we saw that for certain parameter values the MM model generates a very distinct pattern of clustered data. In the case of binary data, we saw that for class probabilities close to 0 or 1 the BPM model would exhibit its own distinct behavior. It may be helpful to keep these distinctions in mind when considering these models for different applications.

7.4 Semiparametric Correlated Bayesian Partial Membership Model: Derivation and Estimation Algorithms

We extend the BPM model in two ways. First, one limitation of this model is its inability to flexibly accommodate correlations among an observational unit's membership in the classes. The Dirichlet prior induces a small negative correlation among the class memberships in individuals. Blei and Lafferty (2007) addressed this shortcoming in mixed membership topic models by replacing the Dirichlet prior for individual membership scores with a logistic normal prior. Under this model, draws from the multivariate normal prior are then transformed to map to the probability simplex so that the values are positive and constrained to add to 1,

$$\boldsymbol{\eta}_{\mathbf{g}_i} \sim \mathcal{N}(\boldsymbol{\rho}, \boldsymbol{\Sigma}), \quad (7.33)$$

$$g_{ik} = \frac{\exp(\eta_{g_{ik}})}{\sum_l \exp(\eta_{g_{il}})}. \quad (7.34)$$

Because of the constraints that $\sum_k g_{ik} = 1$, we fix the K th element of $\boldsymbol{\eta}_{\mathbf{g}_i}$ to 0 so that the vector contains only $K - 1$ free elements and $\boldsymbol{\rho}$ and $\boldsymbol{\Sigma}$ have dimension $K - 1$ and $(K - 1) \times (K - 1)$ respectively. Atchison and Shen (1980) discuss properties and uses of the logistic normal, including a comparison with the Dirichlet distribution. They suggest that the logistic normal can suitably approximate the Dirichlet distribution so that little, if anything, would be lost if we applied the logistic normal in cases where a Dirichlet prior would be appropriate.

Secondly, while the use of exponential family distributions in the BPM model allows for

great flexibility in the modeling of outcomes, the need to choose the parametric distributions appropriate for different outcomes can be a nuisance and may lead to model misspecification as has been discussed in Chapters 4-6. To be able to handle non-nominal outcomes of an arbitrary type, we again turn to the extended rank likelihood method introduced by Hoff (2007). As has been discussed in previous chapters, multivariate sets of mixed outcomes are pervasive and a method to accommodate mixed outcomes requiring little specification on the part of the practitioner is useful, particularly in the case of the BPM when the focus is on the membership scores, the number of classes, or differences between the classes. With these modifications to the original BPM, we propose a semiparametric correlated partial membership model.

To formalize the model, consider a multivariate set of p outcomes for n observational units with each outcome denoted as before by $y_{ij}, i = 1, \dots, n, j = 1, \dots, p$. Initially, we focus on a model where the variance of the latent responses does not vary by class and the outcomes are conditionally independent given the class memberships and the class-specific means. Incorporating the modifications proposed above, the semiparametric correlated BPM is constructed as follows,

$$y_{ij} = h_j(z_{ij}), \quad (7.35)$$

$$z_{ij} | \boldsymbol{\mu}_j, \mathbf{g}_i \sim N(\mathbf{g}_i^T \boldsymbol{\mu}_j, 1), \quad (7.36)$$

$$g_{ik} = \frac{\exp(\eta_{g_{ik}})}{\sum_l \exp(\eta_{g_{il}})}, \quad (7.37)$$

$$\boldsymbol{\eta}_{\mathbf{g}_i} \sim N(\boldsymbol{\rho}, \boldsymbol{\Sigma}), \quad (7.38)$$

$$\boldsymbol{\rho} \sim N(\mathbf{m}_\rho, \mathbf{S}_\rho), \quad (7.39)$$

$$\boldsymbol{\Sigma} \sim \text{Inv. Wish}(\nu_\Sigma, \mathbf{S}_\Sigma), \quad (7.40)$$

$$\boldsymbol{\mu}_j \sim N(\mathbf{m}_\mu, \mathbf{S}_\mu), \quad (7.41)$$

where $h_j(\cdot)$ is an unspecified monotone transformation, $\boldsymbol{\mu}_j = (\mu_{1j}, \dots, \mu_{jK})$ are class specific mean parameters and $\mathbf{g}_i = (g_{i1}, \dots, g_{iK})$. Here, $\boldsymbol{\eta}_{\mathbf{g}_i} = (\eta_{g_{i1}}, \dots, \eta_{g_{i(K-1)}})$ is a vector of membership scores prior to transformation, $\boldsymbol{\rho} = (\rho_1, \dots, \rho_{K-1})$ is vector representing the mean membership scores prior to transformation, and $\boldsymbol{\Sigma}$ is a $(K-1) \times (K-1)$ covariance matrix for pre-transformation membership scores. Again, we designate $\boldsymbol{\eta}_{\mathbf{g}_i}$ as length $K-1$

instead of length K as η_{iK} is set to 0 to ensure identifiability.

Equation (7.37) defines the transformation of the multivariate normal membership scores $\boldsymbol{\eta}_{\mathbf{g}_i}$ for individual i to the constrained vector $\mathbf{g}_i$. Equations (7.38) and (7.41) define the priors for $\boldsymbol{\eta}_{\mathbf{g}_i}$ and $\boldsymbol{\mu}_j$ while equations (7.39) and (7.40) define hyperpriors for the parameters in equation (7.38).

Note that the variance in equation (7.36) is set to 1. As discussed in Chapter 5, the latent responses do not have a fixed location or scale. As a result, we fix the variance of the latent responses conditional on $\mathbf{G}$ and $\boldsymbol{\mu}_j$ for each outcome to 1. Similarly, we must fix the location in some manner. We elect to fix the class-specific means for the 1st class, $\mu_{j1} = 0, j = 1, \dots, p$.

Estimates of M are on the latent response scale and cannot be directly translated to the observed data scale. The vector $\boldsymbol{\mu}_j$ does impart information about the ordering of the class-specific means on the observed data scale. If we define the class-specific mean on the observed data scale for outcome j and class k as $E[y_{ij}|g_{ik} = 1]$, we could use the posterior predictive distribution to estimate this quantity. Specifically, to estimate means for class k , we make posterior predictions for the case where $g_{ik} = 1, g_{il} = 0$ for $l \neq k$. We make these predictions for each MCMC draw of parameter values and average over the predictions. To make posterior predictions from the semiparametric correlated partial membership model, we rely on a similar procedure to that discussed in Section 5.2.

7.4.1 Estimation of Correlated Partial Membership Model for Normally Distributed Outcomes

Before covering estimation of the semiparametric correlated partial membership model, we first examine the case of the correlated partial membership model for normally distributed outcomes with all class variances set to 1. Let $\boldsymbol{\Omega} = \{\mathbf{m}_\rho, \mathbf{S}_\rho, \nu_\Sigma, \mathbf{S}_\Sigma, \mathbf{m}_\mu, \mathbf{S}_\mu\}$ denote all hyperparameters in the BPM and let $\boldsymbol{\Theta}$ denote all parameters for which we would like to approximate posterior distributions. The joint log probability of $\mathbf{Y}$ and $\boldsymbol{\Theta}$ conditional

upon $\mathbf{\Omega}, \mathbf{\Sigma}$ is

$$p(\mathbf{Y}, \mathbf{\Theta} | \mathbf{\Omega}, \mathbf{\Sigma}) = \prod_i \prod_j \left[(2\pi)^{-1/2} \exp \left(-\frac{1}{2} \left(y_{ij} - \sum_k g_{ik} \mu_{jk} \right)^2 \right) \right] \quad (7.42)$$

$$\cdot \prod_i \left[(2\pi)^{-(K-1)/2} |\mathbf{\Sigma}|^{-1/2} \exp \left(-\frac{1}{2} \left(\boldsymbol{\eta}_{g_i} - \boldsymbol{\rho} \right)^T \mathbf{\Sigma}^{-1} \left(\boldsymbol{\eta}_{g_i} - \boldsymbol{\rho} \right) \right) \right] \quad (7.43)$$

$$\cdot \prod_j \prod_k \left(2\pi s_{\mu_{jk}}^2 \right)^{-1/2} \exp \left(-\frac{1}{2s_{\mu_{jk}}^2} (\mu_{jk} - m_{\mu_{jk}})^2 \right) \quad (7.44)$$

$$\cdot \prod_k \left(2\pi s_{\rho_k}^2 \right)^{-1/2} \exp \left(-\frac{1}{2s_{\rho_k}^2} (\rho_k - m_{\rho_k})^2 \right). \quad (7.45)$$

As Heller et al. (2008) noted, all of the parameters in $\mathbf{\Theta}$ are continuous and moreover we may take the derivatives of the log of the above probability expression. As a result, the problem of Bayesian estimation for this model lends itself to Hybrid (Hamiltonian) Monte Carlo (HMC). HMC uses the derivative of the log joint probability to inform its proposals. As a result, in high dimensions, this algorithm may outperform more traditional algorithms such as MH or Gibbs sampling. For a thorough introduction to HMC, see Section 2.2 and Neal (2010).

We do not rely on HMC to draw $\mathbf{\Sigma}$ but rather draw $\mathbf{\Sigma}$ in a separate Gibbs step for the correlated partial membership model. Thus to sample $(\mathbf{\Theta}, \mathbf{\Sigma})$, we apply a Gibbs sampling algorithm where the first step involves sampling $\mathbf{\Theta}$ via HMC and then $\mathbf{\Sigma}$ from its full conditional distribution.

The log joint probability density of which we will take derivatives is

$$\log p(\mathbf{Y}, \mathbf{\Theta} | \mathbf{\Omega}, \mathbf{\Sigma}) = \sum_i \sum_j \left[-\frac{1}{2} \log(2\pi) - \frac{1}{2} \left(y_{ij} - \sum_k g_{ik} \mu_{jk} \right)^2 \right] \quad (7.46)$$

$$+ \sum_i \left[-\frac{K-1}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{\Sigma}| - \frac{1}{2} \left(\boldsymbol{\eta}_{g_i} - \boldsymbol{\rho} \right)^T \mathbf{\Sigma}^{-1} \left(\boldsymbol{\eta}_{g_i} - \boldsymbol{\rho} \right) \right] \quad (7.47)$$

$$+ \sum_j \sum_k \left[-\frac{1}{2} \log(2\pi s_{\mu_{jk}}^2) - \frac{1}{2s_{\mu_{jk}}^2} (\mu_{jk} - m_{\mu_{jk}})^2 \right] \quad (7.48)$$

$$+ \sum_k \left[-\frac{1}{2} \log(2\pi) - \frac{1}{2} \log s_{\rho_k}^2 - \frac{1}{2s_{\rho_k}^2} (\rho_k - m_{\rho_k})^2 \right]. \quad (7.49)$$

The derivative of the log joint probability with respect to μ_{jk} remains the same as above,

$$\frac{d \log p}{d \mu_{jk}} = \sum_i \left(y_{ij} - \sum_l g_{il} \mu_{jl} \right) g_{ik} - s_{\mu_{jk}}^{-2} (\mu_{jk} - m_{\mu_{jk}}). \quad (7.50)$$

The derivative of the log joint probability with respect to $\eta_{g_{ik}}$ is

$$\frac{d \log p}{d \eta_{g_{ik}}} = \sum_j \left(y_{ij} - \sum_l g_{il} \mu_{jl} \right) \sum_l \mu_{jl} \frac{d g_{il}}{d \eta_{g_{ik}}} - \left(\Sigma^{-1} (\boldsymbol{\eta}_{\mathbf{g}_i} - \boldsymbol{\rho}) \right)_k \quad (7.51)$$

where $(\Sigma^{-1} (\boldsymbol{\eta}_{\mathbf{g}_i} - \boldsymbol{\rho}))_k$ is the k -th element of the vector $\Sigma^{-1} (\boldsymbol{\eta}_{\mathbf{g}_i} - \boldsymbol{\rho})$. Finally, the derivative of the log joint probability with respect to ρ_k is

$$\frac{d \log p}{d \eta_{\rho_k}} = \left(\sum_i \Sigma^{-1} (\boldsymbol{\eta}_{\mathbf{g}_i} - \boldsymbol{\rho}) \right)_k - s_{\rho_k}^{-2} (\rho_k - m_{\rho_k}) \quad (7.52)$$

where, as above, the subscript k outside the parentheses denotes k -th element of the vector of values, $\Sigma^{-1} (\boldsymbol{\eta}_{\mathbf{g}_i} - \boldsymbol{\rho})$.

7.4.2 Estimation of Semiparametric Correlated Partial Membership Model

To make posterior draws from the semiparametric Bayesian partial membership model, we use a Gibbs sampling algorithm with three steps. Because the extended rank likelihood $\Pr(\mathbf{Z} \in D(\mathbf{Y}) | \mathbf{G}, \mathbf{M}, \boldsymbol{\rho}, \boldsymbol{\Sigma})$ involves a complicated integral, any expressions involving the extended rank likelihood will be difficult to compute. We avoid having to compute this integral by drawing from the joint posterior of $(\mathbf{Z}, \mathbf{G}, \mathbf{M}, \boldsymbol{\rho}, \boldsymbol{\Sigma})$ via Gibbs sampling. Given $\mathbf{Z} = \mathbf{z}$ and $\mathbf{Z} \in D(\mathbf{Y})$, the full conditional density of $\mathbf{G}, \mathbf{M}, \boldsymbol{\rho}$ can be written as

$$p(\mathbf{G}, \mathbf{M}, \boldsymbol{\rho} | \boldsymbol{\Sigma}, \mathbf{Z} = \mathbf{z}, \mathbf{Z} \in D(\mathbf{Y})) = p(\mathbf{G}, \mathbf{M}, \boldsymbol{\rho} | \boldsymbol{\Sigma}, \mathbf{Z} = \mathbf{z})$$

because the current draw values $\mathbf{Z} = \mathbf{z}$ imply $\mathbf{Z} \in D(\mathbf{Y})$. A similar simplification may be made with the full conditional density of $\boldsymbol{\Sigma}$. Given $\boldsymbol{\Sigma}, \mathbf{G}, \mathbf{M}, \boldsymbol{\rho}, \mathbf{Z} \in D(\mathbf{Y})$ and $\mathbf{Z}_{(-i)(-j)}$, the full conditional density of z_{ij} is proportional to a normal density with mean $\boldsymbol{\mu}_j^T \mathbf{g}_i$ and variance 1 that is restricted to the region specified by $D(\mathbf{Y})$.

Our Gibbs sampling algorithm with the HMC step proceeds as follows:

1. **Draw underlying latent response $\mathbf{Z}$.** For each i and j , sample z_{ij} from $p(z_{ij}|\mathbf{G}, \mathbf{M}, \boldsymbol{\rho}, \boldsymbol{\Sigma}, \mathbf{Z}_{(-i)(-j)}, \mathbf{Z} \in D(\mathbf{Y}))$. More specifically,

$$z_{ij}^{(m)} \sim \text{TN}_{(z_l, z_u)}(\boldsymbol{\mu}_j^T \mathbf{g}_i, 1) \quad (7.53)$$

where TN denotes truncated normal and z_l, z_u define the lower and upper truncation points,

$$z_l = \max\{z_{kj} : y_{kj} < y_{ij}\} \quad (7.54)$$

$$z_u = \min\{z_{kj} : y_{kj} > y_{ij}\}. \quad (7.55)$$

2. **Draw jointly class memberships, class means, membership means $\mathbf{G}, \mathbf{M}, \boldsymbol{\rho}$.**

In this step, we jointly sample $\mathbf{G}, \mathbf{M}, \boldsymbol{\rho}$ from $p(\mathbf{G}, \mathbf{M}, \boldsymbol{\rho}|\boldsymbol{\Sigma}, \mathbf{Z}, \boldsymbol{\Omega})$ using HMC. To do so, we use the log joint probability in equation (7.46) and the derivatives in equations (7.50), (7.51) and (7.52), replacing y_{ij} with z_{ij} .

3. **Draw membership correlation matrix $\boldsymbol{\Sigma}$.** Sample from $p(\boldsymbol{\Sigma}|\mathbf{Z}, \mathbf{G}, \mathbf{M}, \boldsymbol{\rho}, \boldsymbol{\Omega})$.

$$\boldsymbol{\Sigma} \sim \text{Inv. Wishart} \left(\nu_{\boldsymbol{\Sigma}} + n, S_{\boldsymbol{\Sigma}} + (\mathbf{H}_G - \mathbf{1}_n \boldsymbol{\rho}^T)^T (\mathbf{H}_G - \mathbf{1}_n \boldsymbol{\rho}^T) \right) \quad (7.56)$$

where $\mathbf{H}_G$ is an $n \times (K - 1)$ matrix with rows $\boldsymbol{\eta}_{g_i}$.

7.5 A Simulated Data Example

To test the semiparametric correlated partial membership model, we simulated parameters and data according to the model in equations (7.35)-(7.41) and subsequently fit the model to the simulated data. The only exceptions were the class mean and the population membership mean parameters which we manually set to well-dispersed values. We generated data for 400 individuals, 3 classes and 30 outcomes. For each outcome, we drew a random number of thresholds for each outcome that we used to discretize the latent responses into the “observed” ordinal responses. Table 7.7 displays summary statistics for the simulated individual class memberships. In each class, individuals have a wide range of membership. Class 1 is the dominant class, with the largest value for all quantities presented.

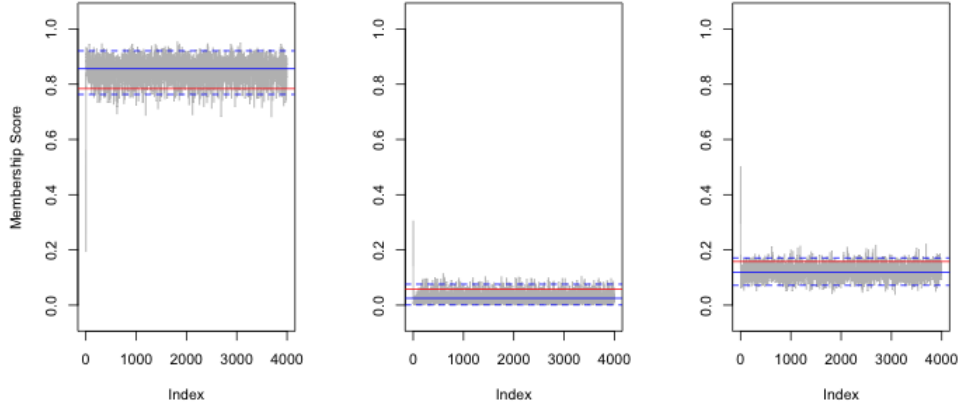
Table 7.7: Summary of simulated class memberships.

	Class 1	Class 2	Class 3
Mean	0.50343	0.29715	0.19942
Min	0.00044	0.00002	0.00042
1st Quartile	0.16749	0.03295	0.04661
Median	0.49960	0.17812	0.13424
3rd Quartile	0.84707	0.51746	0.28839
Max	0.99906	0.99521	0.88146

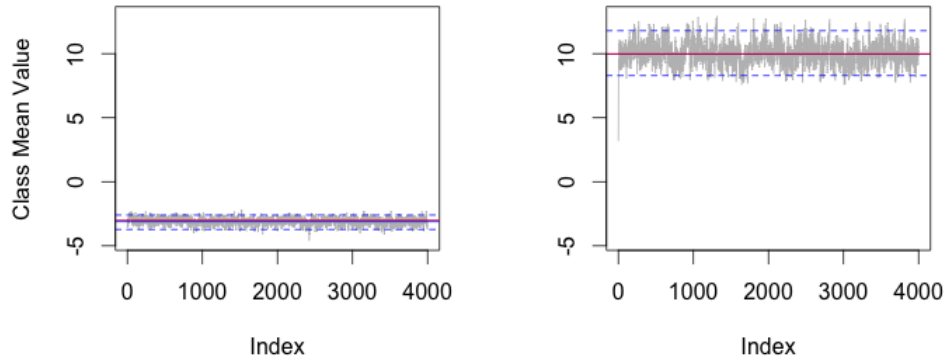
To fit the model, we drew 200,000 MCMC samples of the parameters, keeping every 50th draw. We discarded the first quarter of these retained samples as burn-in, resulting in 3,000 samples from the posterior distribution. We examined trace plots and used the Geweke (Geweke, 1992) and Raftery-Lewis (Raftery and Lewis, 1995) diagnostic tests to assess convergence. Figure 7.8 displays trace plots for the membership scores for a randomly selected individual and the class means for a random selected outcome. For each plot, the true value is displayed as the solid red line while the blue solid line represents the posterior mean and the blue dashed lines are the 2.5% and 97.5% posterior quantiles. In each case, we do not see any evidence of a lack of convergence and the data-generating parameters appear to be recovered.

Tables 7.8, 7.9 and 7.10 present summaries of the posterior distributions for $\mathbf{M}$, $\boldsymbol{\rho}$, and $\boldsymbol{\Sigma}$. Recall that the class means for the first class are restricted to 0 so only two columns are displayed. In general, one can see that the posterior means and medians are close to the data-generating values. The 95% posterior credible intervals contained the truth in 93.3% of the cases.

Based on the results summarized in Tables 7.9 and Tables 7.10, the model also recovers the untransformed population mean membership score, $\boldsymbol{\rho}$, and covariance, $\boldsymbol{\Sigma}$, from the logistic-normal prior well. Finally, although we do not present the results due to the number



(a) Membership Scores, Simulated Individual 272



(b) Class Means, Simulated Outcome 15

Figure 7.8: Trace plots for membership scores and class means for a randomly selected individual and outcome in the simulated data example. In each plot, the blue solid line marks the posterior mean while the dashed lines mark the 2.5% and 97.5% posterior quantiles. The red solid line represents the data generating value.

of parameters, the 95% posterior credible intervals for the membership scores, $\mathbf{G}$, contained the truth for 94.9% of the simulated membership scores.

We subsequently performed the posterior predictive model checks discussed in Section 2.4. To gain a better understanding of whether these model checks would offer any insight in the case where the number of classes was misspecified, we also fit two class ($K = 2$) and four class ($K = 4$) models. In fitting these model for the wrong number of classes, it should be noted that a much larger number of MCMC iterations (approximately 1,000,000) was required to attain convergence.

Figure 7.9 presents histograms for outcomes 14 and 19 for the three models $K = 2, 3, 4$. The three class ($K = 3$) and four class ($K = 4$) models appear to do a good job of replicating the observed marginal distributions for these outcomes and there is little to suggest that we could distinguish between these models based on these histograms. The two class model ($K = 2$) however does a poor job of approximating the observed marginal distributions. We chose these two outcomes in particular because the misfit in the two class ($K = 2$) model was particularly evident. However, these outcomes were hardly the only two outcomes for which misfit was evident. Meanwhile, the three class ($K = 3$) and four class ($K = 4$) models fit the other outcomes in a similar manner.

In Figure 7.10, we present the correlation distance and eigenvalue plot used as part of our posterior predictive model checking methods. The three class ($K = 3$) and four class ($K = 4$) models both appear to replicate the observed rank correlations well. Again, the two class ($K = 2$) model does not fit the observed data well. Thus, the posterior predictive model checks on which we have relied throughout appear (in this case) to help identify a model with two few classes. We see no differences in fit however in the three class ($K = 3$) and four class ($K = 4$) models using these diagnostics. A reasonable approach would be to default to the model with the fewest classes that fits the observed data well. In this simulated data case, the model estimates however helped suggest the three class ($K = 3$) model. For one of the elements of $\boldsymbol{\rho}$ in the four class ($K = 4$) model implied that the mean membership in that class was zero suggesting that the fourth class was unnecessary. In applications to real data, we do not necessarily expect the results to be so clear-cut.

Table 7.8: Posterior summaries for class means for classes 2 and 3.

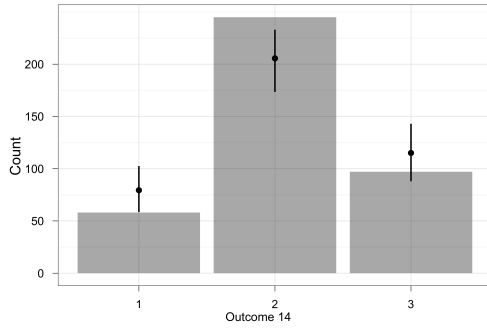
	Class 2			Class 3		
	Truth	Mean	Median	Truth	Mean	Median
1	-10	-9.695	-9.647	15	15.771	15.707
2	-3	-3.584	-3.582	3	3.558	3.555
3	-4	-4.130	-4.129	-14	-14.770	-14.741
4	6	5.685	5.672	-4	-4.053	-4.050
5	-3	-2.878	-2.870	-10	-10.031	-10.026
6	-10	-10.126	-10.093	-2	-1.895	-1.902
7	-12	-12.284	-12.212	-3	-2.688	-2.679
8	-12	-11.338	-11.311	13	13.277	13.208
9	-3	-3.097	-3.087	14	15.281	15.197
10	-11	-12.003	-11.986	10	11.045	11.018
11	-14	-14.795	-14.784	10	11.666	11.603
12	-13	-10.280	-10.256	10	8.730	8.715
13	-10	-9.755	-9.729	8	9.160	9.134
14	10	11.471	11.456	-9	-11.281	-11.151
15	-3	-3.121	-3.112	10	9.952	9.932
16	-11	-10.624	-10.616	10	10.144	10.107
17	10	8.483	8.455	4	2.950	2.951
18	-10	-10.529	-10.474	18	19.061	19.077
19	-10	-9.632	-9.610	10	10.047	10.011
20	-3	-2.756	-2.752	-10	-9.393	-9.372
21	3	3.150	3.148	-3	-3.482	-3.475
22	10	10.134	10.082	-10	-11.910	-11.881
23	11	9.800	9.719	-10	-10.379	-10.277
24	-15	-13.672	-13.647	11	10.627	10.575
25	7	6.265	6.230	-9	-9.879	-9.840
26	-11	-11.781	-11.729	15	17.378	17.367
27	-11	-11.439	-11.431	7	7.927	7.894
28	3	2.739	2.734	-3	-3.204	-3.185
29	-4	-3.638	-3.643	4	4.227	4.214
30	-10	-10.732	-10.683	10	11.796	11.742

Table 7.9: Posterior summary for membership score means, $\boldsymbol{\rho}$.

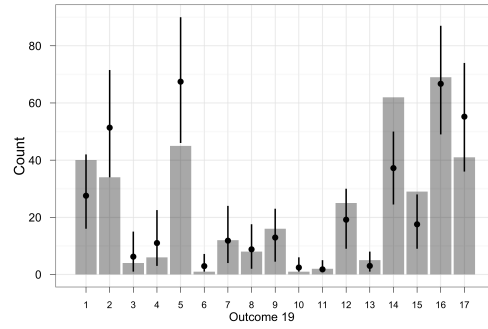
	Truth	Post. Mean	Post. Median	Post. 95% CI
ρ_1	1.000	0.904	0.900	(0.597, 1.236)
ρ_2	-0.100	-0.014	-0.011	(-0.288, 0.265)

Table 7.10: Posterior summary for membership score covariances, $\boldsymbol{\Sigma}$.

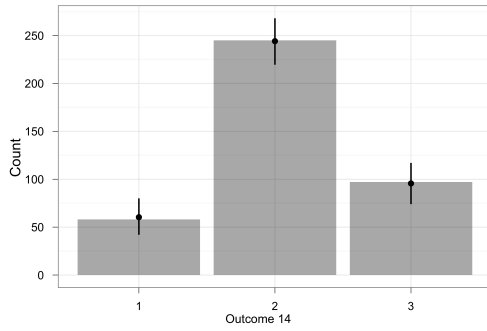
	Truth	Post. Mean	Post. Median	Post. 95% CI
$\Sigma_{1,1}$	4.147	5.195	5.128	(3.649, 7.069)
$\Sigma_{2,1}$	-0.379	-1.053	-1.051	(-1.805, -0.345)
$\Sigma_{1,2}$	-0.379	-1.053	-1.051	(-1.805, -0.345)
$\Sigma_{2,2}$	3.717	3.355	3.302	(2.385, 4.587)



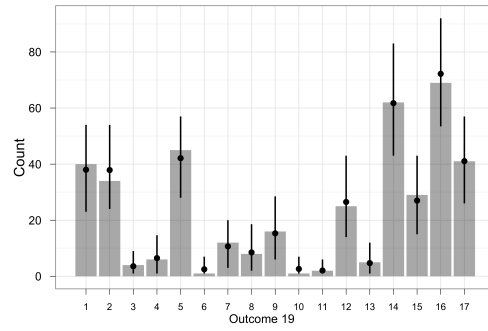
(a) Outcome 14, K=2



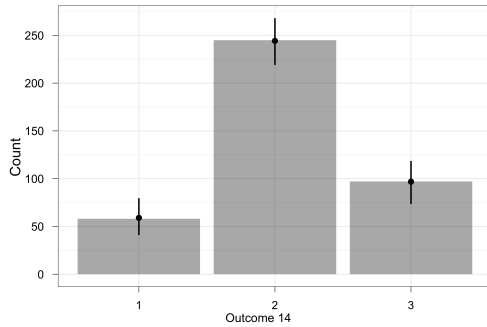
(b) Outcome 19, K=2



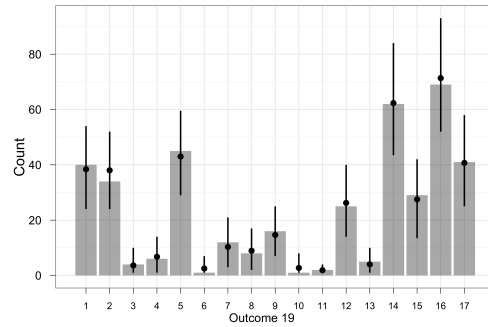
(c) Outcome 14, K=3



(d) Outcome 19, K=3



(e) Outcome 14, K=4



(f) Outcome 19, K=4

Figure 7.9: Histograms of the observed scores for simulated outcomes 14 and 19. The black points indicate the mean count across replicated datasets for each score. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.

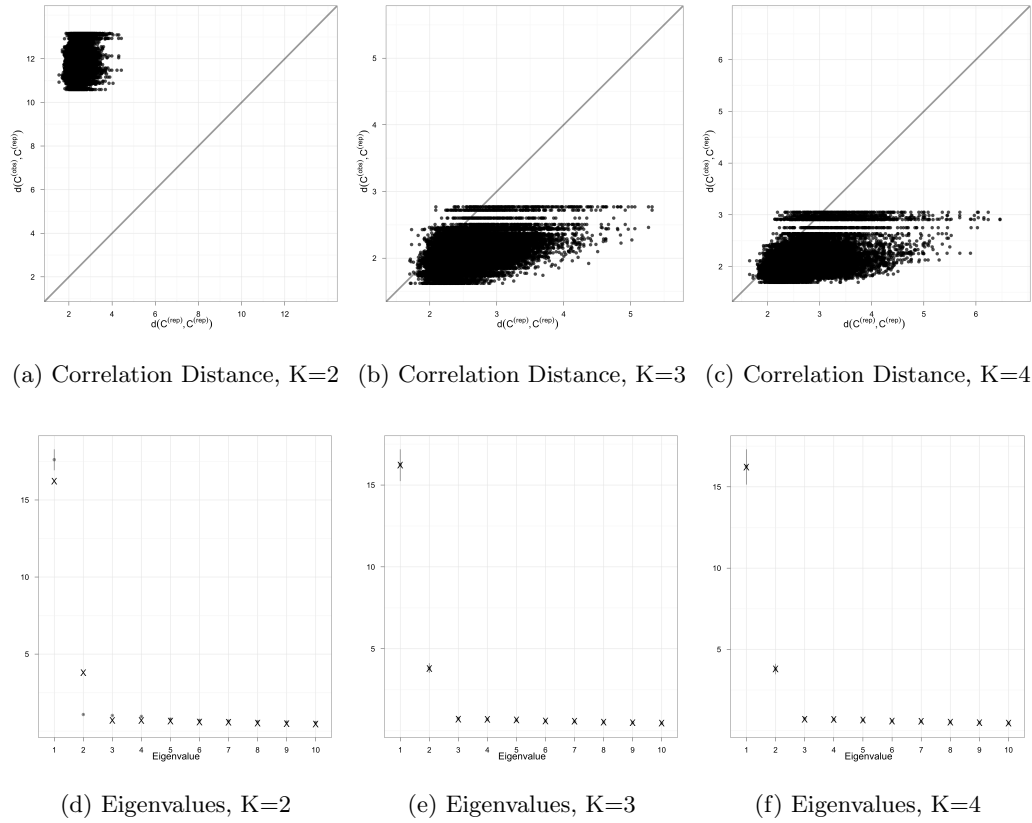


Figure 7.10: Correlation distance and eigenvalue plots for varying K for simulated data under the BPM model. The top row plots present scatterplots of $d_{sld}(C^{obs}, C^{rep,m})$ versus $d_{sld}(C^{rep,m}, C^{rep,m'})$ for all replicated datasets. The grey line represents the 45 degree line. The bottom row plots display the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) of the largest ten eigenvalues calculated using replicated data. Eigenvalues computed from the observed data are denoted by a black “X”.

7.6 Application to 2010-11 NBA Player Data

In the New York Times basketball blog *Off The Dribble* on February 28, 2012, Joshua Brustein highlighted some of the NBA-related research being presented at the increasingly visible MIT Sloan Sports Analytics Conference. Team chemistry and composition appeared to be recurring themes in the research with the intent of understanding how chemistry and composition might relate to winning. In understanding team composition, comparing it across teams and ultimately relating it to game outcomes, it is helpful to be able to group players by playing style and/or ability. Currently, basketball players are typically assigned to one of five positions: point guard (PG), shooting guard (SG), small forward (SF), power forward (PF) and center (C). NBA management and observers may commonly use a more informal typing of players with 3 categories that consolidates the above positions by physical attributes and function on the court: point guard, wings (shooting guards and small forwards), and bigs (power forwards and centers). However, current positions and player assignments to those positions may not fully reflect the variety of playing styles (Lutz, 2012). One approach to identifying positions for players would be to use a finite mixture model to perform a cluster analysis on players based on their season statistics.

Rather than assign players specifically to one playing style or another or identify classes that are themselves mixtures of more extreme classes, we would argue that players themselves are compositions of different extreme playing styles. For instance, the term “combo guard” is regularly used to describe a player who combines the skills and playing style of a point guard and shooting guard.

To apply the semiparametric correlated partial membership model to identify playing styles and player membership in those styles, we used 13 different statistics from the 2010-11 NBA season. We collected the data from hoopdata.com (Hoopdata, 2012). The statistics included minutes played per game, percent of made field goals that are assisted, assist rate, turnover rate, offensive rebound rate, defensive rebound rate, steals per 40 minutes, blocks per 40 minutes, and the number of shots attempted per 40 minutes at each of the following locations: at the rim, from 3-9 feet, from 10-15 feet, from 16-23 feet, and beyond the 3-point line. These statistics overlap with 13 of the 14 used in a cluster analysis by Lutz (2012).

Table 7.11: Variables, abbreviations and formulas.

Variable	Abbreviation	Formula (if calculated)
Minutes played per game (minimum of 10)	Min	-
Percent of made field goals that are assisted	% Ast made	$\frac{\text{field goals that are assisted}}{\text{total made field goals}}$
Assist Ratio	AR	$\frac{\text{Assists} \times 100}{\text{FGA} + (\text{FTA} \times .44) + \text{Turnovers}}$
Turnover Ratio	TOR	$\frac{\text{Turnovers} \times 100}{\text{FGA} + (\text{FTA} \times .44) + \text{Turnovers}}$
Offensive Rebound Rate	ORR	$\frac{100 \times (\text{Player ORebs} \times (\text{Team Min}/5))}{(\text{Player Min} \times (\text{Team ORebs} + \text{Opp DRebs}))}$
Defensive Rebound Rate	DRR	$\frac{100 \times (\text{Player DRebs} \times (\text{Team Min}/5))}{(\text{Player Min} \times (\text{Team DRebs} + \text{Opp ORebs}))}$
Attempted field goals at the rim per 40 minutes	Rim	-
Attempted field goals from 3-9 feet per 40 minutes	3-9	-
Attempted field goals from 10-15 feet per 40 minutes	10-15	-
Attempted field goals from 16-23 feet per 40 minutes	16-23	-
3-point field goals attempted per 40 minutes	3s	-
Steals per 40 minutes	Stls	-
Blocks per 40 minutes	Blks	-

Table 7.11 presented as in Lutz (2012) lists these variables, their abbreviations and formulas of calculated statistics.

We included only players who had played 30 or more games and averaged 10 or more minutes per game. This restriction limited the number of players in the sample to 332. The statistics are continuous but some, such as minutes played per game (maximum of 48) or percent of made field goals (0-100), are restricted in their range. By relying on the orderings of the outcomes to make inference, the semiparametric correlated partial membership model can readily handle these different types of outcomes.

We considered models with different numbers of classes ($K = 3, 4, 5, 6$). The four class, five class and six class ($K = 4, 5, 6$) models produced similar results in that four of the classes in the five and six class models ($K = 5, 6$) bore a strong resemblance to the four in the four class ($K = 4$) model. Although the five and six class ($K = 5, 6$) models appeared to fit the data better than the four class ($K = 4$) model, both of these models continued to show signs of a lack of convergence even after a large number of iterations. The chains for the

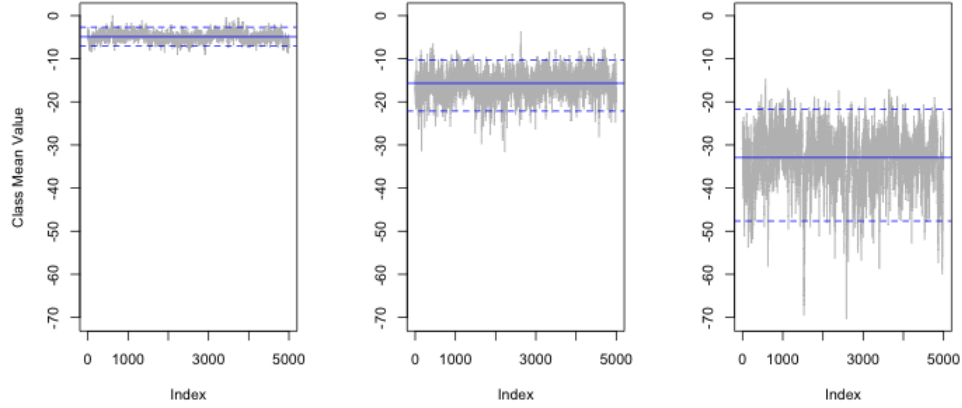
membership scores appeared relatively stable but the chains for the class mean parameters displayed a lack of convergence for the five and six class ($K = 5, 6$) models. The four class ($K = 4$) model, however, did not show signs of a lack of convergence according to trace plots and the Geweke diagnostic test (Geweke, 1992) and still showed reasonable fit according to posterior predictive model checks. Thus, we present here the results for the four class ($K = 4$) model. Figure 7.11 displays trace plots for the class means for a randomly selected statistic, the field goals attempted from 10-15 feet per 40 minutes, and trace plots for the membership scores for a randomly selected player, Kirk Hinrich. Both sets of trace plots are representative of the trace plots for the class means of the other statistics and for the membership scores of the other players.

We used the posterior predictive model checking methods introduced in Section 2.4 to evaluate model fit. Figure 7.12 displays histograms of the observed data for turnover rate and field goals attempted at the rim per 40 minutes as well as black points and intervals indicating the posterior predictive means and 95% posterior predictive intervals. In general, the model appeared to approximate the marginal distributions reasonably well. The model approximated the marginal distributions in similar fashion for the other outcomes not shown here.

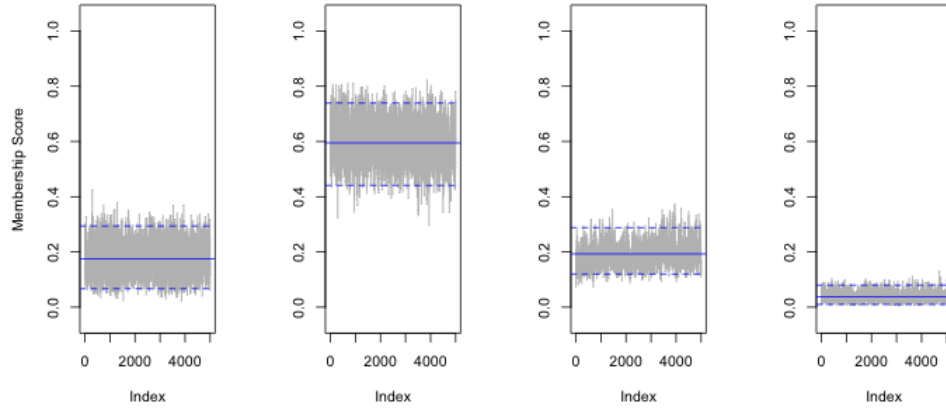
We proceeded to check the ability of the model to replicate the observed rank correlations using the plots in Figure 7.13. In the correlation distance plot, there appeared to be some lack of fit. In the eigenvalue plot, the first 5 eigenvalues were reasonably well fit but the 95% posterior predictive interval for the 6th eigenvalue did not contain the observed value. Plots of the pairwise rank correlations for individual outcomes, however, showed generally good fit. Preliminary¹ posterior predictive model checks for a five class ($K = 5$) model showed generally superior fit for the eigenvalue and correlation distance plots but the differences were fairly small for the marginal distribution and pairwise correlation plots. We do not present the results for the five class model, however, due to some evidence of a lack of convergence for some of the class mean parameters.

Satisfied with the model fit, we analyzed the model results. Table 7.12 presents the

¹“Preliminary” because the posterior predictive checks were based on values from chains that did not all show convergence.



(a) Class Means, FGA from 10-15 ft. per 40 min.



(b) Membership Scores, Kirk Hinrich

Figure 7.11: Trace plots for class means and membership scores for field goal attempted (FGA) from 10-15 feet per 40 minutes and Kirk Hinrich in the NBA player data example. In each plot, the blue solid line marks the posterior mean while the dashed lines mark the 2.5% and 97.5% posterior quantiles.

posterior mean for $\mathbf{M}$. Recall that the means are on the latent response scale and that the first column was fixed to 0 to set the location of the latent responses. As a result,

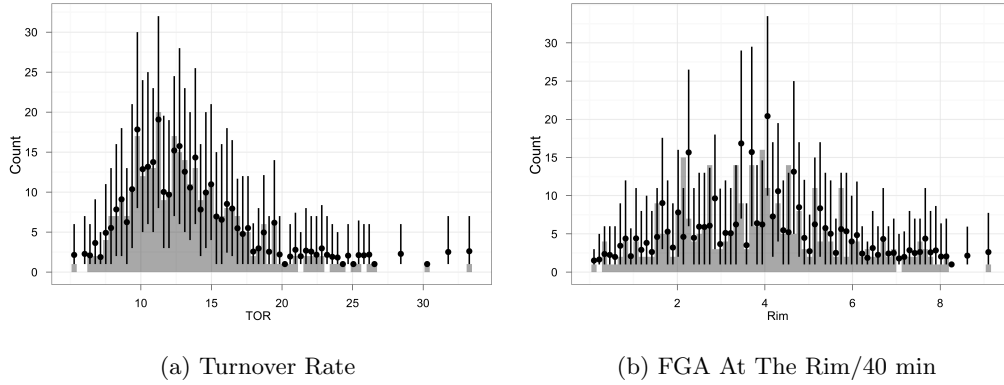


Figure 7.12: Histograms of the observed scores for the Turnover Rate and Field Goal Attempts (FGA) At The Rim per 40 minutes. The black points indicate the mean count across replicated datasets for each interval. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.

we focus on the ordering of the classes in Table 7.12. Looking across the classes, we were able to differentiate the types of players that the classes represent immediately. In no particular order, class 2 appeared to be a ball handler/point guard class with players who are characterized by high assist rates, higher than average turnover rates, very few assisted field goals and high steals per 40 minutes. Class 3 could reasonably be designated a spot-up shooter class characterized by players who primarily take shots beyond the three point line and for whom most of their field goals are assisted (meaning they do not create their own shots in contrast to class 2). Class 1 was characterized by players who shoot at all ranges inside the 3 point line and who also rebound and block shots fairly well. This class is typified by offensively well-rounded big men. Class 4 described players who rebound well (particularly on the offensive end of the floor), block shots, make steals and are limited in both their shooting range and minutes. This defensive big men class could just as well be called active or limited big men.

Tables 7.13-7.16 list the players with top ten membership scores in each class. In general, we believe it is reasonable to say that the qualitative assessments of the player by NBA

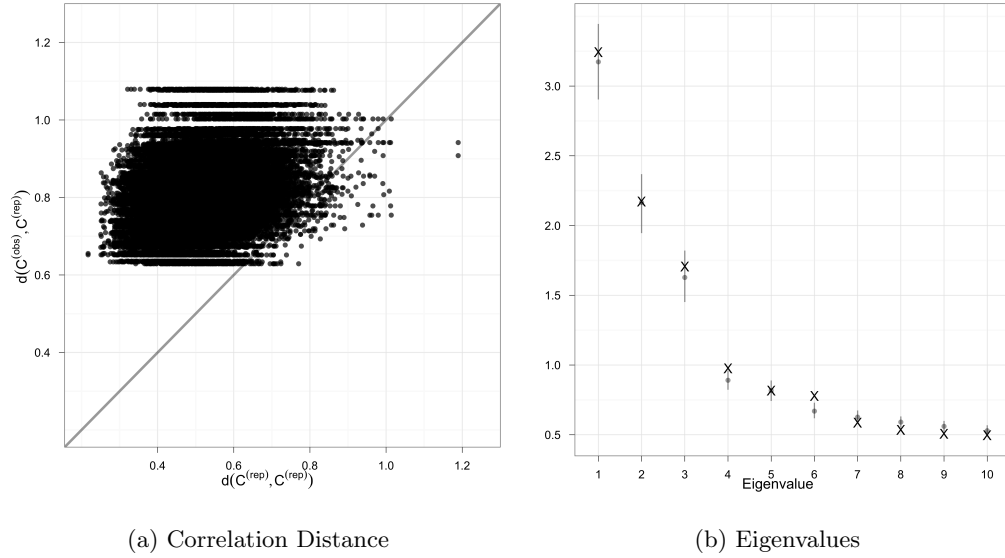


Figure 7.13: The left plot presents a scatterplot of $d_{sld}(C^{obs}, C^{rep,m})$ versus $d_{sld}(C^{rep,m}, C^{rep,m'})$ for all replicated datasets. The grey line represents the 45 degree line. On the right, a plot of top ten eigenvalues depicting the mean prediction (grey point) and 95% prediction intervals (grey line segment) of the eigenvalues calculated using replicated data. Eigenvalues computed from the observed data are denoted by a black “X”.

observers would agree quite well with the model-derived class that they typify. For instance, in class 2, it was comforting to see Steve Nash, considered by many to be the prototypical NBA point guard (at least offensively), at the top of the list. Similarly, James Jones, whose job in the Miami Heat offense in 2010-11 was largely limited to sitting at the three point line and shooting open shots created for him by his teammates, typified class 3. Class 1 was generally dominated by well-known big men with diverse offensive games such as Al Jefferson, Zach Randolph, Elton Brand, Tim Duncan and Paul Gasol. Class 4, the defensive big men, proved interesting in that, even for the highest ranking members in that class, the class 4 membership typically is not the largest class membership for that player.

Looking through Tables 7.13-7.16, however, we did not see the names of the players who are considered the league’s most dynamic stars, players such as LeBron James, Kobe Bryant,

Table 7.12: Posterior means for class means in the $K = 4$ model. Bolded figures indicate that the corresponding posterior 95% CI did not include 0.

	Class 1 (Well-Rounded Big Men)	Class 2 (Ball Handlers)	Class 3 (Spot-up Shooters)	Class 4 (Defensive Big Men)
Min	0.00	-0.53	-3.18	-7.11
% Ast	0.00	-11.65	27.78	-2.63
AR	0.00	7.19	0.97	6.83
TOR	0.00	5.61	-7.11	16.41
ORR	0.00	-6.69	-19.19	11.39
DRR	0.00	-5.59	-8.29	1.73
Rim	0.00	-0.99	-16.81	1.46
3-9	0.00	-3.27	-22.41	-13.66
10-15	0.00	-4.90	-15.69	-32.91
16-23	0.00	-2.94	-1.05	-22.35
3s	0.00	1.78	14.56	-7.86
Stls	0.00	2.99	-0.56	2.69
Blks	0.00	-4.29	-7.42	2.87

Dwyane Wade, Kevin Durant or Derrick Rose. Table 7.17 presents the membership scores for these players. The players’ “dynamic” style was evidenced by their significant membership in multiple classes. Although much larger than the typical point guard, LeBron James’ rare ability to handle the ball for a man of his size was reflected in his high membership in the Ball Handlers class. Yet he still maintained a significant membership in the Well-Rounded Big Men class. Kobe Bryant’s membership was almost perfectly split between the Well-Rounded Big Men class and the Ball Handlers class while Dwyane Wade’s membership scores flip the membership scores of LeBron James in the Well-Rounded Big Men and Ball Handlers classes. Kevin Durant’s versatility offensively was reflected in his non-trivial membership scores in the Well-Rounded Big Men, Ball Handlers and Spot-up Shooters

Table 7.13: Players with top ten membership scores for class 1 (Well-Rounded Big Men).

Player Name	Team	Pos	Class			
			1	2	3	4
Al Jefferson	UTH	C	0.788	0.057	0.077	0.079
Nazr Mohammed	CHA	C	0.756	0.028	0.074	0.142
Zach Randolph	MEM	PF	0.751	0.058	0.065	0.126
Chris Kaman	LAC	C	0.744	0.068	0.106	0.082
Brook Lopez	NJN	C	0.733	0.094	0.094	0.079
Elton Brand	PHI	PF	0.732	0.085	0.108	0.074
J.J. Hickson	CLE	PF	0.731	0.059	0.088	0.122
Tim Duncan	SAS	C	0.730	0.084	0.087	0.098
Dwight Howard	ORL	C	0.729	0.031	0.059	0.181
Pau Gasol	LAL	PF	0.727	0.085	0.085	0.103

classes.

We compared our results to some of the results from the cluster analysis by Lutz (2012) in one of the papers selected for the proceedings at the MIT Sloan Sports Analytics conference. Lutz (2012) sought to re-examine the concept of NBA player positions and determine the association between cluster membership and winning games during the regular season. Using a finite mixture of Gaussians, he identified 10 clusters reflecting these different playing styles (double the standard 5 positions to which players are typically assigned).

Lutz (2012) summarized these clusters in a table which we reproduced in Table 7.18. Note that some of the names of the clusters used by Lutz (2012) seem to suggest that certain groups are characterized by players who represent a mixture of playing styles: “combo” guards and “versatile” swingmen.

We considered the membership scores for players who typify the Combo Guard cluster of Lutz (2012). Table 7.19 presents the membership scores for these players as well. The

Table 7.14: Players with top ten membership scores for class 2 (Ball Handlers).

Player Name	Team	Pos	Class			
			1	2	3	4
Steve Nash	PHO	PG	0.085	0.805	0.095	0.016
Devin Harris	NJN	PG	0.094	0.764	0.122	0.020
Chris Paul	NOR	PG	0.108	0.763	0.111	0.018
T.J. Ford	IND	PG	0.120	0.756	0.098	0.025
Jonny Flynn	MIN	PG	0.083	0.753	0.138	0.026
Will Bynum	DET	PG	0.119	0.741	0.114	0.026
Tony Parker	SAS	PG	0.149	0.728	0.101	0.021
Jose Juan Barea	DAL	PG	0.139	0.727	0.107	0.026
Raymond Felton	NYK	PG	0.136	0.691	0.138	0.035
John Wall	WAS	PG	0.170	0.689	0.111	0.030

semiparametric correlated partial membership model appeared to view these players as combo guards as well, not by necessitating a separate class for these players but by reflecting their split membership primarily between classes 2 and 3, the Ball Handlers and Spot-up Shooters classes. In this sense, the semiparametric correlated partial membership model reduced the number of classes as compared to the cluster analysis of Lutz (2012) and reflected heterogeneity in the population by modeling the heterogeneity in the individual with partial memberships.

One feature of the semiparametric correlated partial membership model that we highlighted in Section 7.4 was the use of the logistic normal prior to flexibly model the correlations among the memberships. Although we have posterior draws for Σ , these values can not be directly translated into posterior correlations for the membership scores. Instead, for each posterior draw of $\mathbf{G}$, we calculated the correlations of the membership scores. We then took the mean of these correlations across the different draws of $\mathbf{G}$ to obtain estimates

Table 7.15: Players with top ten membership scores for class 3 (Spot-up Shooters).

Player Name	Team	Pos	Class			
			1	2	3	4
James Jones	MIA	SF	0.274	0.122	0.441	0.163
Brian Cardinal	DAL	PF	0.196	0.130	0.434	0.240
Daequan Cook	OKC	SG	0.262	0.187	0.413	0.138
James Posey	IND	SF	0.262	0.151	0.406	0.181
Keith Bogans	CHI	SG	0.238	0.171	0.394	0.197
Kyle Korver	CHI	SG	0.293	0.247	0.381	0.079
Eddie House	MIA	PG	0.265	0.300	0.362	0.073
Jodie Meeks	PHI	G	0.309	0.253	0.347	0.091
DeShawn Stevenson	DAL	SG	0.226	0.305	0.337	0.132
Raja Bell	UTH	SG	0.300	0.256	0.334	0.110

of the membership score correlations. The results are presented in Table 7.20.

In Table 7.20, there are moderate to large negative correlations between the Well-Rounded Big Men (class 1) and the Ball Handlers (class 2) and Spot-up Shooters (class 3) as well as between the Defensive Big Men (class 2) and the Ball Handlers (class 3). Well-Rounded Big Men (class 1) and Defensive Big Men (class 4) showed moderate positive correlation between them. The negative correlations between Ball Handlers (class 2) membership and Well-Rounded and Defensive Big Men (classes 1, 4) memberships are not unexpected as these playing styles tend to be mutually exclusive in players. The positive correlation between the Well-Rounded and Defensive Big Men (classes 1, 4) memberships also makes sense as many players tend to exhibit a mix of these playing styles.

Earlier we mentioned that the five and six class ($K = 5, 6$) models showed better model fit. We noted that four of the classes in the five and six class ($K = 5, 6$) models corresponded very closely to the four classes above. The fifth class in the five and six class models was

Table 7.16: Players with top ten membership scores for class 4 (Defensive Big Men).

Player Name	Team	Pos	Class			
			1	2	3	4
Joel Przybilla	POR	C	0.306	0.123	0.202	0.369
Omer Asik	CHI	C	0.435	0.060	0.136	0.369
Joey Dorsey	TOR	PF	0.488	0.051	0.106	0.356
DeAndre Jordan	LAC	C	0.496	0.052	0.132	0.320
Erick Dampier	MIA	C	0.434	0.099	0.179	0.288
Shaquille O'Neal	BOS	C	0.524	0.050	0.149	0.278
Andris Biedrins	GSW	C	0.549	0.061	0.114	0.276
Ben Wallace	DET	C	0.438	0.160	0.129	0.273
Semih Erden	BOS	C	0.465	0.079	0.184	0.272
Aaron Gray	NOR	C	0.593	0.051	0.090	0.265

Table 7.17: Membership scores for commonly regarded dynamic players

Player Name	Team	Pos	Class			
			1	2	3	4
Derrick Rose	CHI	PG	0.286	0.581	0.101	0.032
Dwyane Wade	MIA	SG	0.482	0.361	0.096	0.061
Kevin Durant	OKC	SF	0.515	0.264	0.175	0.046
Kobe Bryant	LAL	SG	0.424	0.429	0.118	0.029
LeBron James	MIA	SF	0.365	0.482	0.108	0.045

typified by big men with extended shooting range (in the 16-23 ft. range). The highest posterior mean membership in this fifth class was around 0.27. The sixth class in the six class ($K = 6$) model was a class with generally low membership; the highest posterior mean

Table 7.18: Cluster labels and typical members found in Lutz (2012).

Cluster	Cluster Label	Typical Members
Combo Guards	1	Steve Blake / Mario Chalmers / Rudy Fernandez
Backup Bigs	2	Deandre Jordan / Ben Wallace / Brendan Haywood
Skilled Swingmen	3	Shane Battier / Lamar Odom / Paul George
Floor Spacers	4	Channing Frye / Matt Bonner / Mike Miller
Elite Bigs	5	Amare Stoudemire / Elton Brand / Brook Lopez
Big Bodies	6	Jason Collins / Antonio McDyess / Kurt Thomas
Aggressive Bigs	7	Brandon Bass / Carlos Boozer / Tyler Hansbrough
Ball Handlers	8	Chris Paul / Kyle Lowery / Devin Harris
Perimeter Scorers	9	Rudy Gay / Dwyane Wade / Eric Gordon
Versatile Shooters	10	James Harden / Ray Allen / Nicolas Batum

Table 7.19: Membership scores for players typifying the Lutz (2012) combo guard cluster.

Player Name	Team	Pos	Class			
			1	2	3	4
Mario Chalmers	MIA	PG	0.178	0.475	0.268	0.079
Rudy Fernandez	POR	SG	0.238	0.407	0.258	0.097
Steve Blake	LAL	PG	0.200	0.377	0.311	0.112

membership in this class was only 0.044. The top ten members of this class did not seem to be well characterized other than perhaps by mid-sized players who have some tendency to shoot around the rim.

In this section, we applied the semiparametric correlated partial membership model to NBA player statistics from the 2010-11 season. The four class example served to demonstrate the features of the semiparametric correlated partial membership as well as provide

Table 7.20: Posterior mean of membership score correlations.

	1	2	3	4
1	1.000	-0.877	-0.438	0.395
2	-0.877	1.000	0.076	-0.678
3	-0.438	0.076	1.000	-0.074
4	0.395	-0.678	-0.074	1.000

sensible results consistent with more qualitative characterizations of player styles. We compared our results to those of a comparable cluster analysis (Lutz, 2012) and found that our model was able to reduce the number of classes identified by characterizing individuals as compositions of different playing styles.

7.7 Application to the SIVD Study

In this section, we apply the semiparametric correlated partial membership model to the SIVD executive functioning data introduced in Chapter 3. We began by applying a two class ($K = 2$) model to the data. A two class model would be reasonable when, for instance, one class was characterized by participants with normal or excellent executive functioning and a second class was typified by individuals with deficient executive functioning, with many individuals a composition of two. We made 200,000 draws, keeping every 40th draw and discarding the first half as burn-in. We used trace plots and the Geweke (Geweke, 1992) and Raftery-Lewis (Raftery and Lewis, 1995) diagnostic tests to assess convergence.

Figure 7.14 presents plots assessing the two class ($K = 2$) model's ability to fit the observed marginal distributions. In each case, the model did a good job of fitting the marginal distributions. We saw similarly good fit for the other outcomes.

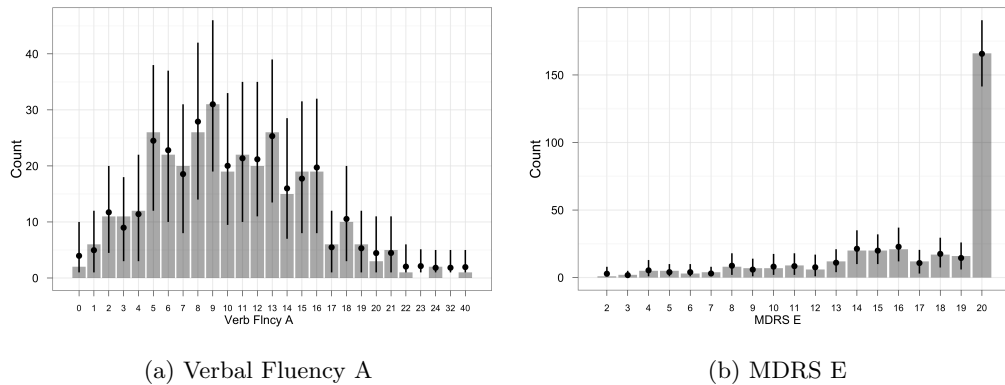


Figure 7.14: Histograms of the observed scores for the Verbal Fluency A and MDRS E. The black points indicate the mean count across replicated datasets for each score. The black vertical segment indicates the interval from the 2.5% to 97.5% quantiles across replicated datasets.

Figures 7.15(a) and 7.15(c) present the correlation distance and eigenvalue plots for the two class ($K = 2$) model. The model, much like the single factor semiparametric latent

variable model, did not approximate the observed rank correlations well. This misfit may suggest that we require more classes as we saw in the simulated data example in Section 7.5. However, in contrast to the simulated data example, we did not see evidence of misfit for the marginal distributions.

We subsequently examined the pairwise correlations. Figures 7.16(a) and 7.16(b) show the pairwise rank correlations for Mattis Dementia Rating J and Visual Span Backwards. Again, much like our results with the single factor semiparametric latent variable model, we see that the majority of pairwise correlations are well approximated but there are notable cases of misfit. For instance, we see misfit for the correlations of Mattis Dementia Rating Scale J with Mattis Dementia Rating Scale I and K. As noted in Section 5.6, these items are conceptually related. Likewise, for Visual Span Backwards, the correlation with Visual Span Forwards was not accurately approximated by the two class ($K = 2$) model.

To improve the fit of the model, we tried increasing the number of classes. A six class ($K = 6$) model was the model with the fewest number of classes for which we attained a seemingly sufficient fit to the observed rank correlations as can be seen in Figures 7.15(b), 7.15(d), 7.16(c) and 7.16(d). We do not present any of the posterior predictive checks for the marginal distributions as there was little to no difference from the corresponding posterior predictive checks for the two class ($K = 2$) model.

In Tables 7.21 and 7.22, posterior means for the class means, $\mathbf{M}$, are listed along with the 95% posterior credible intervals or, in the case of Table 7.22, an indication of whether the 95% posterior credible interval contained 0. Recall that the means for class 1 in both models are fixed to 0 and, as a result, are not presented in the tables. For the $K = 2$ model, the interpretation of the two classes is clear: class 1 is typified by individuals with high levels of executive functioning and class 2 is typified by people with very low levels of executive functioning. For the $K = 6$ model, the interpretation of the six classes is less clear.

It is possible that the six class ($K = 6$) model produces classes for which there is no obvious corresponding substantive meaning in an attempt to fit the data well. The inability to fit the data well with fewer classes may reflect the restrictions of the model as the covariances between outcomes are set to 0. The addition of more classes may aid the model

Table 7.21: Posterior summary for class means for class 2 in the $K = 2$ model.

	Class 2	Class 2	Class 2
	Post. Mean	Post. Median	Post. 95% CI
Digit Sp Fwd	-6.349	-6.303	(-8.815, -4.267)
Digit Sp Bwd	-8.210	-8.000	(-11.339, -5.878)
Visual Sp Fwd	-5.924	-5.817	(-8.488, -3.904)
Visual Sp Bwd	-6.942	-6.840	(-9.357, -4.818)
Verb Flncy F	-11.357	-11.179	(-15.343, -8.295)
Verb Flncy A	-14.122	-13.931	(-18.768, -10.141)
Verb Flncy S	-13.662	-13.497	(-18.329, -10.041)
MDRS E	-8.333	-8.139	(-11.523, -5.889)
MDRS G	-10.898	-10.486	(-17.067, -6.257)
MDRS H	-5.857	-5.698	(-9.964, -2.554)
MDRS I	-6.535	-6.422	(-10.193, -3.641)
MDRS J	-7.330	-7.058	(-12.453, -3.590)
MDRS K	-6.668	-6.374	(-12.161, -2.839)
MDRS L	-6.594	-6.525	(-9.493, -4.024)
MDRS O	-6.301	-6.125	(-10.132, -3.447)
MDRS V	-5.543	-5.440	(-8.069, -3.518)
MDRS W	-8.024	-7.833	(-11.357, -5.237)
MDRS X	-9.493	-9.484	(-13.241, -6.125)
MDRS Y	-17.615	-17.328	(-26.370, -10.743)

in representing the dependence structure in the data under the covariance restrictions. Given that the two class ($K = 2$) appeared to fit the marginal distributions well and that the model misfit occurred in approximated residual correlations among conceptually related items, it is conceivable that removing the assumption of independence among items conditional on

the membership scores and class means would result in a two class model that fits the data very well.

Table 7.22: Posterior means for class means for classes 2-6 in the $K = 6$ model. Bolded figures indicate that the corresponding posterior 95% CI did not include 0.

	Class 2	Class 3	Class 4	Class 5	Class 6
Digit Sp Fwd	-9.293	-10.255	-11.112	-8.347	0.209
Digit Sp Bwd	-5.793	-7.694	-12.859	-1.919	4.156
Visual Sp Fwd	-3.518	-1.760	-9.010	11.223	-4.579
Visual Sp Bwd	0.919	2.670	-9.819	17.176	0.013
Verb Flncy F	-4.052	11.692	0.407	3.190	24.470
Verb Flncy A	-14.654	0.728	-6.658	-1.630	16.053
Verb Flncy S	-19.676	3.247	-5.531	-1.375	23.824
MDRS E	-3.726	2.986	-5.826	9.603	0.985
MDRS G	-8.843	8.698	-5.303	1.427	0.178
MDRS H	-0.568	5.713	-0.732	10.893	5.617
MDRS I	11.820	-4.517	-15.140	2.056	4.742
MDRS J	12.365	-2.687	-17.668	-3.060	4.400
MDRS K	12.763	-4.008	-16.375	-1.177	1.579
MDRS L	1.581	3.175	-7.203	12.236	0.422
MDRS O	0.126	1.228	-6.396	9.300	0.943
MDRS V	2.926	4.453	-2.268	10.472	7.269
MDRS W	-3.084	18.386	-16.560	-10.017	-9.919
MDRS X	-6.867	18.379	-13.795	-14.612	-6.850
MDRS Y	-13.690	13.626	-11.931	4.959	-0.843

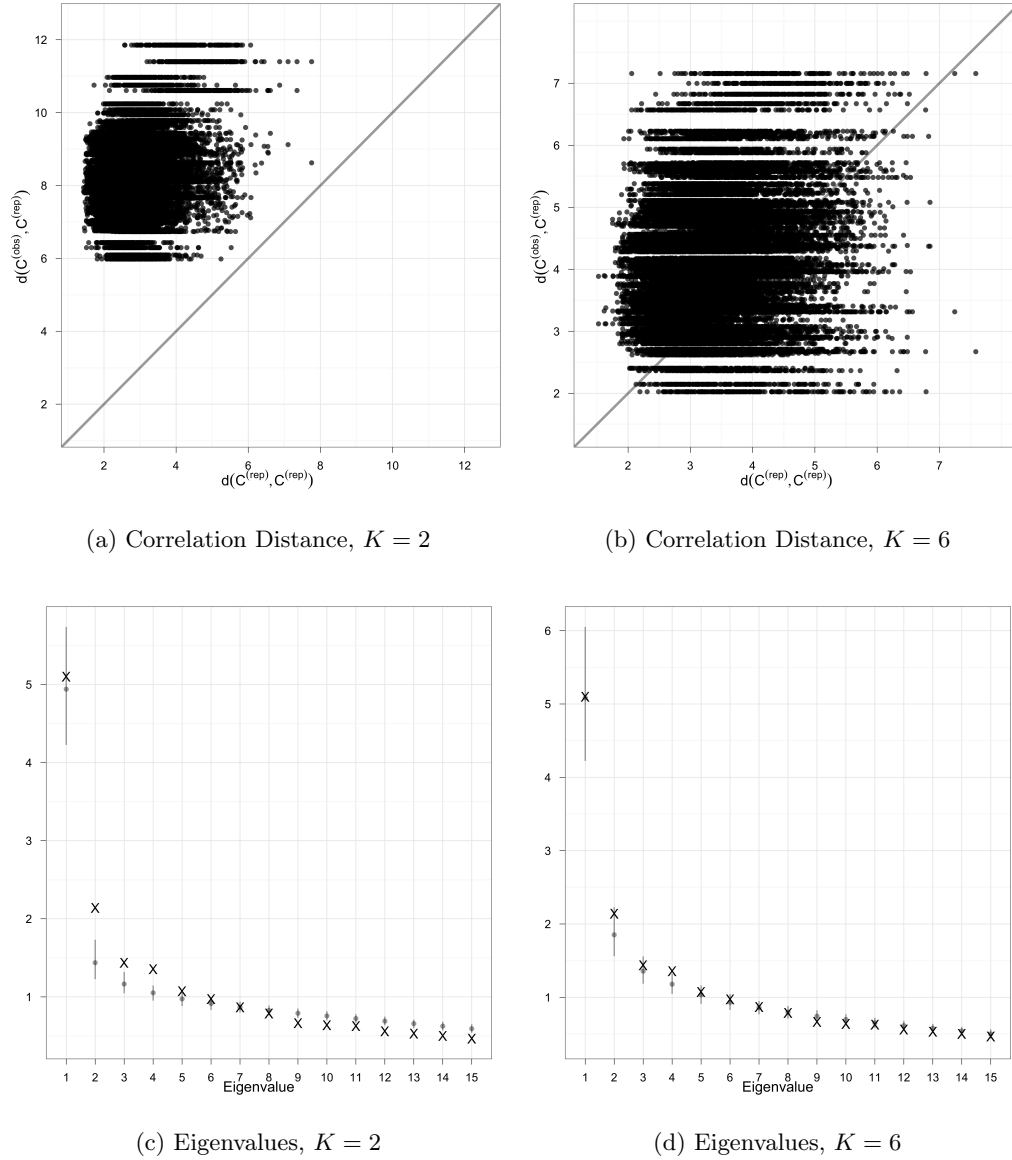


Figure 7.15: Correlation distance and eigenvalue plots for the $K = 2$ and $K = 6$ models. The left plot presents scatterplots of $d_{sld}(C^{obs}, C^{rep,m})$ versus $d_{sld}(C^{rep,m}, C^{rep,m'})$ for all replicated datasets. The grey line represents the 45 degree line. The right plot displays the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) of the largest ten eigenvalues calculated using replicated data. Eigenvalues computed from the observed data are denoted by a black “X”.

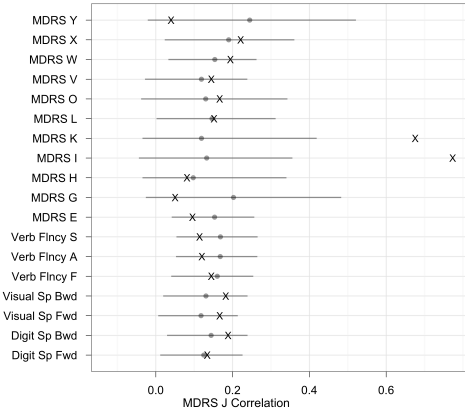
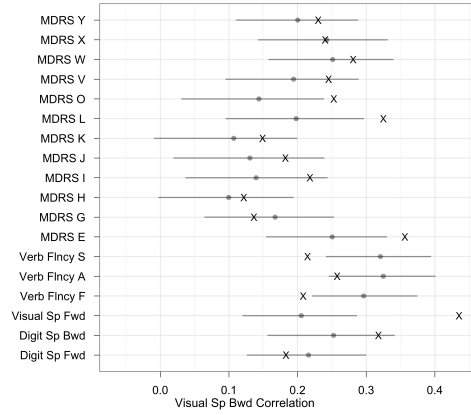
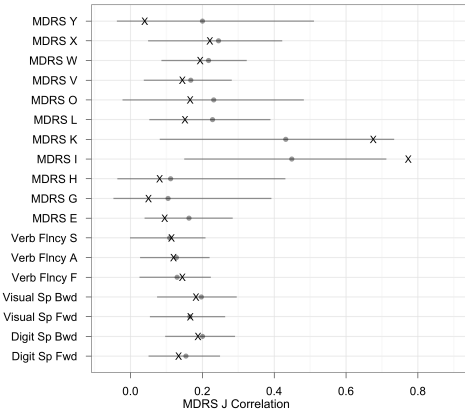
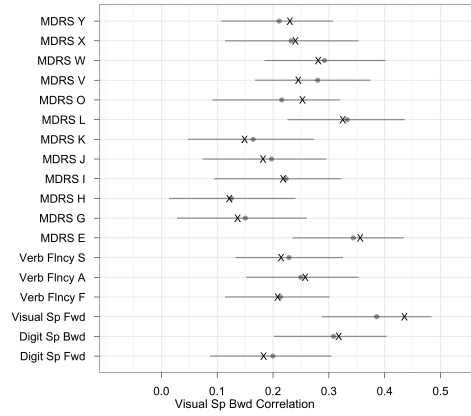
(a) MDRS J, $K = 2$ (b) Visual Span Backwards, $K = 2$ (c) MDRS J, $K = 6$ (d) Visual Span Backwards, $K = 6$

Figure 7.16: Pairwise correlation plots for the $K = 2$ and $K = 6$ models. Each pairwise correlation plot depicts the mean posterior prediction (grey point) and 95% posterior prediction intervals (grey line segment) for Kendall's τ values calculated using replicated data. Kendall's τ values computed from the observed data are denoted by a black "X".

7.8 Discussion

In this chapter, we extended the Bayesian partial membership (BPM) model of Heller et al. (2008) in two ways. First, we followed the developments of Blei and Lafferty (2007) for the mixed membership topic model and used a logistic normal prior for untransformed membership scores. Use of this prior allowed us to model and estimate correlations among membership scores. Second, we augmented the model’s ability to accommodate mixed outcomes by applying the extended rank likelihood model of Hoff (2007).

Prior to discussing the extensions of the model, we examined the similarities and differences of Mixed Membership (MM) models and BPM models. One of the differences between the models that we highlighted was the fact that the BPM model need not impose a local independence assumption. We demonstrated the different sets of data produced by MM and BPM models with the same distributions for each class. Moreover, in the case of binary data, we highlighted the fact that the models could be very similar in some instances before diverging significantly at extreme values of the class parameters.

We applied the semiparametric correlated partial membership model to simulated data and two real datasets. Each of the examples proved illuminating. In demonstrating the semiparametric model on simulated data, we showed that the model could recover the data-generating parameters but we also examined whether the effect of misspecifying the number of classes would be reflected in the posterior predictive model checks. In the simulated data case, we found that specifying too few classes resulted in poor fit but the fit with too many classes was comparable to that with the correct number of classes. However, in the cases of too many classes, we saw in the simulated data example that the mean class membership for one of the classes was close to zero. We noted that we do not necessarily expect this result across different applied examples as we have not run repeated simulations and do not have a complete understanding of how the model performs when the number of classes is misspecified.

The NBA player data example illustrated a number of features of the semiparametric correlated partial membership model including the ability to flexibly model correlations among classes, the elimination of the need to specify conditional distributions for the out-

comes and the utility of individual-level mixture models in general. Compared to a cluster analysis of the same data (Lutz, 2012), we found fewer types were needed to fit the data because the playing styles of certain players could successfully be represented by partial membership. Estimation of the model however was plagued by extremely slow mixing and convergence, highlighting the need for further research into methods that will improve the computational efficiency of estimating the model.

In the SIVD study example with executive functioning indicators, we found that a two class model appeared to fit the observed marginal distributions but struggled with replicating the observed rank correlations. We found a model with a larger number of classes could fit the observed correlations well but did not produce classes with clear substantive interpretations. To represent a larger array of data patterns (as demonstrated in Section 7.3), we could extend the model to allow for covariances among outcomes common across all classes, class specific variances or ideally class specific covariances. We have implemented the Bayesian partial membership model for normally distributed data with class-specific variances as well as means. Extension to the semiparametric version of the model is straightforward but in our experiments we encountered to an even greater extent some of the computation challenges witnessed in the NBA player data example.

Although we addressed some identifiability issues in Section 7.4, we have not thoroughly demonstrated the identifiability of the model and, as we consider more complex models, the issue of identifiability will become more important. Frühwirth-Schnatter (2006) provides an overview of identifiability for population-level mixture models. Holzman, Munk, and Gneiting (2006) note that “identifiability often is tacitly assumed to hold while proofs remain unavailable.” While many mixtures of univariate continuous densities have been shown to be identifiable (not including nonidentifiability due to permutation of class labels) as well as mixtures of multivariate normals, some discrete mixtures have been shown to be nonidentifiable (Frühwirth-Schnatter, 2006). Teicher (1967) has proven identifiability results for products of densities. These results may be useful in establishing identifiability of the partial membership model going forward.

Chapter 8

CONCLUSION AND FUTURE WORK**8.1 Conclusion**

In this dissertation, we proposed novel methods for the analysis of multivariate mixed outcomes. Our primary aims were to model the interdependence among mixed outcomes and make inference on the association between mixed outcomes and a set of covariates. We considered both parametric and semiparametric approaches to modeling mixed outcomes. In the parametric case, we extended the standard IRT model to accommodate mixed outcomes specific to cognitive testing data. In the semiparametric case, we based our approach on the extended rank likelihood method of Hoff (2007), originally developed for the estimation of copulas. We introduced this method for the estimation of semiparametric models in the areas of factor analysis, multivariate regression, and model-based clustering for mixed outcomes.

The Subcortical Ischemic Vascular Dementia (SIVD) study, a longitudinal study collecting brain imaging and neuropsychological testing data from elderly participants, motivated the development of these models. Specifically, we sought to analyze the indicators of executive functioning in the neuropsychological testing data and study the association between these outcomes and the MRI-measured volume of white matter hyperintensities in the frontal lobe. The neuropsychological tests employed in the SIVD study and related studies produce mixed data, directing our focus on methods for outcomes of different type. We now summarize our methodological contributions before presenting some ideas for future work in the next section.

Following a parametric approach to accommodating mixed outcomes, we extended a latent variable model with standard IRT parameterization to handle mixed outcomes specific to cognitive testing data. We utilized the generalized latent trait model developments of Sammel et al. (1997), Moustaki and Knott (2000) and Dunson (2003) to handle mixed data

by specifying different distributions for each outcome conditional on the latent variable. We extended their models to allow for censored count outcomes, duration outcomes and censored duration outcomes that often appear in tests of cognitive functioning. A hierarchical version of the model specified the latent variable, or ability parameter in IRT terminology, as a function of covariates.

While the parametric approach proved flexible, we observed that the specification of the conditional distributions of the outcomes may be a nuisance in the sense that it is not a primary inferential interest and may be just a necessary step in the parametric approach for mixed outcomes. In many cases, our primary interest may lay in the latent variable estimates and their relation to covariates of interest. We proposed a hierarchical semiparametric latent variable model with bifactor structure for mixed outcomes. The model allowed us to avoid specification of the parametric distributions for outcomes conditional on the latent variables. Moreover, the bifactor structure enabled us to more accurately approximate the correlations observed in the data than we could with a single factor model. We used the extended rank likelihood, a method that relies on the partial orderings of the observed outcomes, to estimate the semiparametric bifactor model.

In practice, applications of the semiparametric bifactor model require identification of a suitable bifactor structure to approximate the item dependencies in the observed data. To avoid specifying the latent structure among the outcomes while retaining the mean structure of the hierarchical semiparametric latent variable model, we developed a semiparametric multivariate regression model using the extended rank likelihood for estimation. In the semiparametric multivariate regression model, we specified the latent responses as direct functions of the covariates rather than being related to the covariates through a latent factor. Moreover, we employed an unstructured correlation matrix for the latent responses that we estimated directly rather than assuming a structured form as we do in the case of the semiparametric latent variable model. Lastly, we used a reduced rank representation for the regression coefficient matrix in the semiparametric multivariate regression model. This reduced rank representation aided in interpretation of the coefficients and we examined the relationship of the reduced rank regression model to the semiparametric latent variable model.

Given the utility of the semiparametric approach for latent variable and multivariate regression models for mixed outcomes, we considered its application with another type of analysis of multivariate mixed data, namely, model-based clustering. Specifically, we focused on the Bayesian partial membership model of Heller et al. (2008). We extended the Bayesian partial membership model of Heller et al. (2008) in two ways. First, we abandoned the Dirichlet prior for the membership scores in favor of a logistic normal prior that allowed for the estimation of class membership correlations. Second, we applied the extended rank likelihood to model mixed outcomes. We demonstrated the resulting semiparametric correlated partial membership model on NBA player data from the 2010-11 season to identify playing styles as well as on the SIVD executive functioning data. In addition, we compared the data generated by the Bayesian partial membership model with the mixed membership model for continuous and discrete outcomes.

Throughout this dissertation, we relied on the Bayesian approach to estimation. We developed several posterior predictive model checking methods that allowed us to examine model fit for multivariate mixed outcomes. We found the posterior predictive model checks proposed in Section 2.4 to be useful in detecting model misfit. In the case of the semiparametric latent variable model, posterior predictive model checks were also helpful in informing specification of the latent structure in the model development process.

8.2 Future Work

8.2.1 Further Development of the Semiparametric Correlated Partial Membership Model

Relax Restrictions on Class-Specific Covariances

In Section 7.3, we demonstrated the varying patterns of data that could be generated when the class-specific distributions were normal with class-specific covariance matrices. However, in our initial implementation of the semiparametric correlated partial membership model presented in Section 7.4, we restricted the covariance matrix (for the outcomes, not the class memberships) to be equal across classes and we restricted this common covariance matrix to be diagonal. Moving forward, there are a number of ways to relax this restriction. The most general approach allows for class specific covariance matrices without restrictions (other than

those required to set the scale of the latent responses). As intermediate steps to this model, it may be useful to consider the following extensions: (a) a semiparametric correlated partial membership model with class-specific variances, or (b) a semiparametric correlated partial membership with the common covariance matrix that is no longer restricted to be diagonal. In the analysis of the SIVD study data for example, the model described in (b) may be sufficient to resolve the misfit evident in the dependence-related posterior predictive model checks.

Study Association Between Class Membership and Covariates

We are also interested in understanding how class memberships vary with covariates of interest. Currently, individual class memberships are specified as the transformed draws from a multivariate distribution with mean class membership vector $\boldsymbol{\rho}$. Extending the semiparametric correlated partial membership model equations (7.35)-(7.41) hierarchically, one way to incorporate covariates is to specify the elements of the mean class membership vector $\boldsymbol{\rho}$ to vary by individual as a function of covariates so that

$$\begin{aligned}\rho_{i1} &= \mathbf{x}_i^T \boldsymbol{\beta}_1 \\ \rho_{i2} &= \mathbf{x}_i^T \boldsymbol{\beta}_2 \\ &\vdots \\ \rho_{i(K-1)} &= \mathbf{x}_i^T \boldsymbol{\beta}_{K-1},\end{aligned}\tag{8.1}$$

where $\mathbf{x}_i$ and $\boldsymbol{\beta}_k$ are P -length vectors. Given the often slow convergence of the semiparametric correlated partial membership and the addition of a potentially large number of parameters in $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{(K-1)}$, it is unclear how well we could estimate such a model.

Improving Computational Efficiency

In some of the examples of the semiparametric correlated partial membership model, we observed that the parameter chains were very slow to mix, much more so than the standard BPM model. In the case of the semiparametric latent variable model, we found that a parameter expanded approach to Gibbs sampling (Liu et al., 1998) could be very helpful in

improving the mixing of the chains. We plan to explore a similar approach in the case of the semiparametric correlated partial membership model. We will also consider a number of methods for improving the efficiency of HMC, many of which are discussed in Neal (2010).

8.2.2 Comparing Parametric and Semiparametric Approaches to Mixed Outcome Latent Variable Models

In Chapters 4 and 5, we presented parametric and semiparametric approaches to latent variable modeling for mixed outcomes. However, other than their respective applications to the SIVD study data, we have not compared the two models to gain a better understanding of the advantages and disadvantages of each approach. To better comprehend the differences in the models under different data-generating scenarios, we are interested in comparing the models in a simulation study setting. A plan for comparison of the two models could proceed as follows. First, we generate a large number of datasets according to the IRT model for mixed outcomes using a variety of parameter values and conditional distributions for the outcomes. Second, we generate data in a manner that favors the semiparametric latent variable model; that is, by the methods used in the simulated data example for the semiparametric latent variable model detailed in Section 5.5.

We then evaluate the models' ability to fit the data using the posterior predictive model checks under the two data generating scenarios. Specifically, we compare the frequency with which the 95% posterior predictive intervals for the score counts and the pairwise rank correlations contain the corresponding value computed from the simulated data. In addition, we can compare the average width of these 95% posterior predictive intervals as a means of comparing efficiency under different data-generating scenarios.

Because we hesitate to compare the latent variable values across different models as the values in one model may not be easily translated to values in the other, we will compare whether one model produces posterior estimates for the latent variables for which the order is more consistent with the order of the true latent variable values from the data generating model. If data is generated using the hierarchical model, we can further compare with what frequency the posterior credible intervals for the regression parameters lead to conclusions

about the regression parameters that are consistent with the data generating regression coefficients.

8.2.3 Examining Properties of Test Items with Semiparametric Latent Variable Model

Item Difficulty

One of the advantages often cited for item response theory modeling is that it allows you to characterize both the individuals taking the test and the items composing the test. A key parameter characterizing the questions is the difficulty parameter, b in the notation used in Chapter 4. In the item response theory parameterization used in Chapter 4 (see equation 4.1), the probability of a response for an individual i on outcome j , y_{ij} , conditional on the item parameters and the latent variables depends solely on the item difficulty parameter, b_j , when the latent variables are set to zero. Similarly, the expected value of a response y_{ij} conditional on the item parameters and the latent variables is a function of only the item difficulty parameter, b_j , when the latent variables are set to zero.

This parameter sets the location of the item responses and, as its name implies, is used to compare the difficulty of the item with other items on the test. As discussed in Chapter 5, such an outcome-specific location parameter is unable to be estimated directly in the semiparametric latent variable model. Nonetheless, we can estimate a proxy by estimating the posterior predictive mean response for an outcome for a hypothetical individual whose latent variables are set to 0. If the outcomes are normally distributed, the posterior predictive mean response for individuals whose latent variables were set to 0 corresponds directly to the difficulty parameter, or the outcome specific mean as it might be referred to in factor analysis. If the outcomes are binary, the posterior predictive mean response conditional on latent variable values of 0 corresponds to a function of only the difficulty parameter. We can compare this posterior predictive mean conditional on latent variable values equal to 0 across items to evaluate the items' relative difficulties.

Differential Item Functioning

The identification of differential item functioning evaluates whether item properties are consistent across different groups of examinees. The development of methods for the detection of differential item functioning in the semiparametric latent variable model is of interest. One approach to detect differential item functioning is to extend the semiparametric variable model hierarchically by allowing the loadings and an item-specific mean to vary by group membership. For example, to examine whether the loadings on the primary factor vary by gender, we specify the mean of the product, $\lambda_{j1}\eta_{i1}$, as

$$E[\lambda_{j1}\eta_{i1}] = (\alpha_{0j} + \alpha_{1j}\text{Gender}_i)\eta_{i1}. \quad (8.2)$$

If we wanted to simultaneously evaluate differences in item-specific mean by gender, we could specify

$$E[\mu_j + \lambda_{j1}\eta_{i1}] = (\gamma_{0j} + \gamma_{1j}\text{Gender}_i) + (\alpha_{0j} + \alpha_{1j}\text{Gender}_i)\eta_{i1}. \quad (8.3)$$

Recall that the outcome-specific means are not identifiable for the semiparametric latent variable model. As a result, we need to fix a parameter to obtain identifiability; for instance, we might set $\gamma_{0j} = 0$.

8.2.4 Extensions for Longitudinal Data

Latent Variable Models

In recent decades, the methods for analyzing longitudinal data and the number of publications using longitudinal data have steadily increased (van Montfort, Oud, and Satorra, 2010). Studies of cognition often follow participants longitudinally to investigate the evolution of cognition in individuals and study the factors influencing this evolution. Our motivating example, the SIVD study, follows participants longitudinally. Similarly, the Alzheimer's Disease Neuroimaging Study collects brain imaging, biomarker and neuropsychological testing results from elderly participants on a longitudinal basis. To model longitudinal data, we consider longitudinal extensions of the latent variable models presented in Chapters 4 and 5.

Recent literature on longitudinal latent variable models includes Dunson (2003), Andrade and Tavares (2005), Cagnone, Moustaki, and Vasdekis (2009), Liu and Hedeker (2006), te Marvelde, Glas, Van Landeghem, and Van Damme (2006) and van Montfort et al. (2010). A common theme of these models is that the latent variable(s) is allowed to vary across time, often as a function of time-varying covariates and with some autoregressive error dependence structure. Many of the models, although not all, assume that the individuals are observed at the same time points for the same number of occasions. Although not specifically designed for longitudinal data, Fox and Glas (2001) present a comprehensive formulation of a Bayesian multilevel IRT model. We suggest the formulation of Fox and Glas (2001) as the template for the longitudinal latent variable model for mixed outcomes in order to accommodate irregularly observed outcomes and induce within-individual dependence over time through individual-specific coefficients. In this model, we additionally include a random effect particular to the individual and outcome as a means of inducing additional dependence between responses to the same item by the same individual at different time points.

As our proposed longitudinal data extensions for the parametric and semiparametric latent variable models for mixed outcomes are very similar, we focus on the longitudinal data extension for the semiparametric model. Let $i = 1, \dots, I$ denote the i th participant, $j = 1, \dots, J$ denote the j th item and let t denote time. Age of the individual would be one measure of time that we would likely consider. The response for individual i on item j at time t is denoted as y_{ijt} . Covariates for participant i at time t are denoted as $\mathbf{x}_{it}$. Finally let $g_j(\cdot)$ be an unspecified monotone transformation particular to item j of a normally distributed random variable z_{ijt} so that $y_{ijt} = g_j(z_{ijt})$.

We propose the following semiparametric longitudinal latent variable model:

$$y_{ijt} = g_j(z_{ijt}), \quad (8.4)$$

$$z_{ijt} \sim N(\boldsymbol{\lambda}_j^T \boldsymbol{\eta}_{it} + b_{ij}, 1), \quad (8.5)$$

$$\eta_{itq} \sim N(\mathbf{x}_{it}^T \boldsymbol{\beta}_{iq}, 1), \quad (8.6)$$

$$b_{ij} \sim N(b_j, \sigma_b^2), \quad (8.7)$$

$$\boldsymbol{\beta}_i \sim MVN(\boldsymbol{\mu}_{\boldsymbol{\beta}_i}, \Sigma_{\boldsymbol{\beta}}^2), \quad (8.8)$$

$$\mu_{\beta_{iq}} = \mathbf{w}_{iq}^T \boldsymbol{\gamma}_q. \quad (8.9)$$

This model has a number of useful features. First, the latent variables (indexed by $q = 1, \dots, Q$) vary across time. Second, the responses for the same individual are correlated across time. Moreover, the model allows for responses to the same item for the same individual to be more correlated than responses to different items across time for the same individual. Finally, the model allows the number of longitudinal observations to vary across individuals and to be unequally spaced.

Now consider the specific elements of the model. Equation (8.5) for the mean of z_{ijt} resembles the common 2PL linear predictor with a couple of exceptions. The latent variables, $\boldsymbol{\eta}_{it}$, will vary over time as suggested by the index t affixed to it. Furthermore, there is a outcome- and individual-specific random effect, b_{ij} , that did not exist in our cross-sectional model. In the longitudinal model, we include this parameter to induce dependence over time between responses for the same item by the same individual. As for the outcome-specific mean in the cross-sectional model, there would again be identifiability issues but these may be alleviated by incorporating a restriction on one of the values.

Although the latent variables in equation (8.5) are allowed to vary with time, we would like to induce dependence in these variables across time within the individual. This then would also be a source of dependence between responses within an individual over time. This is achieved in equation (8.6) for the latent variables where x_{it} is a P length vector of time-varying covariates at time t for individual i . The individual-specific coefficient vector $\boldsymbol{\beta}_i$ induces dependence across time in the latent abilities for an individual. The vector $\boldsymbol{\beta}_i$ is specified by equation (8.8) where Σ_{β} is a covariance matrix for the random coefficients and $\boldsymbol{\mu}_{\beta_i}$ is the mean vector.

We specify the means of the individual-specific coefficients $\boldsymbol{\beta}_i$ as functions of non time-varying covariates denoted by $\mathbf{w}_{ip}$. The mean function for the p th of the P individual-specific coefficients is then set in equation (8.9) where $\boldsymbol{\gamma}_p$ is a vector of population-level coefficients.

Semiparametric Multivariate Regression

Although this type of model has not been applied using the semiparametric approach, it has been applied to latent variable models using the parametric approach to mixed outcomes (Dunson, 2003). A more novel development might then be the application of the semiparametric multivariate regression model to longitudinal multivariate data. Although multivariate regression models have been applied to longitudinal data, these applications typically model repeated univariate measurements. In our case, we seek to model repeated multivariate measurements. We can take advantage of the reduced rank representation used in our semiparametric regression model and incorporate individual-specific coefficients as in equation (8.8) for the semiparametric multivariate regression model. Assuming such a model can be estimated, this development would enable the modeling of irregularly-spaced repeated multivariate measurements using the semiparametric multivariate regression model.

In conclusion, we feel that the work in this dissertation not only introduces useful new models for the analysis of multivariate mixed data but also provides plenty of interesting directions for future research and developments. We look forward to pursuing these directions.

BIBLIOGRAPHY

- E.M. Airoldi, D.M. Blei, S.E. Fienberg, and E.P. Xing. Mixed Membership Stochastic Blockmodels. *The Journal of Machine Learning Research*, 9:1981–2014, 2008.
- J. Aitchison and S.D. Silvey. The Generalization of Probit Analysis to the Case of Multiple Responses. *Biometrika*, 44(1/2):131–140, 1957.
- R. Alvarez, G. Katz, R. Llamosa, and H. Martinez. Assessing Voters Attitudes Towards Electronic Voting in Latin America: Evidence from Colombias 2007 E-Voting Pilot. *E-Voting and Identity*, pages 75–91, 2009.
- T.W. Anderson. Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions. *The Annals of Mathematical Statistics*, pages 327–351, 1951.
- T.W. Anderson. On Estimation of Parameters in Latent Structure Analysis. *Psychometrika*, 19(1):1–10, 1954.
- T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience, 2003.
- T.W. Anderson and H. Rubin. Statistical Inference in Factor Analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 5, pages 111–150, 1956.
- D.F. Andrade and H.R. Tavares. Item Response Theory for Longitudinal Data: Population Parameter Estimation. *Journal of Multivariate Analysis*, 95(1):1–22, 2005.
- A. Armagan, D.B. Dunson, and M. Clyde. Generalized Beta Mixtures of Gaussians. *Arxiv Preprint arXiv:1107.4976*, 2011.
- J.R. Ashford and R.R. Sowden. Multivariate Probit Analysis. *Biometrics*, 26:535–546, 1970.
- J. Aitchison and S.M. Shen. Logistic-Normal Distributions: Some Properties and Uses. *Biometrika*, 67(2):261–272, 1980.
- A. Banerjee, S. Merugu, I. Dhillont, and I. Ghosh. Clustering with Bregman Divergences. In *Proceedings of the Fourth SIAM International Conference on Data Mining*, volume 117, page 234. Society for Industrial Mathematics, 2004.

- J. Barnard, R. McCulloch, and X.L. Meng. Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage. *Statistica Sinica*, 10(4):1281–1312, 2000.
- D. Bartholomew, M. Knott, and I. Moustaki. *Latent Variable Models and Factor Analysis: A Unified Approach*. John Wiley & Sons, 2011.
- M.J. Bayarri and J.O. Berger. P Values for Composite Null Models. *Journal of the American Statistical Association*, 95(452):1127–1142, 2000. ISSN 0162-1459.
- James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, 1981.
- P.J. Bickel and Y. Ritov. Local Asymptotic Normality of Ranks and Covariates in Transformation Models. *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, pages 43–54, 1997.
- D.M. Blei and J.D. Lafferty. A Correlated Topic Model of Science. *Annals of Applied Statistics*, 1(1):17–35, 2007.
- D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.
- R.D. Bock. A Brief History of Item Response Theory. *Educational Measurement: Issues and Practice*, 16(4):21–33, 1997.
- RD Bock, R. Gibbons, SG Schilling, E. Muraki, DT Wilson, and R. Wood. *TEST-FACT 4.0 [Computer Software]*. Scientific Software International, 2003.
- K.A. Bollen. *Structural Equations with Latent Variables*. John Wiley & Sons, 1989.
- E.T. Bradlow, H. Wainer, and X. Wang. A Bayesian Random Effects Model for Testlets. *Psychometrika*, 64(2):153–168, 1999.
- S. Cagnone, I. Moustaki, and V. Vasdekis. Latent Variable Models for Multivariate Longitudinal Ordinal Responses. *British Journal of Mathematical and Statistical Psychology*, 62(2):401–415, 2009.
- V.A. Cardenas, F. Ezekiel, V. Di Sclafani, B. Gomberg, and G. Fein. Reliability of Tissue Volumes and Their Spatial Distribution for Segmented Magnetic Resonance Images. *Psychiatry Research: Neuroimaging*, 106(3):193–205, 2001.
- C.M. Carvalho, J. Chang, J.E. Lucas, J.R. Nevins, Q. Wang, and M. West. High-Dimensional Sparse Factor Modeling: Applications in Gene Expression Genomics. *Journal of the American Statistical Association*, 103(484):1438–1456, 2008.

- R.B. Cattell. The Scree Test for the Number of Factors. *Multivariate Behavioral Research*, 1(2):245–276, 1966.
- L. Chen, Z. Qin, and J.S. Liu. Exploring Hybrid Monte Carlo in Bayesian Computation. In *Bayesian Methods with Applications to Science, Policy, and Official Statistics, Selected Papers from ISBA 2000: The Sixth World Meeting of the International Society for Bayesian Analysis*, pages 71–80, 2001.
- S. Chib and E. Greenberg. Analysis of Multivariate Probit Models. *Biometrika*, 85(2):347, 1998.
- H.C. Chui, C. Zarow, W.J. Mack, W.G. Ellis, L. Zheng, W.J. Jagust, D. Mungas, B.R. Reed, J.H. Kramer, C.C. DeCarli, et al. Cognitive Impact of Subcortical Vascular and Alzheimers Disease Pathology. *Annals of Neurology*, 60(6):677, 2006.
- W.S. Cleveland and S.J. Devlin. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, pages 596–610, 1988.
- P. Congdon. *Applied Bayesian Modelling*, volume 394. John Wiley & Sons Inc, 2003.
- P. Congdon. *Bayesian Statistical Modelling*, volume 670. Wiley, 2006.
- S.R. Cook, A. Gelman, and D.B. Rubin. Validation of Software for Bayesian Models using Posterior Quantiles. *Journal of Computational and Graphical Statistics*, 15(3): 675–692, 2006.
- C.M. Crespi and W.J. Boscardin. Bayesian Model Checking for Multivariate Outcome Data. *Computational Statistics & Data Analysis*, 53(11):3765–3772, 2009. ISSN 0167-9473.
- S.M. Curtis. BUGS Code for Item Response Theory. *Journal of Statistical Software*, 36(c01), 2010.
- M.J. Daniels and S.T. Normand. Longitudinal Profiling of Health Care Units Based on Continuous and Discrete Patient Outcomes. *Biostatistics*, 7(1):1, 2006. ISSN 1465-4644.
- G.S. Datta. On Priors Providing Frequentist Validity of Bayesian Inference for Multiple Parametric Functions. *Biometrika*, 83(2):287–298, 1996.
- J.V. Davis. *Mining Statistical Correlations with Applications to Software Analysis*. PhD thesis, The University of Texas at Austin, 2008.

- M.G. De Jong, J.B.E.M. Steenkamp, J.P. Fox, and H. Baumgartner. Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation. *Journal of Marketing Research*, 45(1):104–115, 2008.
- A. Dobra and A. Lenkoski. Copula Gaussian Graphical Models and Their Application to Modeling Functional Disability Data. *Annals of Applied Statistics*, 5(2A):969–993, 2011.
- G. Doyle and C. Elkan. Financial Topic Models. In *Working Notes of the NIPS-2009 Workshop on Applications for Topic Models: Text and Beyond Workshop*, 2009.
- S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, 1987.
- S.D. Dubey. Compound Gamma, Beta and F Distributions. *Metrika*, 16(1):27–31, 1970.
- D.B. Dunson. Dynamic Latent Trait Models for Multidimensional Longitudinal Data. *Journal of the American Statistical Association*, 98(463):555–564, 2003.
- D.B. Dunson. Efficient Bayesian Model Averaging in Factor Analysis. Technical report, Duke University, 2006.
- Y.D. Edwards and G.M. Allenby. Multivariate Analysis of Multiple Response Data. *Journal of Marketing Research*, 40(3):321–334, 2003.
- W.J. Ehlenbach, C.L. Hough, P.K. Crane, S.J.P.A. Haneuse, S.S. Carson, J.R. Curtis, and E.B. Larson. Association Between Acute Care and Critical Illness Hospitalization and Cognitive Function in Older Adults. *JAMA: The Journal of the American Medical Association*, 303(8):763–770, 2010.
- E.A. Erosheva. *Grade of Membership and Latent Structure Models with Application to Disability Survey Data*. PhD thesis, Carnegie Mellon University, 2002.
- E.A. Erosheva. Comparing Latent Structures of the Grade of Membership, Rasch, and Latent Class Models. *Psychometrika*, 70(4):619–628, 2005.
- E.A. Erosheva and S.M. Curtis. Specification of Rotational Constraints in Bayesian Confirmatory Factor Analysis. Technical Report 589, University of Washington, 2011.
- E.A. Erosheva and S.E. Fienberg. Bayesian Mixed Membership Models for Soft Clustering and Classification. *Classification - The Ubiquitous Challenge*, pages 11–26, 2005.

- E.A. Erosheva, S.E. Fienberg, and C. Joutard. Describing Disability Through Individual-Level Mixture Models for Multivariate Binary Data. *Annals of Applied Statistics*, 1(2):502–537, 2007.
- L. Fahrmeir and A. Raach. A Bayesian Semiparametric Latent Variable Model for Mixed Responses. *Psychometrika*, 72(3):327–346, 2007.
- J.P. Fox. *Bayesian Item Response Modeling: Theory And Applications*. Springer Verlag, 2010.
- J.P. Fox and C.A.W. Glas. Bayesian Estimation of a Multilevel IRT Model Using Gibbs Sampling. *Psychometrika*, 66(2):271–288, 2001.
- S. Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer Verlag, 2006.
- F. Galton. Co-Relations and Their Measurement, Chiefly from Anthropometric Data. *Proceedings of the Royal Society of London*, 45(273-279):135–145, 1888.
- A. Gelman. Prior Distributions for Variance Parameters in Hierarchical Models. *Bayesian Analysis*, 1(3):515–533, 2006.
- A. Gelman. Comment: Bayesian Checking of the Second Levels of Hierarchical Models. *Statistical Science*, 22(3):349–352, 2007.
- A. Gelman and D.B. Rubin. Inference from Iterative Simulation using Multiple Sequences. *Statistical Science*, 7(4):457–472, 1992.
- A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL., 2004.
- J. Geweke. Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments. In J.M. Bernardo, J. Berger, A.P. Dawid, and J.F.M. Smith, editors, *Bayesian Statistics 4*, pages 169–193. Oxford University Press, 1992.
- J. Ghosh and D.B. Dunson. Default Prior Distributions and Efficient Posterior Computation in Bayesian Factor Analysis. *Journal of Computational and Graphical Statistics*, 18(2):306–320, 2009.
- M. Girolami and B. Calderhead. Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.

C.A.W. Glas and R.R. Meijer. A Bayesian Approach to Person Fit Analysis in Item Response Theory Models. *Applied Psychological Measurement*, 27(3):217, 2003. ISSN 0146-6216.

A.S. Goldberger and R. Hauser. The Treatment of Unobservable Variables in Path Analysis. *Sociological Methodology*, 3(8):81–117, 1971.

I.C. Gormley and T.B. Murphy. A Grade of Membership Model for Rank Data. *Bayesian Analysis*, 4:265–296, 2009.

J.E. Griffin and P.J. Brown. Inference with Normal-Gamma Prior Distributions in Regression Problems. *Bayesian Analysis*, 5(1):171–188, 2010.

J. Gruhl, E. Erosheva, and P. Crane. Analyzing Cognitive Testing Data with Extensions of Item Response Theory Models, 2010. Joint Statistical Meetings.

J. Gruhl, E. Erosheva, and P. Crane. A Semiparametric Bayesian Latent Trait Model for Multivariate Mixed Type Data, 2011. International Meeting of the Psychometric Society.

L. Guttman. Some Necessary Conditions for Common-Factor Analysis. *Psychometrika*, 19(2):149–161, 1954.

V. Hachinski, C. Iadecola, R.C. Petersen, M.M. Breteler, D.L. Nyenhuis, S.E. Black, W.J. Powers, C. DeCarli, J.G. Merino, R.N. Kalaria, et al. National Institute of Neurological Disorders and Stroke–Canadian Stroke Network Vascular Cognitive Impairment Harmonization Standards. *Stroke*, 37(9):2220–2241, 2006.

JA Hartigan. Note on the Confidence-Prior of Welch and Peers. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 55–56, 1966.

A. Heinen and E. Rengifo. Multivariate Reduced Rank Regression in Non-Gaussian Contexts, Using Copulas. *Computational Statistics & Data Analysis*, 52(6):2931–2944, 2008.

T. Heinen. *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Sage Publications, Inc, 1996.

K.A. Heller, S. Williamson, and Z. Ghahramani. Statistical Models for Partial Membership. In *Proceedings of the 25th International Conference on Machine Learning*, pages 392–399, 2008.

- M. Herdin, N. Czink, H. Ozelik, and E. Bonek. Correlation Matrix Distance, A Meaningful Measure for Evaluation of Non-Stationary MIMO Channels. In *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, volume 1, pages 136–140. IEEE, 2005. ISBN 0780388879.
- P.D. Hoff. Extending the Rank Likelihood for Semiparametric Copula Estimation. *Annals of Applied Statistics*, 1(1):265–283, 2007.
- P.D. Hoff. Rank Likelihood Estimation for Continuous and Discrete Data. *ISBA Bulletin*, 15(1):8–10, 2008.
- P.D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer Verlag, 2009.
- P.D. Hoff, X. Niu, and J.A. Wellner. Information Bounds for Gaussian Copulas. *Arxiv preprint arXiv:1110.3572*, 2011.
- K.J. Holzinger and F. Swineford. The Bi-Factor Method. *Psychometrika*, 2(1):41–54, 1937.
- Hajo Holzman, Axel Munk, and Tilman Gneiting. Identifiability of Finite Mixtures of Elliptical Distributions. *Scandinavian Journal of Statistics*, 33(4):753–763, 2006.
- Hoopdata. NBA Player Statistics, March 2012. URL <http://www.hoopdata.com>.
- H. Hung and C.C. Wang. Matrix-Variate Logistic Regression Analysis. *Arxiv Preprint arXiv:1105.2150*, 2011.
- H. Ishwaran. Applications of Hybrid Monte Carlo to Bayesian Generalized Linear Models: Quasicomplete Separation and Neural Networks. *Journal of Computational and Graphical Statistics*, pages 779–799, 1999.
- A.J. Izenman. Reduced-Rank Regression for the Multivariate Linear Model. *Journal of Multivariate Analysis*, 5(2):248–264, 1975.
- R.I. Jennrich and P.M. Bentler. Exploratory Bi-Factor Analysis. *Psychometrika*, pages 1–13, 2011.
- N. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions, Vol. 1*. John Wiley & Sons, Inc., 1994a.
- N. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions, Vol. 2*. John Wiley & Sons, Inc., 1994b.
- V.E. Johnson and J.H. Albert. *Ordinal Data Modeling*. Springer Verlag, 1999.

K.G. Jöreskog. A General Approach to Confirmatory Maximum Likelihood Factor Analysis. *Psychometrika*, 34(2):183–202, 1969.

K.G. Jöreskog. A General Method for Analysis of Covariance Structures. *Biometrika*, 57(2):239–251, 1970.

K.G. Jöreskog and A.S. Goldberger. Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable. *Journal of the American Statistical Association*, pages 631–639, 1975.

M. Jorgensen and L. Hunt. Mixture Model Clustering of Data Sets with Categorical and Continuous Variables. In *Proceedings of the Conference ISIS*, volume 96, pages 375–384, 1996.

C. Klüppelberg and G. Kuhn. Copula Structure Analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):737–753, 2009.

D. Knowles and Z. Ghahramani. Nonparametric Bayesian Sparse Factor Models with Application to Gene Expression Modeling. *The Annals of Applied Statistics*, 5(2B):1534–1552, 2011.

B. Kuczynski, E. Targan, C. Madison, M. Weiner, Y. Zhang, B. Reed, H.C. Chui, and W. Jagust. White Matter Integrity and Cortical Metabolic Associations in Aging and Dementia. *Alzheimer's and Dementia*, 6(1):54–62, 2010.

D.N. Lawley. The Estimation of Factor Loadings by the Method of Maximum Likelihood. *Proceedings of the Royal Society of Edinburgh*, 60:64–82, 1940.

E. Lawrence, D. Bingham, C. Liu, and V.N. Nair. Bayesian Inference for Multivariate Ordinal Data Using Parameter Expansion. *Technometrics*, 50(2):182–191, 2008.

Paul F. Lazarsfeld and Henry W. Neil. *Latent Structure Analysis*. Houghton Mifflin, 1968.

R. Levy, R.J. Mislevy, and S. Sinharay. Posterior Predictive Model Checking for Multidimensionality in Item Response Theory. *Applied Psychological Measurement*, 33(7):519, 2009. ISSN 0146-6216.

Y. Li and D.W. Schafer. Likelihood Analysis of the Multivariate Ordinal Probit Regression Model for Repeated Ordinal Responses. *Computational Statistics & Data Analysis*, 52(7):3474–3492, 2008.

Y. Li, D.M. Bolt, and J. Fu. A Comparison of Alternative Models for Testlets. *Applied Psychological Measurement*, 30(1):3–21, 2006.

- C. Liu. [The Art of Data Augmentation]: Discussion. *Journal of Computational and Graphical Statistics*, 10(1):75–81, 2001.
- C. Liu, D.B. Rubin, and Y.N. Wu. Parameter Expansion to Accelerate EM: The PX-EM Algorithm. *Biometrika*, 85(4):755, 1998.
- J.S. Liu and Y.N. Wu. Parameter Expansion for Data Augmentation. *Journal of the American Statistical Association*, pages 1264–1274, 1999.
- L.C. Liu and D. Hedeker. A Mixed-Effects Regression Model for Longitudinal Multivariate Ordinal Data. *Biometrics*, 62(1):261–268, 2006.
- X. Liu and M.J. Daniels. A New Algorithm for Simulating a Correlation Matrix Based on Parameter Expansion and Reparameterization. *Journal of Computational and Graphical Statistics*, 15(4):897–914, 2006.
- F.M. Lord, M.R. Novick, and A. Birnbaum. *Statistical theories of mental test scores*. Addison-Wesley, 1968.
- I.R.R. Lu, D.R. Thomas, and B.D. Zumbo. Embedding IRT in Structural Equation Models: A Comparison with Regression Based on IRT Scores. *Structural Equation Modeling*, 12(2):263–277, 2005.
- D. Lutz. A Cluster Analysis Of NBA Players. In *MIT Sloan Sports Analytics Conference*, 2012.
- D.J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
- D.L. Miglioretti. Latent Transition Regression for Mixed Outcomes. *Biometrics*, 59(3):710–720, 2003. ISSN 1541-0420.
- R.E. Millsap. When Trivial Constraints Are Not Trivial: The Choice of Uniqueness Constraints in Confirmatory Factor Analysis. *Structural Equation Modeling*, 8(1): 1–17, 2001.
- J.F. Monahan and D.D. Boos. Proper Likelihoods for Bayesian Analysis. *Biometrika*, 79(2):271–278, 1992.
- J.C. Morris. The Clinical Dementia Rating (CDR): Current Version and Scoring Rules. *Neurology*, 1993.
- J.C. Morris. Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging Measure for Dementia of the Alzheimer Type. *International Psychogeriatrics*, 9(1): 173–176, 1997.

- I. Moustaki and M. Knott. Generalized Latent Trait Models. *Psychometrika*, 65(3):391–411, 2000.
- I. Moustaki, K.G. Jöreskog, and D. Mavridis. Factor Models for Ordinal Variables with Covariate Effects on the Manifest and Latent Variables: A Comparison of LISREL and IRT Approaches. *Structural Equation Modeling*, 11(4):487–513, 2004.
- S.G. Mueller, M.W. Weiner, L.J. Thal, R.C. Petersen, C. Jack, W. Jagust, J.Q. Trojanowski, A.W. Toga, and L. Beckett. The Alzheimers Disease Neuroimaging Initiative. *Neuroimaging Clinics of North America*, 15(4):869, 2005.
- D. Mungas, D. Harvey, BR Reed, WJ Jagust, C. DeCarli, L. Beckett, WJ Mack, JH Kramer, MW Weiner, N. Schuff, et al. Longitudinal Volumetric MRI Change and Rate of Cognitive Decline. *Neurology*, 65(4):565–571, 2005.
- E. Muraki. A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, 16(2):159, 1992.
- E. Muraki and D. Bock. *Parscale 4 [Computer Program]*. Scientific Software International, 2008.
- J.S. Murray, D.B. Dunson, L. Carin, and J.E. Lucas. Bayesian Gaussian Copula Factor Models for Mixed Data. *Arxiv Preprint arXiv:1111.0317*, 2011.
- R.M. Neal. MCMC Using Hamiltonian Dynamics. In S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, editors, *Handbook of Markov Chain Monte Carlo*. Boca Raton: CRC Press, 2010.
- A. Nicolaou. Bayesian Intervals with Good Frequentist Behaviour in the Presence of Nuisance Parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 377–390, 1993.
- Y. Ochi and R.L. Prentice. Likelihood Inference in a Correlated Probit Regression Model. *Biometrika*, 71(3):531, 1984.
- D.W. Osgood, B.J. McMorris, and M.T. Potenza. Analyzing Multiple-Item Measures of Crime and Deviance I: Item Response Theory Scaling. *Journal of Quantitative Criminology*, 18(3):267–296, 2002.
- T. Park and G. Casella. The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.
- R.J. Patz and B.W. Junker. Applications and Extensions of MCMC in IRT: Multiple Item Types, Missing Data, and Rated Responses. *Journal of Educational and Behavioral Statistics*, 24(4):342–366, 1999a.

- R.J. Patz and B.W. Junker. A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models. *Journal of Educational and Behavioral Statistics*, 24(2):146, 1999b.
- K. Pearson. Contributions to the Mathematical Theory of Evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110, 1894.
- A.N. Pettitt. Inference for the Linear Model Using a Likelihood Based on Ranks. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 234–243, 1982.
- M.M. Pitt, S.R. Khandker, and J. Cartwright. Empowering Women with Micro Finance: Evidence from Bangladesh. *Economic Development and Cultural Change*, 54(4):791–831, 2006.
- N.G. Polson and J.G. Scott. Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction. *Bayesian Statistics*, 9:76, 2010.
- A.E. Raftery and S.M. Lewis. The Number of Iterations, Convergence Diagnostics and Generic Metropolis Algorithms. In D.J. Spiegelhalter W.R. Gilks and S. Richardson, editors, *Practical Markov Chain Monte Carlo*. Chapman and Hall, London, U.K., 1995.
- P. Rai and H. Daumé III. The Infinite Hierarchical Factor Regression Model. *Arxiv Preprint arXiv:0908.0570*, 2009.
- M.D. Reckase. *Multidimensional Item Response Theory*. Springer Verlag, 2009.
- S.P. Reise, J. Morizot, and R.D. Hays. The Role of the Bifactor Model in Resolving Dimensionality Issues in Health Outcomes Measures. *Quality of Life Research*, 16: 19–31, 2007.
- J.M. Robins, A. van der Vaart, and V. Ventura. Asymptotic Distribution of P-Values in Composite Null Models. *Journal of the American Statistical Association*, 95(452): 1143–1156, 2000. ISSN 0162-1459.
- P.E. Rossi and G.M. Allenby. Bayesian Statistics and Marketing. *Marketing Science*, pages 304–328, 2003.
- F. Samejima. Estimation of Latent Ability Using a Response Pattern of Graded Scores. *Psychometrika*, 35(1):139–139, 1970.
- F. Samejima. A General Model for Free Response Data. *Psychometrika Monograph Supplement*, 37(1), 1972.

- M.D. Sammel, L.M. Ryan, and J.M. Legler. Latent Variable Models for Mixed Discrete and Continuous Outcomes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 667–678, 1997.
- H. Schmidli. Bayesian Analysis of Reduced Rank Regression. *Test*, 5(1):159–186, 1996.
- J.Q. Shi and S.Y. Lee. Bayesian Sampling-Based Approach for Factor Analysis Models with Continuous and Polytomous Data. *British Journal of Mathematical and Statistical Psychology*, 51(2):233–252, 1998.
- S. Sinharay. Assessing Fit of Unidimensional Item Response Theory Models Using a Bayesian Approach. *Journal of Educational Measurement*, 42(4):375–394, 2005.
- S. Sinharay and M.S. Johnson. Simulation Studies Applying Posterior Predictive Model Checking for Assessing Fit of the Common Item Response Theory Models (ETS RR-03-28). *Princeton, NJ: Educational Testing Service*, 2003.
- S. Sinharay, M.S. Johnson, and H.S. Stern. Posterior Predictive Assessment of Item Response Theory Models. *Applied Psychological Measurement*, 30(4):298, 2006.
- A. Skrondal and S. Rabe-Hesketh. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. CRC Press, 2004.
- M. Stephens. Dealing with Label Switching in Mixture Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):795–809, 2000.
- A. Tabet. *Bayesian Inference in the Multivariate Probit Model*. PhD thesis, The University Of British Columbia, 2007.
- Y. Takane and J. de Leeuw. On the Relationship Between Item Response Theory and Factor Analysis of Discretized Variables. *Psychometrika*, 52(3):393–408, 1987.
- J.M. te Marvelde, C.A.W. Glas, G. Van Landeghem, and J. Van Damme. Application of Multidimensional Item Response Theory Models to Longitudinal Data. *Educational and Psychological Measurement*, 66(1):5–34, 2006.
- H. Teicher. Identifiability of Mixtures of Product Measures. *The Annals of Mathematical Statistics*, 38(4):1300–1302, 1967.
- L.L. Thurstone. *Multiple Factor Analysis*. University Of Chicago Press, 1947.
- W.J. Van der Linden and R.K. Hambleton. *Handbook of Modern Item Response Theory*. Springer Verlag, 1997.

- K. van Montfort, J.H.L. Oud, and A. Satorra. *Longitudinal Research with Latent Variables*. Springer Verlag, 2010.
- J.N. Variyam, J. Blaylock, and D. Smallwood. Reduced-Rank Models for Nutrition Knowledge Assessment. *Biometrics*, pages 1654–1661, 1998.
- M.W. Weiner, P.S. Aisen, C.R. Jack, W.J. Jagust, J.Q. Trojanowski, L. Shaw, A.J. Saykin, J.C. Morris, N. Cairns, L.A. Beckett, et al. The Alzheimer’s Disease Neuroimaging Initiative: Progress Report and Future Plans. *Alzheimer’s and Dementia*, 6(3):202–211, 2010.
- M. West. On Scale Mixtures of Normal Distributions. *Biometrika*, 74(3):646, 1987.
- M.A. Woodbury, J. Clive, and A. Garson Jr. Mathematical Typology: A Grade of Membership Technique for Obtaining Disease Definition. *Computers and Biomedical Research*, 11(3):277–298, 1978.
- T.W. Yee and T.J. Hastie. Reduced-Rank Vector Generalized Linear Models. *Statistical Modelling*, 3(1):15, 2003.
- MF Zimowski, E. Muraki, RJ Mislevy, and RD Bock. *BILOG-MG: Multiple-Group IRT Analysis and Test Maintenance for Binary Items*. Scientific Software International., 2008.

Appendix A

INDUCED PRIORS

In applying parameter expansion (PX) techniques to improve MCMC convergence, one may induce priors for the identified parameters that differ from those in the original model (non parameter expanded model). Liu and Wu (1999) note that the marginal posterior for a parameter will be the same under the parameter expanded and original models if and only if the marginal prior from the parameter expanded approach is the same as that specified in the working model. In the model of Ghosh and Dunson (2009), the parameter expansion induces t prior distributions and folded- t prior distributions for the elements of the factor loading matrix, $\mathbf{\Lambda}$. In the case of the semiparametric model where the variances of the factors denoted by the diagonal $Q \times Q$ matrix $\mathbf{\Psi}$ are allowed to vary but the variances of the latent responses, $\mathbf{\Sigma}$, are restricted, the induced prior on the factor loadings is a t -distribution. We show this here to help understand how the induced prior is obtained. We will then discuss the prior induced by the PX algorithm where both $\mathbf{\Psi}$ and $\mathbf{\Sigma}$ are unrestricted during the MCMC sampling.

Under the original model,

$$p(\mathbf{\Lambda}, \mathbf{H} | \mathbf{Z} \in D(\mathbf{Y})) \propto p(\mathbf{\Lambda})p(\mathbf{H}) \times \int_{D(\mathbf{Y})} p(\mathbf{Z} | \mathbf{\Lambda}, \mathbf{H}). \quad (\text{A.1})$$

Under the working model in the parameter expanded approach,

$$p(\mathbf{\Lambda}^*, \mathbf{H}^*, \mathbf{\Sigma}, \mathbf{\Psi} | \mathbf{Z}^* \in D(\mathbf{Y})) \propto p_{\mathbf{\Lambda}^*}(\mathbf{\Lambda}^*)p_{\mathbf{H}^*}(\mathbf{H}^*)p(\mathbf{\Sigma})p(\mathbf{\Psi}) \times \int_{D(\mathbf{Y})} p(\mathbf{Z}^* | \mathbf{\Lambda}^*, \mathbf{H}^*, \mathbf{\Sigma}, \mathbf{\Psi}) \quad (\text{A.2})$$

$$\begin{aligned} &= \prod_j \prod_q p_{\lambda_{jq}^*}(\lambda_{jq}^*) \cdot \prod_i \prod_q p_{\eta_{iq}^*}(\eta_{iq}^*) \cdot \prod_j p(\sigma_j^{-2}) \cdot \prod_q p(\psi_q^{-2}) \\ &\quad \times \int_{D(\mathbf{Y})} p(\mathbf{Z}^* | \mathbf{\Lambda}^*, \mathbf{H}^*, \mathbf{\Sigma}, \mathbf{\Psi}). \end{aligned} \quad (\text{A.3})$$

where $p_{\mathbf{\Lambda}^*}(\cdot), p_{\mathbf{H}^*}(\cdot)$ are used to distinguish the prior densities in the parameter expanded model from those in the original model for the parameters $\mathbf{\Lambda}$ and $\mathbf{H}$.

Because

$$\int_{D(\mathbf{Y})} p(\mathbf{Z}|\mathbf{\Lambda}, \mathbf{H}) = \int_{D(\mathbf{Y})} p(\mathbf{Z}^*|\mathbf{\Lambda}^*, \mathbf{H}^*, \mathbf{\Sigma}, \mathbf{\Psi}), \quad (\text{A.4})$$

the marginal posteriors for $\mathbf{\Lambda}$ and $\mathbf{H}$ will be the same under the two models if

$$p(\mathbf{\Lambda})p(\mathbf{H}) = \int p_{\mathbf{\Lambda}^*}(t_{\mathbf{\Psi}, \mathbf{\Sigma}}(\mathbf{\Lambda}))|J_{\mathbf{\Sigma}, \mathbf{\Psi}}(\mathbf{\Lambda})|p_{\mathbf{H}^*}(t_{\mathbf{\Psi}}(\mathbf{H}))|J_{\mathbf{\Psi}}(\mathbf{H})|p(\mathbf{\Sigma})p(\mathbf{\Psi}) d\mathbf{\Sigma}d\mathbf{\Psi} \quad (\text{A.5})$$

where $t_{\mathbf{\Sigma}}(\cdot), t_{\mathbf{\Psi}, \mathbf{\Sigma}}(\cdot)$ are the transformations for the parameters in the inferential model to the working model and $|J_{\mathbf{\Sigma}, \mathbf{\Psi}}(\cdot)|, |J_{\mathbf{\Psi}}(\cdot)|$ are respectively the determinants of the Jacobians of these transformations.

In the case where only $\mathbf{\Psi}$ is unrestricted, the PX algorithm uses the priors

$$\psi_q^{-2} \sim \text{Gamma}(\nu_\psi, \phi_\psi), \quad (\text{A.6})$$

$$\boldsymbol{\lambda}_j^{*'} \sim \text{N}(\mathbf{m}_{\lambda_j^{*'}}, \mathbf{S}_{\lambda_j^{*'}}) \quad (\text{A.7})$$

$$\boldsymbol{\eta}_i^* \sim \text{N}(\mathbf{0}, \mathbf{\Psi}). \quad (\text{A.8})$$

We further set $\mathbf{m}_{\lambda_j^{*'}} to 0. The transformations from the working model to the inferential model are$

$$\boldsymbol{\eta}_i = \mathbf{\Psi}^{-1/2}\boldsymbol{\eta}_i^*, \quad (\text{A.9})$$

$$\mathbf{\Lambda} = \mathbf{\Lambda}^*\mathbf{\Psi}^{1/2}. \quad (\text{A.10})$$

The prior distribution of $\boldsymbol{\eta}_i$ is then

$$\mathbf{\Psi}^{-1/2}\boldsymbol{\eta}_i^* \sim \mathbf{\Psi}^{-1/2}\text{N}(\mathbf{0}, \mathbf{\Psi}) \quad (\text{A.11})$$

$$=_d \text{N}(\mathbf{0}, \mathbf{I}_Q). \quad (\text{A.12})$$

In the above notation,

$$p_{\eta_{iq}^*}(t_{\psi_q}(\eta_{iq}))|J_{\psi_q}(\eta_{iq})| = (2\pi)^{-1/2}(\psi_q^{-2})^{1/2} \exp\left(-\frac{\psi_q^{-2}}{2}(\psi_q\eta_{iq})^2\right) \cdot \psi_q \quad (\text{A.13})$$

$$= (2\pi)^{-1/2} \exp\left(-\frac{1}{2}\eta_{iq}^2\right) \quad (\text{A.14})$$

which does not depend on ψ_q^{-2} and is therefore the prior density $p(\eta_{iq})$.

The prior distribution for λ_{jq} is

$$p(\lambda_{jq}) = \int p_{\lambda_{jq}^*}(t_{\psi_q}(\lambda_{jq})) |J_{\psi_q}(\lambda_{jq})| p(\psi_q^{-2}) d\psi_q^{-2} \quad (\text{A.15})$$

$$= \int (2\pi s_{\lambda_{jq}}^2)^{-1/2} \exp\left(-\frac{1}{2s_{\lambda_{jq}}^2}(\psi_q^{-1}\lambda_{jq})^2\right) \cdot \psi_q^{-1}. \quad (\text{A.16})$$

$$\begin{aligned} & \frac{\phi_{\psi_q}^{\nu_{\psi_q}}}{\Gamma(\nu_{\psi_q})} (\psi_q^{-2})^{\nu_{\psi_q}-1} \exp(-\phi_{\psi_q} \psi_q^{-2}) d\psi_q^{-2} \\ &= \frac{1}{(2\pi s_{\lambda_{jq}}^2)^{1/2} \Gamma(\nu_{\psi_q})} \phi_{\psi_q}^{\nu_{\psi_q}} \int (\psi_q^{-2})^{\nu_{\psi_q}-1/2} \exp\left(-\left(\frac{\lambda_{jq}^2}{2s_{\lambda_{jq}}^2} + \phi_{\psi_q}\right) \psi_q^{-2}\right) d\psi_q^{-2} \end{aligned} \quad (\text{A.17})$$

$$= \frac{\Gamma(\nu_{\psi_q} + \frac{1}{2})}{(2\pi s_{\lambda_{jq}}^2)^{1/2} \Gamma(\nu_{\psi_q})} \frac{\phi_{\psi_q}^{\nu_{\psi_q}}}{\left(\frac{\lambda_{jq}^2}{2s_{\lambda_{jq}}^2} + \phi_{\psi_q}\right)^{\nu_{\psi_q}+1/2}} \quad (\text{A.18})$$

$$\begin{aligned} & \cdot \int \frac{\left(\frac{\lambda_{jq}^2}{2s_{\lambda_{jq}}^2} + \phi_{\psi_q}\right)^{\nu_{\psi_q}+1/2}}{\Gamma(\nu_{\psi_q} + \frac{1}{2})} (\psi_q^{-2})^{\nu_{\psi_q}-1/2} \exp\left(-\left(\frac{\lambda_{jq}^2}{2s_{\lambda_{jq}}^2} + \phi_{\psi_q}\right) \psi_q^{-2}\right) d\psi_q^{-2} \\ &= \frac{\Gamma\left(\frac{2\nu_{\psi_q}+1}{2}\right)}{\Gamma\left(\frac{2\nu_{\psi_q}}{2}\right)} \left(\frac{1}{2\pi s_{\lambda_{jq}}^2 \phi_{\psi_q}}\right)^{1/2} \left(1 + \frac{\lambda_{jq}^2}{2s_{\lambda_{jq}}^2 \phi_{\psi_q}}\right)^{-\frac{2\nu_{\psi_q}+1}{2}}. \end{aligned} \quad (\text{A.19})$$

Thus, λ_{jq} is distributed according to a three-parameter t distribution with degrees of freedom equal to $2\nu_{\psi_q}$, location equal to 0, and scale set to $\sqrt{s_{\lambda_{jq}}^2 \phi_{\psi_q} / \nu_{\psi_q}}$.

In the case where Ψ and Σ are unrestricted (or rather their diagonals are unrestricted),

$$p(\lambda_{jq}) = \int p_{\lambda_{jq}^*}(t_{\psi_q, \sigma_j}(\lambda_{jq})) |J_{\psi_q, \sigma_j}(\lambda_{jq})| p(\psi_q^{-2}) p(\sigma_j^{-2}) d\psi_q^{-2} d\sigma_j^{-2} \quad (\text{A.20})$$

$$= \int p_{\lambda_{jq}^*}\left(\frac{\psi_q^{-1}}{\sigma_j^{-1}} \lambda_{jq}\right) \frac{\psi_q^{-1}}{\sigma_j^{-1}} p(\psi_q^{-2}, \sigma_j^{-2}) d\psi_q^{-2} d\sigma_j^{-2}. \quad (\text{A.21})$$

Let $w = \sigma_j^{-2} / \psi_q^{-2}$, $v = \psi_q^{-2}$ so that $\lambda_{jq}^* = w^{-1/2} \lambda_{jq}$. Then,

$$p(w) = \int v p_{\sigma_j^{-2}, \psi_q^{-2}}(wv, v) dv \quad (\text{A.22})$$

$$= \int v \frac{\phi_{\sigma_j}^{\nu_{\sigma_j}}}{\Gamma(\nu_{\sigma_j})} (wv)^{\nu_{\sigma_j}-1} \exp(-\phi_{\sigma_j} wv) \frac{\phi_{\psi_q}^{\nu_{\psi_q}}}{\Gamma(\nu_{\psi_q})} (v)^{\nu_{\psi_q}-1} \exp(-\phi_{\psi_q} v) dv \quad (\text{A.23})$$

$$= \frac{\phi_{\sigma_j}^{\nu_{\sigma_j}}}{\Gamma(\nu_{\sigma_j})} \frac{\phi_{\psi_q}^{\nu_{\psi_q}}}{\Gamma(\nu_{\psi_q})} w^{\nu_{\sigma_j}-1} \int v^{\nu_{\sigma_j}+\nu_{\psi_q}-1} \exp(-(\phi_{\sigma_j} w + \phi_{\psi_q}) v) dv \quad (\text{A.24})$$

$$= \frac{\Gamma(\nu_{\sigma_j} + \nu_{\psi_q})}{\Gamma(\nu_{\sigma_j}) \Gamma(\nu_{\psi_q})} \frac{\phi_{\sigma_j}^{\nu_{\sigma_j}} \phi_{\psi_q}^{\nu_{\psi_q}}}{(\phi_{\sigma_j} w + \phi_{\psi_q})^{\nu_{\sigma_j} + \nu_{\psi_q}}} w^{\nu_{\sigma_j} - 1} \quad (\text{A.25})$$

$$\cdot \int \frac{(\phi_{\sigma_j} w + \phi_{\psi_q})^{\nu_{\sigma_j} + \nu_{\psi_q}}}{\Gamma(\nu_{\sigma_j} + \nu_{\psi_q})} v^{\nu_{\sigma_j} + \nu_{\psi_q} - 1} \exp(-(\phi_{\sigma_j} w + \phi_{\psi_q}) v) dv \quad (\text{A.26})$$

$$= \frac{\Gamma(\nu_{\sigma_j} + \nu_{\psi_q})}{\Gamma(\nu_{\sigma_j}) \Gamma(\nu_{\psi_q})} \frac{\phi_{\sigma_j}^{\nu_{\sigma_j}}}{\phi_{\psi_q}^{\nu_{\sigma_j}} \left(1 + \frac{w}{\phi_{\psi_q}/\phi_{\sigma_j}}\right)^{\nu_{\sigma_j} + \nu_{\psi_q}}} w^{\nu_{\sigma_j} - 1} \quad (\text{A.27})$$

$$= \frac{\Gamma(\nu_{\sigma_j} + \nu_{\psi_q})}{\Gamma(\nu_{\sigma_j}) \Gamma(\nu_{\psi_q})} \left(\frac{w}{\phi_{\psi_q}/\phi_{\sigma_j}}\right)^{\nu_{\sigma_j} - 1} \left(1 + \frac{w}{\phi_{\psi_q}/\phi_{\sigma_j}}\right)^{-\nu_{\sigma_j} - \nu_{\psi_q}} \frac{\phi_{\sigma_j}}{\phi_{\psi_q}}. \quad (\text{A.28})$$

As a result, w is distributed as a generalized beta prime random variable with second shape parameter set to 1, or equivalently a compound gamma random variable (Johnson, Kotz, and Balakrishnan, 1994a,b; Dubey, 1970).

Returning to the induced prior for λ_{jq} , we now have

$$p(\lambda_{jq}) = \int p_{\lambda_{jq}^*} \left(w^{-1/2} \lambda_{jq}\right) w^{-1/2} p(w) dw \quad (\text{A.29})$$

$$= \int \left(2\pi s_{\lambda_{jq}}^2 w\right)^{-1/2} \exp\left(-\frac{1}{2s_{\lambda_{jq}}^2 w} \lambda_{jq}^2\right) p(w) dw. \quad (\text{A.30})$$

The induced prior for λ_{jq} takes the form of a scale mixture of normals with a compound gamma mixing density.

Appendix B

COMPARISON OF CORRELATION MATRIX DISTANCE MEASURES

This section details an informal and heuristic comparison of the correlation distance measures considered in Section 2.4 for posterior predictive model checks. To gain some better intuition about these measures, we consider an example with the following matrices,

$$\mathbf{A} = \begin{pmatrix} 1 & 0.4 & 0.3 \\ 0.4 & 1 & 0.2 \\ 0.3 & 0.2 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0.6 & 0.3 \\ 0.6 & 1 & 0 \\ 0.3 & 0 & 1 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 & 0.4 & 0.3 \\ 0.4 & 1 & 0.48 \\ 0.3 & 0.48 & 1 \end{pmatrix}. \quad (\text{B.1})$$

We compare matrix $\mathbf{A}$ to matrices $\mathbf{B}$ and $\mathbf{C}$ using the Euclidean, symmetric LogDet and CMD measures. The matrices were selected so that $d_E(\mathbf{A}, \mathbf{B}) = d_E(\mathbf{A}, \mathbf{C})$. Qualitatively, however, the relationship between $\mathbf{A}$ and $\mathbf{B}$ seems different than that between $\mathbf{A}$ and $\mathbf{C}$. $\mathbf{A}$ and $\mathbf{B}$ differ in two unique off-diagonal elements with one of the elements greater in $\mathbf{B}$ and one greater in $\mathbf{A}$. $\mathbf{A}$ and $\mathbf{C}$ on the other hand differ only in one unique off-diagonal element with that off-diagonal element in matrix $\mathbf{C}$ greater than the corresponding element in matrix $\mathbf{A}$ and closer in value to one of the other off-diagonal elements in matrix $\mathbf{C}$.

Table B.1 shows the results of the comparison. The Symmetric LogDet measure appears to be the only measure where there is a noticeable difference between $d(\mathbf{A}, \mathbf{B})$ and $d(\mathbf{A}, \mathbf{C})$, indicating greater similarity between $\mathbf{A}$ and $\mathbf{C}$ than between $\mathbf{A}$ and $\mathbf{B}$. To see if the differences in the matrices would manifest themselves in data generated under a multivariate normal with the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ as the covariance matrices, we employed the following procedure. We generated 10,000 random vectors of length 3 from a multivariate normal distribution with zero mean and covariance set as the identity matrix. We transform each vector three times so that the first transformation results in a randomly drawn vector from a multivariate normal with covariance matrix set to matrix $\mathbf{A}$ above, the second transformation results in a similar vector but with covariance matrix $\mathbf{B}$, and the third with

Table B.1: Comparison of dissimilarity measures for matrices **A**, **B** and **C**.

	$d.(A, B)$	$d.(A, C)$
Euclidean	0.2828	0.2828
Symmetric LogDet	0.4301	0.2647
CMD	0.0205	0.0199

covariance matrix **C**. We then have three sets of 10,000 multivariate normal vectors with covariances set to matrices **A**, **B**, and **C** respectively. We calculated the Euclidean distance between each vector in the set with covariance matrix **A** and the set with covariance matrix **B**. We also did the same with sets of vectors produced by covariance matrices **A** and **C**. The average Euclidean distance between the sets of vectors with covariance matrix **A** and covariance matrix **B** was 0.3373 whereas the average Euclidean distance between the sets of vectors with covariance matrix **A** and covariance matrix **C** is 0.2552. The standard error for each of these means is approximately 0.0019. These results appear to support the notion that **A** and **C** are closer together than **A** and **B** in that vectors generated from a multivariate normal distribution with **A** and **C** as the specified covariances are closer together (according to Euclidean distance) than multivariate normal vectors generated with **A** and **B** as the covariances. Moreover, these results suggest that there are differences in correlation matrices to which Euclidean distance and CMD may not be sensitive and to which symmetric LogDet may be sensitive.

VITA

Jonathan Gruhl earned his Bachelor of Science degree in Industrial Engineering from Stanford University in 1996. After working in management consulting and financial services, he obtained a Master of Science degree in Mathematics in Finance in 2004 from New York University. Jonathan subsequently worked in the quantitative analytics group at Arden Asset Management in New York, NY. In 2012, he earned his Doctor of Philosophy in Statistics from the University of Washington.