



University of Pennsylvania
ScholarlyCommons

Publicly Accessible Penn Dissertations

Fall 12-21-2011

Wrestling With Issues in Scale Development Using Joint Latent Variable Methods

Steffanie M. Halberstadt

University of Pennsylvania, steffhalb@gmail.com

Follow this and additional works at: <http://repository.upenn.edu/edissertations>

 Part of the [Biostatistics Commons](#)

Recommended Citation

Halberstadt, Steffanie M., "Wrestling With Issues in Scale Development Using Joint Latent Variable Methods" (2011). *Publicly Accessible Penn Dissertations*. 440.

<http://repository.upenn.edu/edissertations/440>

This paper is posted at ScholarlyCommons. <http://repository.upenn.edu/edissertations/440>

For more information, please contact libraryrepository@pobox.upenn.edu.

Wrestling With Issues in Scale Development Using Joint Latent Variable Methods

Abstract

Many applications of biomedical science involve unobservable constructs, from measurement of health states to severity of complex diseases. In this dissertation I utilize joint latent variable methods to combine item selection and validation to identify significant items in a symptom scale and determine how these symptoms relate to "gold standard" diagnostic measures. Joint latent variable models eliminate bias inherent in traditional two-stage methods and provide a global test of the association between the underlying construct and a clinical measure. In Chapter 1, a review of latent variable methods for multivariate outcomes is provided. Chapter 2 proposes a Multiple Indicator Multiple Cause (MIMIC) model to perform item reduction and validation simultaneously. A modified Score test for individual factor loadings in the MIMIC model is derived. The methods are motivated by an example from a premenstrual syndrome (PMS) clinical trial in which one objective was to determine a reduced number of core symptoms in the diagnosis of severe PMS and to compare patient-reported symptom information to a clinician-rated "gold standard" diagnostic measure. Chapter 3 applies an extension to the MIMIC model to patient-reported outcomes (PROs) from the Physical Activity and Lymphedema (PAL) clinical trial. PROs are a potentially less expensive and time-consuming measure of diagnosis than some clinical measures. An extension of the MIMIC model for ordered categorical outcomes determines which symptoms are important indicators of lymphedema and how these symptoms compare to clinical endpoints. Finally, in Chapter 4, a multivariate zero-inflated proportional odds (MZIPO) model is proposed to account for excess symptom non-response at baseline. This model adds a latent class component to the traditional MIMIC model. The MZIPO model is applied to the PAL data to obtain more accurate estimates of the latent construct and its association with current measures of lymphedema severity.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Epidemiology & Biostatistics

First Advisor

Mary D. Sammel

Keywords

Factor Analysis, Latent Class, Latent Variable, Multiple Indicator Multiple Cause, Multivariate, Zero-Inflation

Subject Categories

Biostatistics

WRESTLING WITH ISSUES IN SCALE DEVELOPMENT USING JOINT
LATENT VARIABLE METHODS

Steffanie M. Halberstadt

A DISSERTATION

in

Epidemiology and Biostatistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2011

Supervisor of Dissertation

*Signature*_____

Mary D. Sammel, Associate Professor of Biostatistics

Graduate Group Chairperson

*Signature*_____

John H. Holmes, Associate Professor of Medical Informatics in Epidemiology

Dissertation Committee

Ellen W. Freeman, Research Professor of Obstetrics and Gynecology

Knashawn H. Morales, Assistant Professor of Biostatistics

Kathryn H. Schmitz, Associate Professor of Epidemiology

Andrea B. Troxel, Professor of Biostatistics

Acknowledgment

This dissertation could not have been completed without the encouragement of many individuals in my life. First, I would like to acknowledge my dissertation committee and my advisor, Mary Sammel, in particular. Throughout my time at Penn, Mary provided me with insight and expertise that fostered my academic growth.

I would also like to acknowledge my fellow students for their laughter and friendship. Although Blockley is historically well-known as a place of misery, sordidness, and suffering, my classmates made it enjoyable. Specifically, I would like to thank Peter Dawson for being an ideas man, I think he proved that; Vica Gamerman for always being there when I needed a smile and a hug; and Matt Guerra for being my first friend in graduate school and, among other things, for teaching me that maple syrup enhances sausage above which bacon and maple syrup.

Finally I would like to acknowledge my family for always being supportive of my goals and Ben for providing me with the love and support necessary for me to become Dr. Hablerdazzler.

ABSTRACT

WRESTLING WITH ISSUES IN SCALE DEVELOPMENT USING JOINT LATENT VARIABLE METHODS

Steffanie M. Halberstadt

Mary D. Sammel, Advisor

Many applications of biomedical science involve unobservable constructs, from measurement of health states to severity of complex diseases. In this dissertation I utilize joint latent variable methods to combine item selection and validation to identify significant items in a symptom scale and determine how these symptoms relate to “gold standard” diagnostic measures. Joint latent variable models eliminate bias inherent in traditional two-stage methods and provide a global test of the association between the underlying construct and a clinical measure. In Chapter 1, a review of latent variable methods for multivariate outcomes is provided.

Chapter 2 proposes a Multiple Indicator Multiple Cause (MIMIC) model to perform item reduction and validation simultaneously. A modified Score test for individual factor loadings in the MIMIC model is derived. The methods are motivated by an example from a premenstrual syndrome (PMS) clinical trial in which one objective was to determine a reduced number of core symptoms in the diagnosis of severe PMS and to compare patient-reported symptom information to a clinician-rated “gold standard” diagnostic measure.

Chapter 3 applies an extension to the MIMIC model to patient-reported outcomes (PROs) from the Physical Activity and Lymphedema (PAL) clinical trial. PROs are a potentially less expensive and time-consuming measure of diagnosis than some clinical measures. An extension of the MIMIC model for ordered categorical outcomes determines which symptoms are important indicators of lymphedema and how these symptoms compare to clinical endpoints.

Finally, in Chapter 4, a multivariate zero-inflated proportional odds (MZIPO) model is proposed to account for excess symptom non-response at baseline. This model adds a latent class component to the traditional MIMIC model. The MZIPO model is applied to the PAL data to obtain more accurate estimates of the latent construct and its association with current measures of lymphedema severity.

Key Words: Factor analysis, Latent class, Latent variable, Multiple indicator multiple cause, Multivariate, Zero-inflation.

Contents

1	Introduction	1
1.1	Motivation: Symptom Scale Data in Clinical Trials	2
1.2	Latent Variable Models for Multivariate Outcomes	4
1.2.1	Latent Variable Models for Continuous Observed Outcomes	5
1.2.2	Latent Variable Models for Categorical Observed Outcomes	7
1.2.3	Extensions to Classical Latent Variable Models	7
2	A Joint Latent Variable Model for Item Reduction and Validation using Continuous Scale Items	9
2.1	Introduction	9
2.2	Methods	15
2.2.1	MIMIC Model Specification	17
2.2.2	Likelihood and Score Function	20
2.2.3	Score Test	21
2.3	Application to Premenstrual Syndrome Clinical Trial	22

2.4	Results	25
2.5	Simulations	28
2.6	Discussion	31
3	Application of Joint Latent Variable Model for Item Reduction and Validation for Ordered Categorical Scale Items	35
3.1	Introduction	35
3.2	Methods	38
3.2.1	MIMIC Model Specification	38
3.3	Application to Physical Activity and Lymphedema Clinical Trial . .	40
3.4	Results	43
3.5	Discussion	49
4	A Joint Latent Variable Model for Item Reduction and Validation in the Presence of a Preponderance of Zeros	51
4.1	Introduction	51
4.2	Model Specification	56
4.2.1	Measurement Model	57
4.2.2	Structural Model	58
4.2.3	Latent Class Model	59
4.2.4	Complete Data Likelihood	60
4.3	Model Estimation	62

4.3.1	Expectation Step	65
4.3.2	Maximization Step	69
4.3.3	Predictive validity	74
4.4	Example	75
4.5	Results	76
4.5.1	Ordered Categorical MIMIC Models vs. MZIPO Model . . .	78
4.5.2	Predictive Validity	82
4.5.3	Evaluation of Observed Zeros and Latent Class Membership	83
4.6	Discussion	85
5	Conclusions	90
A	Derivations for Score Test of Factor Loadings in MIMIC Model	96
A.1	Score Function for λ	96
A.2	Information Matrix	100
B	Example Mplus Code for Ordered Categorical MIMIC Model	107
C	Derivations for MZIPO Model	108
C.1	Score Function	108
C.2	Information Matrix	110

List of Tables

2.1	Descriptive statistics of Daily Symptom Report stratified by PMS diagnosis.	26
2.2	Standard psychometric measures of item selection for DSR symptoms.	27
2.3	Full MIMIC model for continuous items applied to PMS data. . . .	27
2.4	Simulation 1 results. Data generated under null hypothesis.	29
2.5	Simulation 2 results. Data generated under the alternative hypothesis.	30
3.1	Polychoric correlations of Norman Lymphedema Survey items. . . .	45
3.2	Standard psychometric measures of Norman Lymphedema Survey items.	45
3.3	Univariate cumulative probit models.	47
3.4	Ordered categorical MIMIC model applied to PAL data.	48
4.1	Baseline characteristics for complete data sample of PAL data. . . .	77
4.2	Ordinal MIMIC model vs. MZIPO model for Norman Lymphedema Survey items.	80

4.3	Comparison of latent class assignment and observed symptom response.	84
4.4	Characteristics of predicted latent class assignment and observed symptom response.	84

List of Figures

2.1	Path diagram of general MIMIC model for m continuous observed indicators and one observed cause. Boxes denote observed variables and oval denotes latent variable.	19
3.1	Mean volume difference plots with error bars for Norman Lymphedema Survey items.	44
4.1	Histograms of Norman Lymphedema Survey items for N=278 complete data sample.	79
4.2	Profile plot of mean volume difference by predicted latent class assignment.	83

Chapter 1

Introduction

Many applications of biomedical science involve unobservable constructs, from the measurement of health states to the severity of complex diseases. From a statistical perspective, the goal of measurement is often to combine important pieces of information in a way that thoroughly describes an unobservable construct.

In the scale development process, item selection or reduction determines which items best exemplify the construct of interest. It is undesirable to include “junk” items, any items that fail to contribute useful information about the hypothetical construct, because they obscure the final scale score. The process of removing unnecessary items ultimately improves scale accuracy, reduces burden to participants, and decreases research costs. Next, validation establishes the relationship between a particular scale and other measures of the unobservable concept of interest. Typically item selection and validation are performed separately using methods from

psychometric research. Joint latent variable methods are proposed here to combine item reduction and validation into a single statistical model.

This dissertation is presented as an illustration of the advantages of joint latent variable models and as an example of the applicability of these models to biomedical research. The methods are applied to data from innovative clinical trials in women's health. The first application involves the Penn Daily Symptom Report from a clinical trial for premenstrual syndrome (PMS) at the University of Pennsylvania. The second application involves the Norman Lymphedema Survey from the Physical Activity and Lymphedema (PAL) clinical trial at the University of Pennsylvania. Both motivating examples seek to evaluate current diagnostic tools and incorporate patient-reported symptom information into the diagnostic process.

1.1 Motivation: Symptom Scale Data in Clinical Trials

PMS

PMS is described as a collection of physical and emotional symptoms that present at the luteal phase of the menstrual cycle. Over 200 symptoms of PMS have been previously identified, yet there is no universal diagnosis criteria. Therefore, a variety of "gold standard" measures are used to diagnose PMS. Some include summary measures from patient-reported daily symptom diaries, while others are

based on clinician ratings of PMS severity. In Chapter 2, I identify and evaluate a core set of symptoms from a larger battery to thoroughly describe severe PMS and establish criterion validity by comparing this reduced set of symptoms to a “gold standard” diagnostic measure of PMS using data from a University of Pennsylvania PMS clinical trial directed by Dr. Ellen Freeman.

Lymphedema

For breast cancer survivors, lymphedema is a debilitating chronic condition resulting from the surgical excision of lymph nodes as part of cancer treatment. Its significant impact of daily functioning and quality of life makes lymphedema a serious concern for women recovering from breast cancer. Similar to PMS research, a complication in the study of lymphedema is that the estimates of incidence among breast cancer survivors in the literature vary dramatically, ranging from as small as 6% to as large as 70%. This discrepancy is likely due to the variety of diagnostic measures that are used, many of which may not necessarily capture the same attributes of lymphedema. Furthermore, patient-reported symptoms that could potentially prove useful as indicators of lymphedema are not typically used as diagnostic measures. The objective of Chapter 3 is to determine which items from the Norman Lymphedema Survey are the most important indicators of lymphedema and whether they perform as well as current “gold standard” diagnostic measures. Chapter 4 focuses on how to identify important lymphedema symptoms while simultaneously accounting for significant symptom non-response at baseline.

1.2 Latent Variable Models for Multivariate Outcomes

Although latent variable models originated in psychometrics and education research, their utility has increased recently in biomedical research. The following literature review provides a general overview of classic latent variable methods.

A latent variable is defined as a random variable that cannot be directly measured and instead is inferred through the measurement of other observable variables. Latent variables often represent underlying constructs that are difficult or impossible to quantify. The use of latent variables in the study of immeasurable constructs is extensive in education, psychometrics, and econometrics. For example, latent variable models are used in scale development and revision for high-stakes educational assessment; scales assessing constructs such as personality, happiness, or depression; and as a representation for paradigms like permanent income. Recent examples of biomedical applications for latent variable models include research on diagnostic tests, health-related quality of life scales, and even genetics. These models are flexible in that they accommodate a variety of observed response types and different types of latent variables. They provide a means of aggregating multiple observed variables in a single model and relating them to an intangible construct.

1.2.1 Latent Variable Models for Continuous Observed Outcomes

Factor analysis and structural equation models are two popular latent variable models for multivariate continuous observed variables. Used to describe variability among a set of continuous observed variables, the factor analysis model separates variance among observed variables into variance due to a common factor, or unobserved latent construct, and the residual variance due to the specific observed manifest variables. The classic factor analysis model assumes the latent variable is a continuous, normally distributed random variable. Observed variables are modeled as a linear combination of latent factors and error terms for specific variance. Factor analysis is commonly used to determine which items are the strongest indicators of a particular construct. Factor loadings, the parameters of interest in the factor analysis model, measure the association between an item and the latent factor. Large standardized factor loadings are desired because they indicate that the item is strongly associated with the underlying construct. Other measures involving factor loadings can also help to determine the value of an item, such as the proportion of variance in the outcome explained by the latent variable. This measure can be calculated for each item and provides an estimate of the item's importance in a manner not dependent on the scale.

Factor analysis can be exploratory or confirmatory in nature, and the difference lies in the assumptions made about the structure of the latent factors. Exploratory

factor analysis (EFA) identifies the number and nature of the factors. The ultimate purpose of factor analysis is often to reduce the number of observed variables, so EFA is a natural choice for item selection. Confirmatory factor analysis (CFA) serves as a separate validation procedure once the factor structure has been established. CFA differs from EFA in that it allows for more constraints on the factor structure. For example, in multifactor CFA models, a “clean” solution is frequently assumed where certain factor loadings are set to zero such that each manifest variable loads on only one latent factor. In the methods developed in Chapter 2 and applied in Chapter 3, a CFA model is one of two components of a joint latent variable model developed for item selection and validation.

Structural Equation Modeling

Structural equation modeling extends the factor analysis model to include structural relations among latent variables as well as covariates on latent or observed variables. Originally developed for continuous observed outcomes, structural equation models (SEMs) are two-part models containing both a measurement and a structural model. The measurement model is simply a CFA model with one or more continuous latent variables measured by several continuous items. The structural model specifies associations of latent variables with other observed or potentially latent variables. SEMs feature both latent exposure variables and latent outcome variables, as well as observed exposure and outcome variables. Structural relationships among the latent variables may be formulated as a regression relationship or in terms of correlated

residuals among the latent variables. The models proposed in Chapter 2 can be thought of as an example of an SEM with a single latent variable.

1.2.2 Latent Variable Models for Categorical Observed Outcomes

Latent Class Models

A common latent variable for categorical observed outcomes, the latent class model classifies subjects into unobserved subgroups. In comparison to latent variable models with continuous latent variables like the factor analysis model or SEM, the latent class model contains a discrete latent variable with C categories. Useful for describing population heterogeneity, latent class models assume that any observed responses originate from an underlying discrete latent class. These models have particular relevance to clinical studies because diagnostic decisions can potentially be made based on subgroup classification. In Chapter 4, a latent class component is incorporated into the model to classify subjects into subpopulations based on symptom response or non-response.

1.2.3 Extensions to Classical Latent Variable Models

The models presented above represent the most basic latent variable models that set the foundation for the many advanced latent variable models that have been developed over the past several decades. Extensions to classical models have been for-

mulated for a variety of response types including multilevel, longitudinal, survival, or mixed outcome types. Of particular relevance to this dissertation is the class of latent variable models known as latent variable hybrid models. The confluence of a continuous latent variable model and a latent class model, these models accommodate both cross-sectional and longitudinal outcomes and allow for covariates. Crucial features of these models are that they allow for classification of subjects into unobserved subgroups and also provide factor scores, estimates of the latent variable, within the subgroups. In this dissertation I extend the latent variable hybrid literature in the situation of zero-inflated outcomes.

The remainder of the dissertation is structured as follows: Chapter 2 presents a joint latent variable model for continuous data. A multiple indicator multiple cause model is employed to perform item reduction and validation simultaneously, and a modified Score test for individual factor loadings is developed. This work is illustrated with an example from a University of Pennsylvania PMS clinical trial. Chapter 3 presents an application of an extension to the MIMIC model for ordered categorical outcomes and is illustrated with an example from the PAL clinical trial. Chapter 4 proposes a new joint latent variable model for item reduction and validation in the presence of zero-inflation, which is also applied to the PAL data. Finally, Chapter 5 presents conclusions and future work.

Chapter 2

A Joint Latent Variable Model for Item Reduction and Validation using Continuous Scale Items

2.1 Introduction

This chapter presents a joint latent variable approach to item reduction and validation for continuous data. I employ a multiple indicator multiple cause (MIMIC) model with the objective of identifying important items in a symptom scale and comparing these items to a physician-rated “gold standard” diagnostic measure. The methods in this chapter are illustrated with an example from PMS.

Methods for item reduction originated in psychometrics and education where

constructs such as depression and intelligence provided motivation for the development of measurement techniques to quantify and explain these entities via multiple items in a scale. The literature in these fields offers a variety of latent variable methods that utilize the correlation among items to define a construct. Classical test theory (CTT), a popular psychometric method for item selection, employs measures such as item-total correlations and Cronbach's alpha to judge items (Nunnally et al., 1967; Clark and Watson, 1995). Item-total correlation measures the correlation of a particular item with the scale total when that item is omitted from the scale. High item-total correlations are desired and it is advised that items whose item-total correlation is less than 0.2 be dropped from the scale (Streiner and Norman, 1994). Cronbach's alpha is another popular measure of reliability and it is often advised that alpha be between 0.7 and 0.9. Values of alpha below 0.7 indicate that the items may not be homogeneous and values above 0.9 can be indications of several problems with the scale, such as item redundancy or the presence of more than one distinct construct.

Perhaps the most common statistical model for item reduction is the factor analysis model (Spearman, 1904). As described in Chapter 1, the model separates the variance among a set of observed variables into variance due to a latent factor and variance due to individual observed variables. Item reduction in factor analysis is done by evaluating factor loadings across multiple factors or analyzing the loadings within a particular factor. Factor loading parameters yield information about the

strength of the relationship between an individual item and the underlying construct.

Creating decision rules presents a unique challenge in item selection. Although many arbitrary rules exist, there seems to be little quantitative justification for what constitutes a sufficiently high factor loading, item-total correlation, or Cronbach's alpha. I focus on factor loadings as a primary measure of interest because they are natural measures of the contribution of an item to the latent construct.

Item selection in factor analysis is performed by assessing the magnitude of factor loadings, but a challenge is determining when they are sufficiently large. A review of the literature reveals many different methods, some of which are contradictory to one another. Shortly after the development of factor analysis, Thurstone (1938) proposed to interpret factor loadings on the correlation scale greater than 0.40 as significant and to disregard loadings less than 0.20. Comrey and Lee (1992) offered the following scale: loadings greater than 0.71 are excellent, 0.63 are very good, 0.55 are good, 0.45 are fair, and loadings less than 0.32 are poor. More recently, several authors suggested treating loadings greater than 0.30 as significant (Tabachnick et al., 2001). Cudeck and O'Dell (1994) mentioned that it is common in factor analysis studies with more than one factor to examine the loadings for each item, choose the largest one for each item and disregard the others in order to obtain a "clean" solution. Similarly, Clark and Watson (1995) suggested that items that load highly on the first factor and weakly on subsequent factors are ideal candidates

for consideration and should be chosen after a further correlational and reliability analysis is performed to identify redundant pairs of items. Clearly there is little agreement on what constitutes a significant loading. Furthermore, papers that suggest reference values generally do not provide justification for these choices.

In an effort to make the process of analyzing factor loadings more statistically rigorous, standard errors for factor loadings have been derived to allow for the assessment of variance in the estimated loadings. Lawley and Maxwell (1962) developed asymptotic variances and covariances for unrotated factor loadings for maximum likelihood factor analysis. Subsequently, formulas for various types of unrotated (Jennrich, 1974; Jennrich and Thayer, 1973) and rotated (Archer and Jennrich, 1973; Jennrich and Thayer, 1973; Jennrich, 1974; Cudeck and O'Dell, 1994) loadings were developed. There is a substantial amount of literature devoted to the topic of standard errors in factor analysis and structural equation models given varying assumptions about the distribution of the data, the number of underlying factors and whether the loadings are rotated or unrotated.

Cudeck and O'Dell (1994) proposed one of the few methods of developing a statistical test using standard errors of factor loadings. They provided a thorough review of methods for estimating standard errors and provided some new results for standard errors of rotated loadings. Ogasawara (1996, 1998, 1999) provided standard error estimates for orthomax, promax, and procrustes rotations in a method similar to Jennrich (1974) assuming normality and using an augmented information

matrix. Using a new approach, Hayashi and Kumar Sen (1998) found a matrix form for the covariance matrix of factor loading estimates under the normality assumption. Furthermore, Hayashi and Yung (1999) developed standard errors for factor loadings with an orthomax rotation using the delta method. In addition to the development of standard errors in the case of normal data, Yuan et al. (2002) reviewed standard errors for factor loadings in the presence of missing data, non-normal data, and outliers.

Unfortunately, in most studies that employ factor analysis as a main statistical method, the use of standard errors in determining the significance of factor loadings is far less popular than other *ad hoc* techniques. Furthermore, conventional Wald tests using the standard errors are not always valid. In the model presented in this chapter, factor loadings were required to be non-negative so that items only contributed positively to the estimate of the latent variable. Consequently, the statistical test of a factor loading amounted to a test of a variance component on the boundary of the parameter space. The traditional Wald test was no longer appropriate in this case. This issue is described in detail in Section 2.2.

Once the selection of important items is complete, a process of validation determines how well the scale measures the intended construct of interest. Establishing validity in scale development can be performed in a number of ways, including comparing the model to other models measuring the same construct or comparing a single model on different samples. In this dissertation, establishing criterion va-

lidity was desired. Streiner and Norman (1994) describe criterion validity as the correlation of a scale with another measure of the underlying construct, such as a “gold standard” that has been previously studied or is accepted in the field. Frequently, validation is performed separately from item selection. Unfortunately, the two-stage procedure of item selection and validation ignores additional measurement error inherent in the estimation of the factor (Sammel and Ryan, 1996). In comparison to other approaches such as a multivariate linear mixed model, there is evidence that a two-stage approach leads to bias toward the null.

A structural equation model can be used to combine item selection and validation. It is not only possible to model the association between items and a latent construct but also associations between the validation measure and individual items. To explore the efficiency of performing item selection and validation in a single model, I employed a MIMIC model (Joreskog and Goldberger, 1975) in this chapter. This flexible model is an example of a general structural equation model with one latent variable. As in factor analysis, factor loading estimates from the model provide information on how closely the items are correlated with the latent construct. Additionally, information about the relationship between the items and the validation metric is revealed through the regression parameter on the latent variable. In order to identify unimportant items, I developed a univariate Score test for estimated factor loadings of the MIMIC model. There are several advantages to this model. First, validation is incorporated directly into model development as

opposed to CTT, IRT, or simple factor analysis models where validation occurs as in a second stage. Second, the statistical test of individual factor loadings derived here reflects the assumed model constraints.

A University of Pennsylvania PMS clinical trial was considered as a motivating example. The MIMIC model allowed for both the identification of symptoms from a scale that were ultimately important contributors to a latent measure of PMS severity and the comparison of this core set of symptoms to the “gold standard” diagnostic measure. The primary aim was to produce a clinically relevant set of symptoms that discriminated subjects with severe PMS from those without.

The remainder of this chapter is outlined as follows. Section 2.2 illustrates the MIMIC model and describes model constraints, Section 2.3 describes the PMS clinical trial, Section 2.4 presents results of the PMS example, Section 2.5 presents simulations and results, and Section 2.6 describes our conclusions.

2.2 Methods

An extension to the factor analysis model, the MIMIC model consists of a system of structural equations including both observed indicators and observed causes of a hypothesized latent variable. Observed indicators are random variables assumed to have been generated by the latent variable, similar to the items or manifest variables in a factor analysis model. Observed causes are either fixed or random variables that influence the latent variable, similar to covariates in a regression setting. The single

latent variable is measured by the set of observed indicators and is regressed on the set of observed causes (Zellner, 1970; Hauser and Goldberger, 1971; Sammel and Ryan, 1996). Sammel and Ryan (2002) demonstrate that a test of the regression parameter is a global test of the observed cause on all observed indicators. The MIMIC model can be considered an extension to the factor analysis model because it allows for the inclusion of covariates that serve to validate the latent variable. The model also measures the relationship between covariates and indicator items through the latent variable. In the motivating example, the MIMIC model is formulated such that the single latent variable is interpretable as a continuum of PMS severity. It is natural to constrain factor loading parameters to be non-negative so that each item is a positive addition to the severity score. Allowing negative factor loadings would obscure the interpretation of the severity score allowing it to be a combination of positive and negative standardized items.

A modified Score test was developed to evaluate the significance of individual factor loadings of the MIMIC model under model constraints. Unlike traditional *ad hoc* methods of assessing factor loadings, this statistical test was chosen because it not only accounts for the estimate of the factor loading but also for its variance. A Score test was chosen over a Wald test provided by standard software because a Wald test is no longer appropriate given the model was constrained to allow for positive factor loadings only. Because of this constraint, the test is considered a test of a variance component on the boundary of the parameter space. Details of

the Score test are described below.

2.2.1 MIMIC Model Specification

There are two components to the general MIMIC model: a measurement model that specifies the factor analysis model and a structural model that specifies the regression of the latent variable on the observed causes (see Figure 2.1). Consider a sample of n individuals from whom a set of m outcome measurements, y_{i1}, \dots, y_{im} , and one validation measurement, z_i , are taken. The MIMIC model is written as follows. For subject $i = 1, \dots, n$ and outcome measurements $j = 1, \dots, m$ the measurement model is specified as

$$\mathbf{y}_{i(m \times 1)} = \boldsymbol{\mu}_{(m \times 1)} + \boldsymbol{\lambda}_{(m \times 1)} b_{i(1 \times 1)} + \boldsymbol{\epsilon}_{i(m \times 1)} \quad (2.2.1)$$

where $\boldsymbol{\mu}$ is a vector of means, $\boldsymbol{\lambda}$ is a vector of factor loading parameters, b_i is the latent variable, and $\boldsymbol{\epsilon}_i$ is a vector of specific variances. The structural model with a single “gold standard” is specified as

$$b_{i(1 \times 1)} = \beta_{(1 \times 1)} z_{i(1 \times 1)} + \delta_{(1 \times 1)} \quad (2.2.2)$$

where β is the regression coefficient for the “gold standard” and δ is an error term for the latent variable. Additional assumptions include

$$E(\boldsymbol{\epsilon}) = 0, \quad Cov(\boldsymbol{\epsilon}) = \boldsymbol{\Psi}, \quad Cov(b, \boldsymbol{\epsilon}) = 0, E(\delta) = 0,$$

$$Var(\delta) = \sigma^2, \quad Cov(\delta, z) = 0, \quad Cov(\delta, \boldsymbol{\epsilon}) = 0, Cov(z, \boldsymbol{\epsilon}) = 0$$

where $\boldsymbol{\Psi}$ is a diagonal matrix. This model assumes that the error terms in the measurement model are uncorrelated with the observed causes in the structural model and with the error term in the structural model. Additionally, the error term in the structural model is uncorrelated with the observed causes. Equations (2.2.1) and (2.2.2) imply the following marginal model for \mathbf{y}_i

$$f(\mathbf{y}_i | \boldsymbol{\mu}, \boldsymbol{\lambda}, \beta, \boldsymbol{\Psi}) \sim MVN(\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i, \boldsymbol{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}^\top + \boldsymbol{\Psi}). \quad (2.2.3)$$

Additionally, the constraint that each of the factor loadings must be non-negative, i.e., $\lambda_j > 0$, is placed on the model.

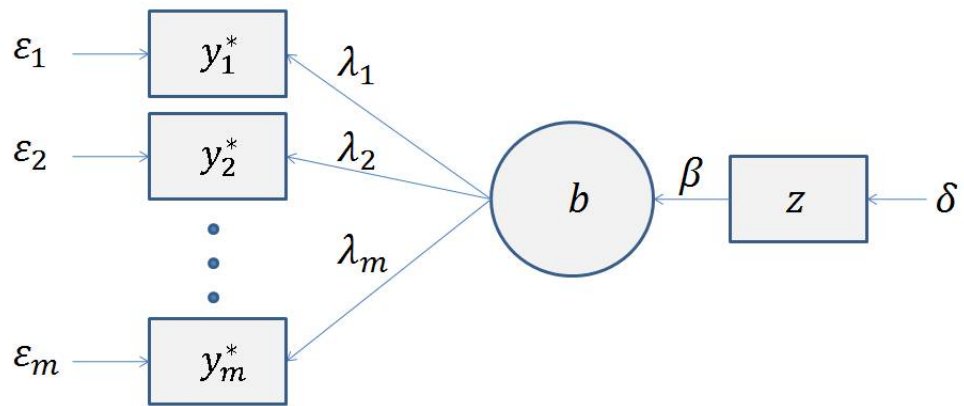


Figure 2.1: Path diagram of general MIMIC model for m continuous observed indicators and one observed cause. Boxes denote observed variables and oval denotes latent variable.

2.2.2 Likelihood and Score Function

For the set of parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\lambda}, \beta, \boldsymbol{\Psi})^\top$, the log-likelihood for the model from Equation (2.2.3) is expressed as

$$\begin{aligned} \ln L(\mathbf{y}; \boldsymbol{\theta}) &= \frac{np}{2} \ln(2\pi) - \frac{n}{2} \ln(|\boldsymbol{\Sigma}|) \\ &\quad - \frac{1}{2} \sum_{i=1}^n [\{\mathbf{y}_i - (\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i)\}^\top \boldsymbol{\Sigma}^{-1} \{\mathbf{y}_i - (\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i)\}]. \end{aligned} \quad (2.2.4)$$

Taking the partial derivative of the log likelihood with respect to $\boldsymbol{\lambda}$, the vector of factor loadings, yields the following portion of the score function

$$\begin{aligned} U_{\boldsymbol{\lambda}} = \frac{\partial l(\mathbf{y}; \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} &= -\frac{n}{2} [2\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \circ \mathbf{I}_m] \star [\mathbf{E}_{(m,1)} \boldsymbol{\lambda}^\top + \mathbf{I}_m \otimes \boldsymbol{\lambda}] \\ &\quad - \frac{1}{2} \sum_{i=1}^n 2[\boldsymbol{\Sigma}^{-1} \beta z_i \{\mathbf{y}_i - (\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i)\} \\ &\quad - \{((\mathbf{y}_i - (\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i))^\top \otimes \mathbf{I}_m)(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_m) \times \\ &\quad (\mathbf{E}_{(m,1)} \boldsymbol{\lambda}^\top + \mathbf{I}_m \otimes \boldsymbol{\lambda}) \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - (\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i))\}] \end{aligned} \quad (2.2.5)$$

where $(\boldsymbol{\Sigma}^{-1} \circ \mathbf{I}_m)$ denotes the elementwise Hadamard product (Styan, 1973). The star product, \star , is used as in MacRae (1974), and $\mathbf{E}_{(m,1)}$ refers to a permuted identity matrix (Rogers, 1980). The Fisher information matrix was derived by evaluating the negative expectation of the matrix of second partial derivatives. The information matrix was simplified by assuming the null hypothesis $H_0 : \lambda_j = 0$ was true.

2.2.3 Score Test

The strength of the relationship between each observed indicator and the underlying latent construct was evaluated by testing each of the factor loadings in the MIMIC model. The null hypothesis can be written as

$$H_0 : \lambda_j = 0 | \boldsymbol{\mu}, \beta, \boldsymbol{\Psi}, \boldsymbol{\lambda}_{j \neq k}. \quad (2.2.6)$$

The Score test for the null hypothesis resulted in the univariate test statistic

$$T_{\lambda_j} = U_{\lambda_j}^2(0) V_{j,j} |_{\beta=\hat{\beta}, \boldsymbol{\mu}=\hat{\boldsymbol{\mu}}, \boldsymbol{\Psi}=\hat{\boldsymbol{\Psi}}, \boldsymbol{\lambda}_{j \neq k}=\hat{\boldsymbol{\lambda}}_{j \neq k}}, \quad (2.2.7)$$

where $U_{\lambda_j}^2(0)$ is the square of the score function of λ_j evaluated under the null hypothesis and $V_{j,j}$ is the element of the inverse of the Fisher information matrix corresponding to λ_j evaluated under the null hypothesis. The parameters $\hat{\beta}$, $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\lambda}}_{j \neq k}$ were estimated using maximum likelihood.

Because of the constraint that λ_j must be non-negative, the parameter space for λ_j was $[0, \infty)$. The test of $H_0 : \lambda_j = 0$ was considered a test of a variance component on the boundary of the parameter space. Therefore, a one-sided Score test that did not have the traditional chi-square distribution was used (Zhang and Lin, 2008). The asymptotic null distribution of the test statistic was a mixture of

chi-square distributions; that is,

$$P(T_{\lambda_j} > c|H_0) = \frac{1}{2}P(\chi_1^2 > c) + \frac{1}{2}P(\chi_0^2 > c) \quad (2.2.8)$$

(Verbeke and Molenberghs, 2003, p. 256). Testing each of the factor loadings allowed for the assessment of whether an item appeared to be uncorrelated with the latent variable and could be considered for removal from the scale. In order to develop a Score test for the MIMIC model, the score function and information matrix of the MIMIC model were derived (see Appendix A) and SAS 9.2 IML was used to evaluate the inverse of the information matrix.

2.3 Application to Premenstrual Syndrome Clinical Trial

Currently, there is little consensus among the medical community regarding a definition of PMS. There are differences in the diagnostic criteria given by the American College of Obstetricians and Gynecologists, the World Health Organization, and the American Psychological Association. A literature review found over 200 symptoms cited as indicators of PMS. One aim of a University of Pennsylvania PMS clinical trial was to reduce the set of symptoms in the Penn Daily Symptom Report (DSR), a validated scale that measures the severity of 17 PMS symptoms (Freeman et al., 1996). Item reduction was an important step in determining a symptom profile for

PMS that discriminated well among patients with differing levels of severity. Furthermore, identifying the core symptoms that describe severe PMS could be useful in clinical practice, but only if symptoms could be shown to be as effective a means of diagnosis as the current “gold standard” diagnosis for severe PMS. Traditional methods (factor analysis, IRT models) base selection on the correlations among a set of variables and do not consider how the proposed latent constructs relate to a “gold standard” or other known variables. It is proposed here that better constructs are developed by combining the identification and validation stages into a joint model.

The sample for the study came from $N = 684$ women with complete data who participated in one of three similar PMS clinical trials at the University of Pennsylvania between 1994 and 2007. Women sought medical treatment for PMS symptoms and were screened for inclusion in the trials. The data presented here came from the screening pre-treatment portion of the trials, which included three standardized menstrual cycles for each woman. PMS symptoms were measured using the DSR. Each symptom was measured on a 5-point Likert scale from 0 = None to 4 = Severe. Symptoms included irritability/anger, mood swings, anxiety/tension, depression, feeling out of control, feeling worthless/guilty, decreased interest in usual activities, poor coordination, insomnia, difficulty concentrating/confusion, fatigue, aches, headache, cramps, breast tenderness, swelling/bloating, and food cravings/increased appetite. The scale aimed to measure several domains of symp-

tom severity, including physical, emotional, and behavioral. Premenstrual symptom scores were calculated by summing the daily symptom score over the 6-day premenstrual period for each symptom and were treated as normally distributed continuous measures.

The “gold standard” severe PMS outcome measure was created using data from the second untreated menstrual cycle. A woman was considered to have severe PMS based on her scores of the clinical global impression (CGI) scale (Guy, 1976). CGI is reported as a secondary outcome in many clinical trials of PMS (Freeman et al., 1999, 2001, 2004) and incorporates a variety of information into a single global measure meant to reflect the overall status of the patient. CGI is measured on a Likert scale with seven categories with higher scores indicating greater severity. PMS diagnosis was based on having a CGI score of five or greater. In the MIMIC model analysis, this “gold standard” PMS diagnosis served as the observed cause and the 17 summed DSR symptom scores served as the observed indicators. We assumed $Var(\delta_i) = 1$, allowing the latent variable to be interpreted on a standardized scale. All analysis was performed on the covariance scale, not the correlation scale. Parameter estimates for the MIMIC model were obtained using PROC CALIS in SAS 9.2.

2.4 Results

The prevalence of PMS as measured by the “gold standard” was about 70% in the sample ($N = 476$). For the full sample, mean item DSR scores ranged from 4.45 for cramps to 12.28 for anxiety (See Table 2.1). Generally larger mean DSR scores were found among mood items, including irritability (mean DSR = 12.17) and mood swings (mean DSR = 12.10). Smaller average DSR scores were found among physical symptoms, such as headaches (mean DSR = 5.70) and lack of coordination (mean DSR = 5.06). There were significant differences in the symptom means between those with severe PMS and those without across all 17 symptoms. Average DSR scores were significantly greater in the severe PMS group. When considered individually, all symptoms were significantly associated with the “gold standard” in univariate linear regressions ($p < 0.0001$ for all).

Correlations among all items were moderate, ranging in size from $\rho = 0.233$ between anxiety and cramps to $\rho = 0.791$ between anxiety and irritability. The largest correlations were found among mood items, while the correlations between mood and physical items tended to be smaller. In general, correlations between two physical items were not as large as the correlations between two mood items.

Table 2.2 provides standard psychometric measures for DSR items. Item-total correlations ranged in magnitude from 0.467 for breast tenderness to 0.749 for mood swings. According to the 0.2 criteria (Streiner and Norman, 1994) none of the items were immediately be candidates for removal. All Cronbach’s alpha values were

Table 2.1: Descriptive statistics of Daily Symptom Report stratified by PMS diagnosis.

Symptom	Full Sample (N=684)		PMS (N=476)		No PMS (N=208)		P
	Mean	SD	Mean	SD	Mean	SD	
Fatigue	11.25	6.193	13.116	5.743	6.981	4.955	<0.0001
Poor Coordination	5.063	6.187	6.347	6.656	2.125	3.494	<0.0001
Hopeless	9.958	6.785	12.013	6.332	5.255	5.286	<0.0001
Guilty	7.202	6.845	8.884	6.972	3.351	4.653	<0.0001
Headache	5.703	5.909	6.857	6.183	3.063	4.174	<0.0001
Anxiety	12.279	6.467	14.319	5.761	7.611	5.509	<0.0001
Aches	7.297	6.798	8.721	6.967	4.038	5.078	<0.0001
Irritability	12.165	6.374	14.143	5.750	7.639	5.351	<0.0001
Mood Swings	12.095	6.730	14.275	5.980	7.106	5.608	<0.0001
Weight Gain	11.133	7.093	12.868	6.809	7.163	6.070	<0.0001
Food Cravings	10.994	6.922	12.710	6.761	7.067	5.559	<0.0001
No Interest in Activities	8.444	6.906	10.391	6.783	3.990	4.812	<0.0001
Cramps	4.446	5.964	5.361	6.43	2.351	4.016	<0.0001
Depression	9.921	6.970	11.918	6.731	5.351	5.117	<0.0001
Breast Tenderness	8.558	7.485	9.849	7.680	5.606	6.078	<0.0001
Insomnia	7.171	6.931	8.578	7.184	3.952	5.011	<0.0001
Difficulty Concentrating	7.950	6.858	9.754	6.893	3.822	4.643	<0.0001

similar to one another and all exceeded 0.9. Although this is a common occurrence, it could indicate the presence of item redundancy.

Table 2.3 displays the results of the full MIMIC model for all 17 symptoms. Estimates of factor loadings, on the covariance scale, ranged from 2.14 for cramps to 4.45 for mood swings. As seen in previous results, estimates for anxiety, irritability and mood swings were larger than the estimates for many of the physical symptoms. The estimate of the regression coefficient on the latent variable, $\hat{\beta}$, was 1.55. This estimate can be interpreted as the mean of the latent variable, PMS severity, for subjects with diagnosed severe PMS according to the “gold standard” measure. The latent variable was standardized such that $Var(\delta_i) = 1$, which implies that subjects

Table 2.2: Standard psychometric measures of item selection for DSR symptoms.

Symptom	Item-Total Correlation	Cronbach's Alpha
Fatigue	0.7083	0.9243
Poor Coordination	0.6538	0.9256
Hopeless	0.7208	0.9240
Guilty	0.6468	0.9258
Headache	0.5024	0.9292
Anxiety	0.7103	0.9242
Aches	0.5845	0.9273
Irritability	0.7173	0.9241
Mood Swings	0.7491	0.9233
Weight Gain	0.6167	0.9265
Food Cravings	0.5802	0.9274
No Interest in Activities	0.7251	0.9239
Cramps	0.4782	0.9298
Depression	0.6956	0.9246
Breast Tenderness	0.4666	0.9301
Insomnia	0.5782	0.9274
Difficulty Concentrating	0.7151	0.9241

without severe PMS had a mean latent severity score of 0.

Table 2.3: Full MIMIC model for continuous items applied to PMS data.

Symptom	$\hat{\lambda}_j$	T_{λ_j}	Corrected P
Fatigue	3.631	127.009	<0.0001
Poor Coordination	3.310	1.697	0.0963
Hopeless	4.390	57.488	<0.0001
Guilty	3.888	3.208	0.0366
Headache	2.342	0.003	0.4765
Anxiety	4.176	242.999	<0.0001
Aches	3.006	8.185	0.0021
Irritability	4.112	239.628	<0.0001
Mood Swings	4.447	242.065	<0.0001
Weight Gain	3.429	162.829	<0.0001
Food Cravings	3.294	144.294	<0.0001
No Interest in Activities	4.188	17.345	<0.0001
Cramps	2.137	0.163	0.3431
Depression	4.209	66.623	<0.0001
Breast Tenderness	2.727	48.008	<0.0001
Insomnia	3.212	6.210	0.0064
Difficulty Concentrating	4.132	12.945	0.0002

Although most items in the full model were highly significant according to the corrected p-value, several items were candidates for elimination based on the Score

test. In particular, lack of coordination ($T = 1.697$, $p = 0.0963$), headaches ($T = 0.003$, $p = 0.4765$), and cramps ($T = 0.163$, $p = 0.3431$) were all non-significant in the full model. The most significant items in the full model were anxiety ($T = 242.9$, $p < 0.0001$), irritability ($T = 239.6$, $p < 0.0001$), and mood swings ($T = 242.1$, $p < 0.0001$).

2.5 Simulations

Simulations were performed to assess the behavior of the Score test of the MIMIC model in comparison to the Wald test given by PROC CALIS under several conditions. A smaller set of items was chosen for simulations due to computing time. This set of items was chosen to represent three highly correlated and significant mood items and one “junk” item that was a candidate for elimination. The MIMIC model also included the “gold standard” diagnosis measure for PMS. Simulation 1 investigated the performance of the Score test under the null hypothesis, in which the “junk” item truly contributed no information to the model. In this simulation, three multivariate normal random variables were generated using the observed moments of anxiety, irritability, and mood swings from the original sample. Simulated data were generated such that there was a one standard deviation shift in the mean of the latent variable between subjects with PMS and no PMS to reflect the association between the correlated items and the gold standard. An additional normal random variable was generated with the mean and standard deviation of cramps

in the original sample. This item was uncorrelated with the simulated mood items and was not associated with the gold standard. In each simulation, 1000 bootstrap samples were taken from a simulated dataset of $N = 684$. Type I error of the Score and Wald tests for the simulated “junk” item was calculated for various levels of correlation among the simulated mood items.

Table 2.4: Simulation 1 results. Data generated under null hypothesis.

ρ_{Mood}	Score Test	Wald Test
	Type I Error Rate	Type I Error Rate
0.8	0.011	0.062
0.6	0.021	0.077
0.5	0.061	0.058
0.4	0.095	0.075
0.3	0.101	0.079
0.15	0.187	0.110
0.075	0.241	0.091

Simulation 2 investigated the performance of the Score and Wald tests under the alternative hypothesis, where the “junk” item was truly correlated with the latent variable and was associated with the “gold standard.” For this simulation four multivariate normal random variables were generated with the means and standard deviations of the items from the original dataset. Each of the simulated items was associated with the gold standard such that there was a 0.5 standard deviation increase in the mean of the latent variable for those with severe PMS. Similar to Simulation 1, for each simulation, 1000 bootstrap samples were taken from the simulated dataset and MIMIC models were fit to each. Power of the Score and Wald tests was calculated for each simulation. Correlations among the simulated mood items and between the junk item and the mood items were varied for Simulation 2.

Table 2.5: Simulation 2 results. Data generated under the alternative hypothesis.

ρ_{Mood}	ρ_{Junk}	Score Test Power	Wald Test Power
0.6	0.60	1.000	1.000
0.6	0.30	1.000	1.000
0.6	0.15	1.000	1.000
0.6	0.10	1.000	1.000
0.6	0.05	0.727	0.588
0.6	0.00	0.855	0.763
0.3	0.60	1.000	1.000
0.3	0.30	1.000	1.000
0.3	0.15	1.000	1.000
0.3	0.10	1.000	1.000
0.3	0.05	1.000	1.000
0.3	0.00	0.853	0.844

Table 2.4 displays results from Simulation 1, where the “junk” item was not correlated with the mood items or associated with the gold standard. When the correlation between the mood items was set relatively high, at $\rho_{Mood} = 0.6$ or above, Type I error rate of the Score test was much better than that of the Wald test. For example, at $\rho_{Mood} = 0.6$, the Type I error for the Score test was only 0.021 as compared to 0.077 for the Wald test. As the correlation among mood items decreased, the Type I error for the Score test increased significantly. For example, when $\rho_{Mood} = 0.3$, the Type I error for the Score test was 0.101 as compared to 0.079 for the Wald test. The Type I error of the Wald test was always larger than the nominal level and increased as the correlation among the items decreased. However, Score test Type I error performance was influenced more by the correlation among the other items than was the Wald test.

Table 2.5 presents results from Simulation 2, where the “junk” item was corre-

lated with the mood items and associated with the “gold standard.” Power remained high under most conditions. For both $\rho_{Mood} = 0.6$ and $\rho_{Mood} = 0.3$, power did not drop substantially until ρ_{Junk} reached below 0.1. For $\rho_{Mood} = 0.6$ and $\rho_{Junk} = 0.05$, power for the Score test was 0.727 and power for the Wald test was 0.588. For $\rho_{Mood} = 0.3$ and $\rho_{Junk} = 0.0$, power for the Score test was 0.853 and power for the Wald test was 0.844, illustrating that the Score test was more powerful than the Wald test. The power of both tests remained high unless ρ_{Junk} was quite small. Also notable is that power for the Score test was consistently higher than power for the Wald test under the condition that ρ_{Junk} was small. Finally, it seemed that power was affected not only by small values of ρ_{Junk} , but also by the magnitude of the difference between ρ_{Mood} and ρ_{Junk} . This was evidenced in that for $\rho_{Mood} = 0.6$ and $\rho_{Junk} = 0.05$, power dropped below 1.0. However, for $\rho_{Mood} = 0.3$ and $\rho_{Junk} = 0.05$, power remained at 1.0. In sum, both the Score and the Wald test were robust to changes in correlations of the mood and junk items.

2.6 Discussion

The objective of this chapter was to evaluate items under consideration for inclusion in a scale while simultaneously comparing the scale under development to a “gold standard” measure. The results of our evaluation of the 17 DSR items showed that the MIMIC model can be useful to inform item reduction. Three of the DSR items in the full MIMIC model were identified as potential items for removal according

to the Score test. Removing these items could help to create a stronger core set of symptoms for PMS diagnosis.

It is not entirely surprising that headaches, lack of coordination, and cramps were nonsignificant in the full model. While premenstrual headaches exist, they are not limited to the premenstrual period, and this could be why the symptom is not able to discriminate between PMS and non PMS groups. Further, poor coordination is considered to an indicator of major depression as well as a potential symptom of PMS. Cramps are diagnostically an indication of dysmenorrhea rather than PMS. Thus, it is reasonable to believe that poor coordination and cramps are not the best indicators of PMS for these data but instead other related conditions.

Standard psychometric measures did not provide a clear picture of which items were most important. Item-total correlations displayed a range of values and allowed ranking of items in terms of importance to the scale, but none of the item-total correlations fell below the standard guideline of 0.2. Cronbach's alpha values were all above the suggested guideline of 0.9, which suggested general item redundancy but did not provide any indication of which particular items were redundant. The MIMIC model provided clearer results as well as an objective test of the significance of the items. Including the validation component into the model indicated that several items could be considered for removal.

In addition to using the MIMIC model generally, the use of the Score test derived in this paper for the MIMIC model is advocated for a number of reasons. Under the

constraint that factor loadings must be non-negative, the statistical test of a factor loading is a test of a variance component on the boundary of the parameter space. As a result, it is important to take into account that the Score test statistic no longer follows the traditional chi-square distribution and that the conventional Wald test is invalid. Instead, the test statistic follows a mixture of chi-square distributions, and the use of the modified Score test is more appropriate because it uses the correct chi-square distribution for this situation. Furthermore, simulations showed that when the null hypothesis was true the Score test performed better than the Wald test in the presence of a set of moderately correlated items and a true uncorrelated “junk” item. When the alternative hypothesis was true and the suspected junk item was actually a meaningful part of the model, although the Score and Wald tests generally yielded high power, the Score test performed better in proximity to the null hypothesis (i.e. when the correlation, ρ_{Junk} , was very small). The Score test presented here provides a preferable alternative to the Wald test for performing item reduction.

Item reduction and validation are common procedures for scale development. Rather than performing these processes separately, it has been demonstrated that it is advantageous to combine them into a single latent variable model. This model eliminates the potential bias induced by using separate item selection and validation procedures with better precision. While latent variable models such as the MIMIC model are more complex than simple factor analysis or psychometric measures such

as Cronbach's alpha, there is software available to estimate these types of latent variable models. An impartial method of analyzing items that is justifiable from a statistical perspective could add credibility to the usually subjective process of item selection, and adding a validation metric to the same model allows for the estimation of the relationship between the items and a "gold standard" measure.

Chapter 3

Application of Joint Latent

Variable Model for Item

Reduction and Validation for

Ordered Categorical Scale Items

3.1 Introduction

As shown in the previous chapter, joint latent variable methods such as the MIMIC model for continuous observed outcomes can be useful in combining item reduction and validation. Fortunately, extensions to classic latent variable models for continuous outcomes are available for other types of responses as well. Symptom

scales often involve Likert scale data that can be modeled as ordered categorical. In this chapter we focus on methods for categorical observed outcomes in which we assume that all observed responses originate from a continuous latent variable. Several extensions to classic latent variable models are highlighted below and an ordered categorical extension to the MIMIC model is presented with an illustration from the PAL clinical trial.

An extension of unidimensional factor analysis for dichotomous or ordered categorical items, item response theory (IRT) is widely used in educational testing and is becoming more prevalent in health measurement (Rasch, 1960). IRT models use a latent variable framework to explain the probability of “correctly” answering test items. Similar to factor loadings for continuous outcome factor analysis, discrimination parameters in IRT models are often used as metrics for item reduction.

Multi-trait multi-method (MTMM) models offer an alternative approach to evaluating measurement error (Campbell and Fiske, 1959). The most basic method, the MTMM matrix, consists of correlations of several concepts or traits measured by each of several methods. This correlation matrix provides estimates of reliability for each trait and method pair as well as estimates of validity for different measures of the same trait. While the MTMM matrix allows for useful estimates of reliability and validity, limitations of the MTMM matrix include the lack of rigorous statistical tests associated with correlation coefficients and the inability of the MTMM matrix to separate method variance from random error. ANOVA and

latent variable models like confirmatory factor analysis are proposed to account for these issues. ANOVA partitions variance into groups defined by person, method, and trait and provides a global estimate of each type of variance (Guilford, 1954). Repeated measures are needed to estimate trait-method interactions. The confirmatory factor analysis model provides estimates of correlation among observed traits and methods through factor loadings (Werts and Linn, 1970). The model contains several components: a trait component, a method component and a random error component. Overall, MTMM models provide an additional approach to assessing validity and can be useful in evaluating a “gold standard.”

As in the previous chapter, to explore the efficiency of performing item selection and validation in one model, we employed a MIMIC model (Joreskog and Goldberger, 1975). An extension to the MIMIC model for categorical items assumes that ordinal items originate from underlying unobserved continuous, normally distributed items. The model relates observed items with underlying unobserved items through a series of threshold relationships (Muthén, 1984a).

Data from the PAL clinical trial was considered as a motivating example (Schmitz et al., 2009, 2010). The primary aim was to identify important and clinically relevant symptoms from ordered categorical items in a scale and to demonstrate that this set of symptoms had a strong association with the “gold standard” arm volume difference. The remainder of the chapter is outlined as follows. Section 3.2 illustrates the MIMIC model and its formulation for Likert scale data, Section 3.3

describes the PAL sample, Section 3.4 presents results of the example, and in Section 3.5 I draw some conclusions.

3.2 Methods

Similar to Chapter 2, the MIMIC model was formulated such that the single latent variable is interpretable as a continuous measure of lymphedema severity. As a means of comparison, lymphedema symptoms were also analyzed using standard psychometric techniques such as item-total correlation and Cronbach’s alpha. The relationship between individual symptoms and the “gold standard” diagnostic measure of lymphedema was assessed with cumulative probit models.

3.2.1 MIMIC Model Specification

Consider a sample of n individuals from whom a set of m ordered categorical Likert scale outcome measurements (items), y_{i1}, \dots, y_{im} , and one continuous validation measurement (“gold standard”), z_i , are taken. The ordinal items are assumed to have originated from continuous, normally distributed items $y_{i1}^*, \dots, y_{im}^*$. For subject $i = 1, \dots, n$ and outcome measurements $j = 1, \dots, m$, the measurement model is specified as

$$\mathbf{y}_{i(m \times 1)}^* = \boldsymbol{\mu}_{(m \times 1)} + \boldsymbol{\lambda}_{(m \times 1)} b_{i(1 \times 1)} + \boldsymbol{\epsilon}_{i(m \times 1)} \quad (3.2.1)$$

where $\boldsymbol{\mu}$ is a vector of means, $\boldsymbol{\lambda}$ is a vector of factor loading parameters, b_i is the latent variable (lymphedema severity), and $\boldsymbol{\epsilon}_i$ is a vector of specific variances. In the classical MIMIC model \mathbf{y}_i^* is assumed to be directly observed, but here it is unobserved. Instead, the relationship between unobserved \mathbf{y}_i^* and observed \mathbf{y}_i is specified by

$$\begin{aligned}
\mathbf{y}_i &= C - 1 && \text{if } \tau_{C-1} < \mathbf{y}_i^* \\
&= C - 2 && \text{if } \tau_{C-2} < \mathbf{y}_i^* \leq \tau_{C-1} \\
&= \vdots \\
&= 1 && \text{if } \tau_1 < \mathbf{y}_i^* \leq \tau_2 \\
&= 0 && \text{if } \mathbf{y}_i^* \leq \tau_1
\end{aligned}$$

where the τ 's are threshold parameters defining category intervals on \mathbf{y}_i^* . The structural model with a single “gold standard” is specified as

$$b_{i(1 \times 1)} = \beta_{(1 \times 1)} z_{i(1 \times 1)} + \delta_{i(1 \times 1)} \tag{3.2.2}$$

where β is the regression coefficient measuring the association between the latent variable and the “gold standard.” The random variable δ_i is an error term for the latent variable, and we assumed $Var(\delta_i) = 1$, allowing the latent variable to be interpreted on a standardized scale. Estimation of the model was performed using the mean and variance adjusted weighted least squares estimator and Delta parameterization in Mplus Version 6.1. Example code for the ordinal version of the

MIMIC model is provided in Appendix B.

3.3 Application to Physical Activity and Lymphedema Clinical Trial

A prominent fear for many breast cancer survivors, lymphedema is a debilitating chronic disease that results from surgical excision of lymph nodes as part of breast cancer treatment. In addition to swelling, lymphedema can cause skin changes, reduction of limb function, and loss of sensation as well as depression, decreased quality of life, decreased physical self-esteem and other physical and psychological morbidities (Ahmed et al., 2008; Cormier et al., 2009; Shih et al., 2009).

Until recently, women at risk for or diagnosed with lymphedema have been encouraged to limit physical activity, even such mundane tasks as lifting grocery bags. This guideline has the effect of inhibiting everyday activities and may even slow physical recovery from cancer. However, results of the PAL trial contradicted these guidelines, showing that a progressive weight-training program was safe for breast cancer survivors. Among women diagnosed with lymphedema, the trial showed that a weight-training intervention was not only safe in that it did not significantly affect lymphedema severity, but it also was shown to effectively reduce the number and severity of arm and hand symptoms and the incidence of lymphedema exacerbations (Schmitz et al., 2009). A subsequent follow-up study indicated that in breast cancer

survivors who were at risk for lymphedema but who had not yet been diagnosed, a much larger group, the same weight-training intervention was not associated with an increased incidence of lymphedema (Schmitz et al., 2010). That weight-lifting was shown to be safe for both women at risk for lymphedema and women already diagnosed with lymphedema was revolutionary as great benefits such as increased muscle strength, decreased weight gain, and increased quality of life have been shown to result from increased physical activity (Schmitz et al., 2010).

Incidence of lymphedema in the literature varies dramatically with estimates between 6% to 70% (Schmitz et al., 2009). One factor attributed to the difference is the criteria used for diagnosis of lymphedema. Several diagnostic measures include: water displacement volumetry, extracellular water in the arm measured by multi-frequency bioelectrical impedance analysis, serial circumference measurements and truncated cone volumetry. Initial evidence indicates that self-reported lymphedema symptoms can be as useful as objective diagnostic measures in discriminating women with lymphedema from those without (Norman et al., 2001). Self-report of symptoms could be a useful measure in diagnosing lymphedema because patients are more aware of acute changes in swelling, skin tone, and function. Furthermore, it is argued that patient pain or distress should be incorporated in the diagnosis of lymphedema and that swelling alone is not sufficient for diagnosis. We explored the relationship among symptoms and a limb volume difference “gold standard” measure to determine the utility of this information in summarizing

lymphedema severity.

Outcome measures included self-reported lymphedema symptoms and an objective “gold standard” measure of lymphedema severity. The self-reported symptoms were measured using the validated Norman Lymphedema Survey (Norman et al., 2001). The severity of 13 symptoms was assessed: rings too tight, watch too tight, bracelets too tight, clothing too tight, puffiness, knuckles not visible, veins not visible, skin feels leathery, arm feels tired, pain, pitting, swelling after exercise, and difficulty writing. Symptoms were measured on a 5-point Likert scale with responses ranging from 0 (no symptom) to 4 (very severe). Water displacement volumetry was chosen as the “gold standard” and was measured as the percent difference in volume between the lymphedema-affected and unaffected arms.

EFA models for ordered categorical outcomes revealed three distinct factors from the Norman Lymphedema Survey items. One symptom, swelling after exercise, did not load strongly on any factor. This symptom was chosen as a “junk” item, potentially eligible for removal from the scale. The factor representing tissue, swelling and function was chosen for further investigation. This factor included the following symptoms: clothing too tight, puffiness, skin feels leathery, arm feels tired, pain, and difficulty writing. In the MIMIC model, Norman lymphedema symptoms served as observed indicators and volume difference served as the observed cause.

3.4 Results

Data for the example came from a subset of the PAL trial. The sample comprised $N = 141$ women diagnosed with lymphedema at baseline. The average volume difference, defined as the percent difference in arm volume between affected and unaffected arms, for the sample was 16.11 percent [95% C.I. (13.59, 18.65)]. Figure 3.1 illustrates the association between individual symptom response categories and average percent volume difference. For clothing too tight, puffiness, skin feels leathery, and arm feels tired, there was a general increase in mean volume difference with increasing levels of symptom severity. Because of the small sample sizes in response category of very severe for pain and difficulty writing ($n = 2$ and $n = 3$, respectively), responses severe and very severe were combined. There did not appear to be a trend in the association between mean volume difference and symptom severity for pain, difficulty writing, or swelling after exercise.

Polychoric correlations among the symptoms are presented in Table 3.1. Correlations among the items in the swelling/function factor were generally moderate to strong, ranging from $\rho_{\text{Pain, Clothing}} = 0.222$ to $\rho_{\text{Puffiness, Clothing}} = 0.742$. Correlations among the junk item and other items were much smaller, ranging from $\rho_{\text{Puffiness, Swelling}} = -0.002$ to $\rho_{\text{Writing, Exercise}} = 0.220$. These correlations provided initial evidence that swelling after exercise could be a candidate for removal because it did not correlate highly with the other items.

Standard psychometric techniques also identified swelling after exercise as po-

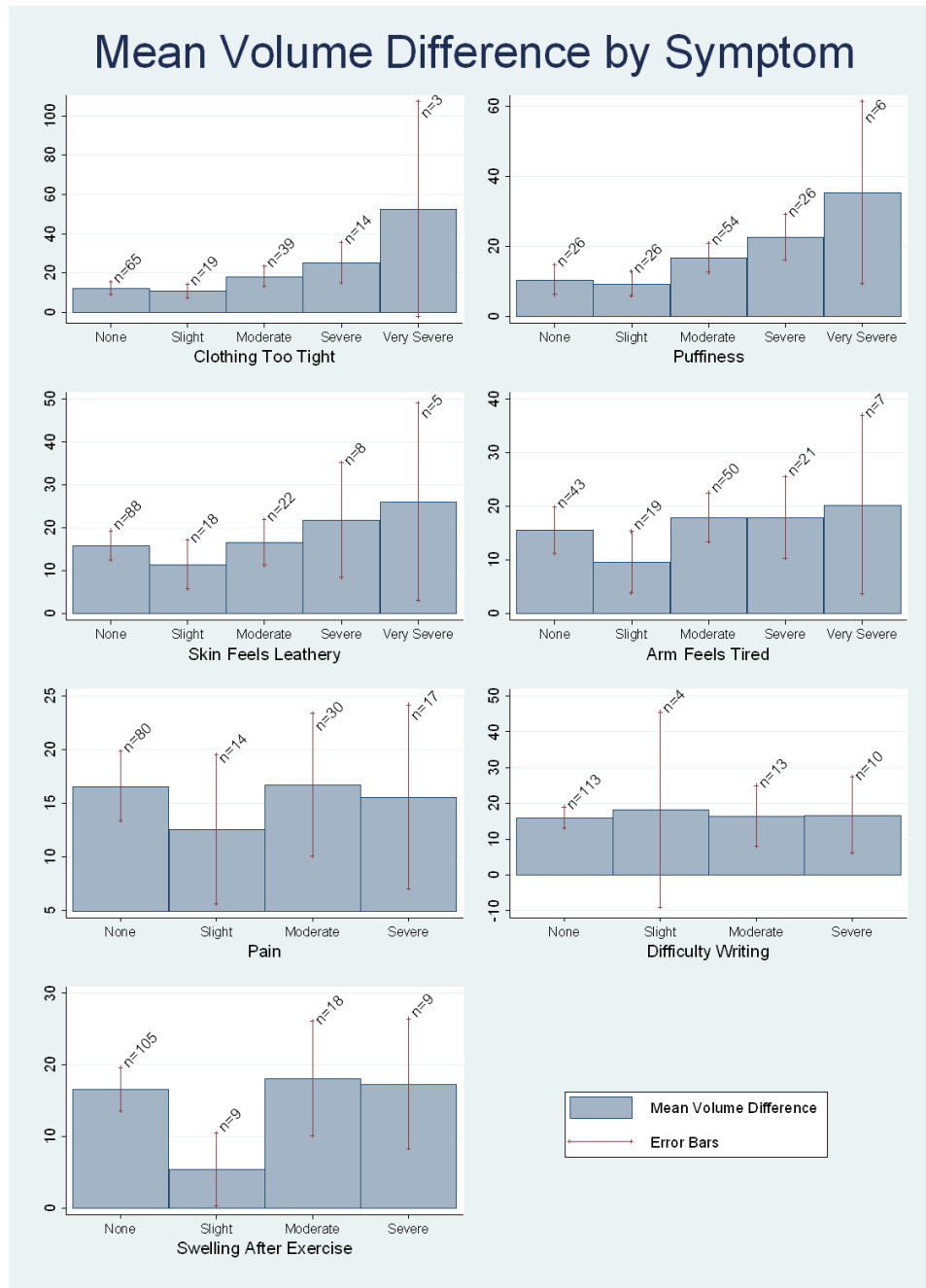


Figure 3.1: Mean volume difference plots with error bars for Norman Lymphedema Survey items.

tential “junk” (see Table 3.2). Item-total correlations when all items were included in the scale ranged from $\rho = 0.111$ for swelling after exercise to $\rho = 0.666$ for arm

Table 3.1: Polychoric correlations of Norman Lymphedema Survey items.

	Clothing too tight	Puffiness	Skin Feels Leathery	Arm Feels Tired	Pain	Difficulty Writing	Swelling
Clothing Too Tight	1.0000						
Puffiness	0.7422	1.0000					
Skin Feels Leathery	0.5219	0.5814	1.0000				
Arm Feels Tired	0.4514	0.6718	0.5426	1.0000			
Pain	0.2217	0.4231	0.3323	0.5455	1.0000		
Difficulty Writing	0.2408	0.4141	0.2257	0.5687	0.3858	1.0000	
Swelling	0.0322	-0.0017	0.1966	0.2169	-0.0112	0.2197	1.0000

Table 3.2: Standard psychometric measures of Norman Lymphedema Survey items.

Symptom	Item-total Correlation	Alpha
Clothing Too Tight	0.474	0.687
Puffiness	0.636	0.647
Skin Feels Leathery	0.495	0.682
Arm Feels Tired	0.666	0.639
Pain	0.382	0.709
Difficulty Writing	0.358	0.715
Swelling After Exercise	0.111	0.769

feels tired. According to the $\rho = 0.2$ guideline (Streiner and Norman, 1994), swelling after exercise could potentially be considered for removal. The overall Cronbach's alpha assuming all seven items in the scale was $\alpha = 0.727$, indicating that the removal of swelling after exercise is unwarranted.

Results from univariate cumulative probit models are featured in Table 3.3. According to these models, there was a statistically significant relationship between volume difference and increasing item severity for clothing too tight and puffiness ($p < 0.001$ for both). In both of these models, the cumulative probability starting at the severe end of the scale increased with higher levels of volume difference ($\beta_{\text{Clothing too tight}} = 0.030, S.E. = 0.006, \beta_{\text{Puffiness}} = 0.029, S.E. = 0.007$). In other words, symptom severity for clothing too tight and puffiness tended to be more intense as volume difference increased. All other items identified in the exploratory

model as well as the potential “junk” item were not significantly associated with volume difference when considered individually.

When symptoms were considered jointly, factor loadings from the ordinal MIMIC model indicated a strong relationship among the candidate items and the latent measure of lymphedema severity (see Table 3.4). For these items identified by EFA, factor loadings exceeded 0.5 and ranged from $\lambda_{\text{Pain}} = 0.512$ to $\lambda_{\text{Puffiness}} = 0.906$. All factor loadings for candidate items were statistically significant ($p < 0.001$). The factor loading for swelling after exercise was small ($\lambda_{\text{Swelling}} = 0.145$) and the test for the factor loading was not statistically significant ($p = 0.132$), indicating that this item did not contribute to the underlying measure of lymphedema severity. The coefficient for the regression of the latent lymphedema severity on volume difference was $\beta = 0.020$ and was statistically significant ($p = 0.001$). This represents a global test of the significance of the relationship between volume difference and all items in the MIMIC model. The regression coefficient can be interpreted as follows: every 1% change in volume difference corresponded to a $\beta = 0.020$ change in latent lymphedema severity on the standard normal scale. In other words, a clinically meaningful change in volume difference of 5% would correspond to a change of $\beta = 0.10$, an effect size of 0.1 standard deviation units in the latent measure of lymphedema severity.

Table 3.3: Univariate cumulative probit models.

Symptom	Estimate	SE	P
Clothing too tight			
α_1	0.356	0.150	0.017
α_2	0.723	0.153	<0.001
α_3	1.746	0.199	<0.001
α_4	2.797	0.396	<0.001
β	0.030	0.006	<0.001
Puffiness			
α_1	-0.503	0.156	0.001
α_2	0.106	0.146	0.467
α_3	1.260	0.176	<0.001
α_4	2.363	0.259	<0.001
β	0.029	0.007	<0.001
Skin feels leathery			
α_1	0.431	0.157	0.006
α_2	0.796	0.169	<0.001
α_3	1.452	0.206	<0.001
α_4	1.940	0.255	<0.001
β	0.007	0.006	0.260
Arm feels tired			
α_1	-0.394	0.151	0.009
α_2	-0.032	0.146	0.829
α_3	0.963	0.159	<0.001
α_4	1.771	0.211	<0.001
β	0.007	0.006	0.243
Pain			
α_1	0.138	0.148	0.352
α_2	0.399	0.151	0.008
α_3	1.140	0.170	<0.001
α_4	2.161	0.286	<0.001
β	-0.002	0.006	0.751
Difficulty writing			
α_1	0.905	0.178	<0.001
α_2	1.014	0.180	<0.001
α_3	1.503	0.206	<0.001
α_4	2.227	0.320	<0.001
β	0.002	0.008	0.771
Swelling after exercise			
α_1	0.632	0.175	<0.001
α_2	0.847	0.192	<0.001
α_3	1.497	0.219	<0.001
α_4	NA	NA	NA
β	-0.002	0.008	0.842

Table 3.4: Ordered categorical MIMIC model applied to PAL data.

Symptom	Estimate	SE	P
Factor Loadings			
Clothing too tight	0.709	0.051	<0.001
Puffiness	0.906	0.038	<0.001
Skin feels leathery	0.631	0.057	<0.001
Arm feels tired	0.779	0.038	<0.001
Pain	0.512	0.069	<0.001
Difficulty writing	0.526	0.075	<0.001
Swelling after exercise	0.145	0.096	0.132
Regression Coefficient			
β	0.020	0.006	0.001
Thresholds			
Clothing too tight			
α_1	0.356	0.150	0.017
α_2	0.723	0.153	<0.001
α_3	1.746	0.199	<0.001
α_4	2.798	0.396	<0.001
Puffiness			
α_1	-0.503	0.156	0.001
α_2	0.106	0.146	0.467
α_3	1.260	0.176	<0.001
α_4	2.363	0.259	<0.001
Skin feels leathery			
α_1	0.431	0.157	0.006
α_2	0.796	0.169	<0.001
α_3	1.452	0.206	<0.001
α_4	1.940	0.255	<0.001
Arm feels tired			
α_1	-0.394	0.151	0.009
α_2	-0.032	0.146	0.829
α_3	0.963	0.159	<0.001
α_4	1.771	0.211	<0.001
Pain			
α_1	0.138	0.148	0.352
α_2	0.399	0.151	0.008
α_3	1.140	0.170	<0.001
α_4	2.161	0.286	<0.001
Difficulty writing			
α_1	0.905	0.178	<0.001
α_2	1.014	0.180	<0.001
α_3	1.503	0.206	<0.001
α_4	2.228	0.320	<0.001
Swelling after exercise			
α_1	0.632	0.175	<0.001
α_2	0.847	0.192	<0.001
α_3	1.497	0.219	<0.001
α_4	NA	NA	NA

3.5 Discussion

In terms of the motivating example, standard psychometric techniques, univariate models, and the MIMIC model provided somewhat conflicting results. Item-total correlations identified swelling after exercise as a potential junk item. While psychometric techniques took into account the relationship among symptoms, they did not take into account the relationship between items and “gold standard” volume difference. Cumulative probit models revealed a significant relationship between volume difference and only two of the items: clothing too tight and puffiness. These models were useful for describing the relationship between volume difference and item severity scores in the univariate setting but did not consider the correlation among the items and how all the items together are associated with volume difference. The MIMIC model showed that six of the seven items in the model were significant components of latent lymphedema severity. The MIMIC model is advocated because it takes into account the correlation among the items as well as the relationship between the latent measure of severity and the “gold standard.” Unlike the standard psychometric techniques, which assume an underlying normal distribution of the items, the categorical formulation of the MIMIC model is an appropriate choice for ordinal data. The results of our evaluation of the Norman symptoms showed that the MIMIC model can be useful to inform item reduction. The MIMIC model suggested that the potential junk item was not statistically significant and could be considered for removal. There are several directions for future work based on the

methods investigated in this chapter. Of particular interest is the influence of scale item distributions on MIMIC model estimates. I noticed that there was a significant proportion of non-response for many of the Norman Lymphedema Survey items. In Chapter 4 I investigate the effects of skewed item distributions and propose a new model to account for significant symptom non-response.

Chapter 4

A Joint Latent Variable Model for Item Reduction and Validation in the Presence of a Preponderance of Zeros

4.1 Introduction

In this chapter I propose an extension to the ordered categorical MIMIC model presented in Chapter 3. In the case of highly skewed item distributions it is possible that estimates from a standard MIMIC model are not entirely accurate, and in the motivating example the Norman Lymphedema Survey scale items exhibited a

preponderance of zeros. The methodological objectives in this chapter are twofold: First, simultaneously and efficiently perform item reduction and validation using a joint latent variable model, and second, determine a method to account for the excess zeros in item distributions. In this chapter I propose a multivariate zero-inflated proportional odds (MZIPO) model, a joint latent variable model that accomplishes both tasks simultaneously. The MZIPO model yields more accurate estimates both of the relationship between individual symptoms and the latent variable and the relationship between the latent variable and the gold standard diagnostic measure than if a standard MIMIC model were used.

The proposed MZIPO model can be thought of as an extension of several statistical models. This section highlights several models that motivate the MZIPO model. Several previous latent variable models formulated for ordered categorical data establish a foundation for the MZIPO model. As discussed in Jöreskog and Moustaki (2000), latent variable models for ordered categorical items are generally framed either from an underlying response function approach where it is assumed that an observed ordered categorical item is a manifestation of an underlying continuous latent variable (Muthén, 1984b; Jöreskog, 1994), or a response function approach, which was originally developed in item response theory (IRT) as described by the graded response model (Samejima, 1970). Extensions to the graded response model accommodate multiple latent variables and a variety of link functions. Moustaki (2000) presents a general class of latent variable models for ordinal

items in a regression framework similar to McCullagh (1980).

Other extensions allow for mixed observed outcome types. Sammel et al. (1997) propose a latent variable model for mixed discrete and continuous outcome types that also allows for the inclusion of covariates. This model accommodates any type of observed outcome that can be framed as a generalized linear model. The estimation method presented in Sammel et al. motivated the EM estimation procedure used in the MZIPO model, which is described in Section 4.3.

In addition to models for mixed observed outcome types, a general class of latent variable models allows for mixed latent variable types. Proposed by Muthén (2008), latent variable hybrid models contain discrete and continuous latent variables and accommodate cross-sectional or longitudinal data. The cross-sectional formulation of these models, also known as factor mixture models (FMMs), combines the classic factor analysis model and the classic latent class model to cluster items into a smaller set of dimensions and subjects into unobserved subpopulations. The main objective of FMMs is to determine structural relations between latent and observed variables in the presence of unobserved population heterogeneity. If subjects could be classified into any number of observed groups then multiple group latent variable methods could be used. However, when population heterogeneity is unobserved, latent classes can be used to infer subpopulations of interest. Additionally, FMMs allow for the inclusion of covariates, which can help to explain the unobserved heterogeneity present in the data. The MZIPO model fits in the framework of

FMMS, but the model is distinct because the latent classes have a slightly different form than FMMS. In the MZIPO approach it is assumed that one of the classes has a degenerate distribution. In the motivating example it was of interest to determine, among subjects who did not experience symptoms at baseline, who was truly susceptible to experiencing lymphedema incidence or exacerbations.

Statistical models that accommodate both continuous and discrete latent variables have been described elsewhere (Arminger et al., 1999; Dolan and van der Maas, 1998; Yung, 1997) but until recently have lacked popularity in psychometrics and biomedical research. These models are extremely useful for our application because they will allow us to distinguish between groups of patients who are truly at a greater risk for lymphedema from those whose symptoms do not indicate a substantial risk. Factor mixture models to accommodate sample heterogeneity for quantitatively measured outcomes are described generally in Muthén and Shedden (1999) and Lubke and Muthén (2005). Several other articles have assessed model performance of the factor mixture model under various conditions (Lubke and Neale (2006), Lubke and Muthén (2007)). Lubke and Neale (2008) builds on Lubke and Neale (2006) by extending the model to deal with binary and ordinal outcomes. In general, FMMS will be useful in comparing the symptom profiles of those at risk to an objective “gold standard” measure of lymphedema severity.

In addition to FMMS, one particular statistical model served as motivation for the MZIPO model. In the motivating example, I utilize the latent class to account

for zero inflation in the symptom battery, thus extending the work of Kelley and Anderson (2008) who addressed this issue in a univariate setting. Their model, a univariate zero-inflated proportional odds (ZIPO) model, is a mixture model that incorporates the probability of non-response into an ordinal regression model. Similar to the zero-inflated Poisson model, the ZIPO model accounts for non-responders using a mixture model approach but is intended for ordered categorical observed outcomes. The ZIPO model is particularly useful because it allows for the simultaneous modeling of symptom frequency and severity in the presence of a preponderance of zeros.

The rationale of the ZIPO model is that the mixture distributions come from a population that can be divided into two unobserved states. The unsusceptible state includes those subjects who experience “true” symptom non-response. On the other hand, the susceptible state includes subjects for whom a particular symptom may not be present at the time of data collection but who is still susceptible to experiencing the symptom at a different time or under different conditions. As a result, the conditional distribution for the ordinal regression may still contain some observed zeros.

Building on the papers cited above, the MZIPO model fits into the general class of factor mixture models as it combines a classic factor analysis model and latent class model components. It is unique from the traditional factor mixture model, however, in that the latent class component is used to account for zero-inflation

in the data, so that one of the mixture distributions is assumed to be degenerate. The MZIPO model can be thought of as an extension to the ZIPO model in several ways. First, we are extending the ZIPO model to accommodate multiple observed ordinal outcomes. This extension allows for the evaluation of an entire battery of symptom items simultaneously rather than each item individually. Next, in order to model the multivariate ordinal outcomes the model also includes a continuous latent variable. Including the continuous latent variable provides a single weighted summary measure of the multivariate outcomes, or latent factor score, for each individual. The MZIPO model incorporates useful features of previously proposed latent variable models to accomplish the objectives set for the motivating example.

The rest of the chapter is outlined as follows: Section 4.2 specifies the MZIPO model, Section 4.3 outlines estimation via the EM algorithm, Section 4.4 describes the motivating example, Section 4.5 presents results of the application of the MZIPO model, and finally conclusions are given in Section 4.6.

4.2 Model Specification

Let y_{ij} be an ordered categorical measure for subject $i = 1, \dots, n$ on item $j = 1, \dots, m$ where each item contains $k = 1, \dots, K_j$ response categories.

4.2.1 Measurement Model

The measurement model is an ordered categorical factor analysis model assuming a proportional odds structure for observed ordinal items as described by Jöreskog and Moustaki (2000). The model measures the relationship between individual scale items and the continuous latent disease severity measure. This model can be thought of as a latent variable formulation of a multivariate generalized linear model. Similar models with a variety of link functions can be found in Moustaki (2000). For a model with a single continuous latent variable and a logit link, the measurement model is specified as

$$\ln \left[\frac{\gamma_k^{(j)}(b)}{1 - \gamma_k^{(j)}(b)} \right] = \alpha_k^{(j)} - \beta_j b_i$$

where $\gamma_k^{(j)}(b) = \Pr(y_{ij} \leq k) = \pi_1^{(j)}(b) + \dots + \pi_k^{(j)}(b)$ is the probability of a response in category k or lower on variable j . It follows that

$$\gamma_k^{(j)}(b) = \frac{\exp \left[\alpha_k^{(j)} - \beta_j b_i \right]}{1 + \exp \left[\alpha_k^{(j)} - \beta_j b_i \right]} = \Psi \left[\alpha_k^{(j)} - \beta_j b_i \right]$$

where $\Psi(\cdot)$ is the distribution function of the logistic distribution. Therefore, the conditional distribution of $(y_{ij} | b_i)$ is given by

$$f(y_{ij} | b_i) = \prod_{k=0}^{K_j} \pi_k^{(j)}(b_i)^{I[y_{ij}=k]} = \prod_{k=0}^{K_j} \left(\Psi \left[\alpha_k^{(j)} - \beta_j b_i \right] - \Psi \left[\alpha_{k-1}^{(j)} - \beta_j b_i \right] \right)^{I[y_{ij}=k]}.$$

$I[y_{ij} = k]$ is an indicator function that equals 1 if $y_{ij} = k$ and 0 otherwise.

4.2.2 Structural Model

The structural model measures the relationship between the continuous measure of latent disease severity and the “gold standard” diagnostic measure. The model is a simple linear regression model

$$b_i = \tau T_i + \delta_i$$

where τ is the coefficient for the regression of the latent variable on the gold standard, T_i is the gold standard covariate measure, and δ_i is the error term for the latent variable. The test of the regression coefficient in the model serves as an indicator of criterion validity in that it measures the strength of the relationship between the “gold standard” diagnostic measure and disease severity as measured by the observed scale items. For purposes of identifiability it is assumed in this case that $\text{Var}(\delta_i) = 1$, although other constraints are possible (Muthén and Shedden, 1999).

4.2.3 Latent Class Model

Define latent class c_i as a categorical latent variable with two classes to represent the susceptible and unsusceptible states such that

$$c_i = \begin{cases} 1, & \text{if } \mathbf{y}_i \text{ is in the unsusceptible state} \\ 0, & \text{if } \mathbf{y}_i \text{ is in the susceptible state.} \end{cases}$$

Note that latent class membership is based on the entire vector of observed item responses.

Observed Response and Latent Class Assignment

As described in the introduction to this chapter, it is assumed that the unobserved unsusceptible state contains only “true” symptom non-response. Define observed responders as those subjects who have a non-zero response to at least one of the scale items. An observed non-responder is then defined as a subject who exhibits a zero response to all scale items. An important element in the MZIPO model is that the classification in the observed responder group means that a subject will automatically be assigned to the susceptible state. The MZIPO model then classifies all observed non-responders into either the unsusceptible or susceptible state. A comparison of observed response and latent class assignment is presented later.

4.2.4 Complete Data Likelihood

For the purpose of model building the latent class assignment is treated as missing data. Suppose it was possible to know which subjects came from the unsusceptible state and which came from the multinomial distribution (susceptible state). Latent class assignment could be defined as $z_i = 1$ when the observed responses for subject i (\mathbf{y}_i) came from the unsusceptible state and $z_i = 0$ when the responses came from the multinomial distribution. For parameter vector $\Theta = [\boldsymbol{\alpha}_k^{(j)}, \boldsymbol{\beta}, \tau, p]$ a complete-data likelihood could be constructed as

$$L = \prod_{i=1}^n f(\mathbf{y}_i, z_i, b_i | \Theta) = \prod_{i=1}^n \Pr(\mathbf{y}_i = \mathbf{0}, z_i = 1)^{z_i} \Pr(\mathbf{y}_i = \mathbf{k}, z_i = 0, b_i)^{(1-z_i)}.$$

Because the unsusceptible class includes only zero responses on all observed outcomes,

$$\begin{aligned} \Pr(\mathbf{y}_i = \mathbf{0}, z_i = 1) &= \Pr(\mathbf{y}_i = \mathbf{0} | z_i = 1)Pr(z_i = 1) \\ &= p. \end{aligned}$$

Note that $\Pr(z_i = 0) = 1 - p, \Pr(z_i = 1) = p$ because the prevalence of the unsusceptible and susceptible states is of interest. In other formulations of factor mixture models and the ZIPO model, it is possible to allow for covariates to explain the probability of latent class membership.

An important assumption in the MZIPO model is that the latent class and the continuous latent variable are independent of one another, i.e. $b_i \perp\!\!\!\perp z_i$. This assumption allows for the separation of components in the model likelihood and facilitates estimation of the model.

For subjects in the susceptible state,

$$\begin{aligned}
\Pr(\mathbf{y}_i = \mathbf{k}, z_i = 0, b_i) &= \Pr(\mathbf{y}_i = \mathbf{k} \mid z_i = 0, b_i) \Pr(z_i = 0 \mid b_i) f(b_i) \\
&= \Pr(\mathbf{y}_i = \mathbf{k} \mid z_i = 0, b_i) \Pr(z_i = 0) f(b_i) \\
&= \Pr(\mathbf{y}_i = \mathbf{k} \mid z_i = 0, b_i) (1 - p) f(b_i),
\end{aligned}$$

where

$$\begin{aligned}
&\Pr(\mathbf{y}_i = \mathbf{k} \mid z_i = 0, b_i) \\
&= \prod_{j=1}^M \prod_{k=0}^{K_j} \Pr(y_{ij} = k \mid z_i = 0, b_i)^{I[y_{ij}=k]} \\
&= \prod_{j=1}^M \prod_{k=0}^{K_j} [\gamma_k^{(j)}(\mathbf{b}) - \gamma_{k-1}^{(j)}(\mathbf{b})]^{I[y_{ij}=k]} \\
&= \prod_{j=1}^M \prod_{k=0}^{K_j-1} \left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^{I[y_{ij}=k]} \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]^{I[y_{ij}=K]}.
\end{aligned}$$

The final element of the complete data likelihood arises from the structural model in which we assume the continuous latent variable to be normally distributed

$$f(b_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(b_i - \tau T_i)^2}{2}}$$

Given the components defined above, the log-likelihood is given by

$$\begin{aligned}
\ell = \ln L = & \sum_{i=1}^n z_i \ln p + (1 - z_i) \ln(1 - p) \\
& + \sum_{i=1}^n (1 - z_i) \sum_{j=1}^M \left[\sum_{k=0}^{K-1} I[y_{ij} = k] \ln \left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right] \right. \\
& \left. + I[y_{ij} = K] \ln \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right] \right] \\
& + \sum_{i=1}^n (1 - z_i) \left[\ln \frac{1}{\sqrt{2\pi}} - \frac{1}{2} (b_i - \tau T_i)^2 \right]
\end{aligned}$$

which, for the purpose of estimation, can be separated into three distinct portions with parameters for the latent class, the measurement model, and the structural model.

4.3 Model Estimation

The EM algorithm is a natural choice for estimation of the MZIPO model because both types of latent variables can be considered missing data. Little and Rubin (1987) demonstrated for data $Y = (Y_{obs}, Y_{mis})$, the EM algorithm proceeds by first factoring the complete data log-likelihood into one piece for the marginal log-likelihood of the observed data and one piece for the log-likelihood of the missing data conditional on the observed data

$$\ell(\theta | Y) = \ell(\theta | Y_{obs}, Y_{mis}) = \ell(\theta | Y_{obs}) + \ln f(Y_{mis} | Y_{obs}, \theta).$$

Solving for $\ell(\theta | Y_{obs})$ yields

$$\ell(\theta | Y_{obs}) = \ell(\theta | Y) - \ln f(Y_{mis} | Y_{obs}, \theta). \quad (4.3.1)$$

As in Little and Rubin, the expectation of both sides of Equation 4.3.1 with respect to the distribution of missing data given observed data is given by

$$E\ell(\theta | Y_{obs}) = Q(\theta | \theta^{(t)}) - H(\theta | \theta^{(t)})$$

for

$$Q(\theta | \theta^{(t)}) = \int_{-\infty}^{\infty} [\ell(\theta | Y_{obs}, Y_{mis})] f(Y_{mis} | Y_{obs}) dY_{mis}.$$

In the MZIPO model, define $\mathbf{Y}_{obs} = (\mathbf{y}_i)$ and $\mathbf{Y}_{mis} = (z_i, b_i)$ for $\Theta = [\boldsymbol{\alpha}_k^{(j)}, \boldsymbol{\beta}, \tau, p]$

so that

$$Q(\Theta | \Theta^{(t)}) = \int_{-\infty}^{\infty} [\ell(\Theta | \mathbf{y}_i, T_i, z_i, b_i)] f(z_i, b_i | \mathbf{y}_i, T_i, \Theta^{(t)}) d(z_i, b_i)$$

where

$$\ell(\Theta | \mathbf{y}_i, T_i, z_i, b_i) = \sum_{i=1}^n \left[\ln f(\mathbf{y}_i | z_i, b_i, \Theta) + \ln f(z_i | \Theta) + \ln f(b_i | \Theta) \right].$$

In the M-step of the EM algorithm $Q(\Theta | \Theta^{(t)})$ is maximized such that

$$\frac{\partial}{\partial \Theta} Q(\Theta | \Theta^{(t)}) = 0.$$

Under certain regularity conditions (Little and Rubin, 1987, p. 136), met in our example because the distribution of \mathbf{y}_i is multinomial, the order of integration and differentiation can be reversed and the following equation can be solved

$$\frac{\partial}{\partial \Theta} Q(\Theta | \Theta^{(t)}) = \int_{-\infty}^{\infty} \frac{\partial}{\partial \Theta} [\ell(\Theta | \mathbf{y}_i, T_i, z_i, b_i)] f(z_i, b_i | \mathbf{y}_i, T_i, \Theta^{(t)}) d(z_i, b_i) = 0.$$

Define $h(b_i, z_i | \mathbf{y}_i) = f(z_i, b_i | \mathbf{y}_i, \Theta^{(t)})$, the posterior distribution of the missing data. Solving $Q(\Theta | \Theta^{(t)})$ is equivalent to solving for the expected score function $S(\theta_j)$ with respect to the posterior distribution of missing data given observed data

$$E_{b,z} S_i(\Theta_j) = \int_{-\infty}^{\infty} \sum_z S_i(\Theta_j) h(b, z | \mathbf{y}_i) db_i. \quad (4.3.2)$$

Because closed form solutions for $E_{b,z} S(\Theta_j) = 0$ are not available, a Fisher scoring algorithm is used to iteratively update estimates of Θ_j . Sammel et al. (1997) and others showed that assuming conditional independence of the observed items given the continuous latent variable, the parameter space can be separated into components corresponding to each outcome. The expectation step, approximating the integral $E_{b,z} S(\Theta_j)$, will also yield estimates of b_i and z_i by setting $E_{b,z} S(\Theta_j) = 0$,

and solving for each parameter in the maximization step will give iterative estimates of each parameter.

4.3.1 Expectation Step

Expectation of z_i

A natural first step of the EM algorithm is the estimation of latent class membership.

An estimate of the expected value of z_i given the posterior distribution is computed as

$$\begin{aligned} E_{h(b,z)}(z_i) &= \int_{-\infty}^{\infty} \sum_{z_i=0}^1 z_i h(b_i, z_i \mid \mathbf{y}_i, \boldsymbol{\theta}) db_i, \\ &= \frac{\int_{-\infty}^{\infty} \sum_{z_i=0}^1 z_i f(\mathbf{y}_i \mid b_i, z_i) f(b_i \mid z_i) f(z_i) db_i}{\int_{-\infty}^{\infty} \sum_{z_i=0}^1 f(\mathbf{y}_i \mid b_i, z_i) f(b_i \mid z_i) f(z_i) db_i}. \end{aligned} \quad (4.3.3)$$

Considering the components separately, the numerator can be re-expressed as

$$\begin{aligned} &\int_{-\infty}^{\infty} \sum_{z_i=0}^1 z_i f(\mathbf{y}_i \mid b_i, z_i) f(b_i \mid z_i) f(z_i) db_i \\ &= \int_{-\infty}^{\infty} (0) f(\mathbf{y}_i \mid b_i, z_i = 0) f(b_i \mid z_i = 0) f(z_i = 0) db_i \\ &+ \int_{-\infty}^{\infty} (1) f(\mathbf{y}_i \mid b_i, z_i = 1) f(b_i \mid z_i = 1) f(z_i = 1) db_i \quad (4.3.4) \\ &= p * \Pr(\mathbf{y}_i = 0 \mid z_i = 1) \int_b f(b_i \mid z_i = 1) db_i \\ &= p, \end{aligned}$$

and the denominator as

$$\begin{aligned}
& \int_{-\infty}^{\infty} \sum_{z_i=0}^1 f(\mathbf{y}_i | b_i, z_i) f(b_i | z_i) f(z_i) db_i \\
&= \int_{-\infty}^{\infty} f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) f(z_i = 0) db_i \\
&+ \int_{-\infty}^{\infty} f(\mathbf{y}_i | b_i, z_i = 1) f(b_i | z_i = 1) f(z_i = 1) db_i \\
&= \int_{-\infty}^{\infty} (1 - p) f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) db_i + p.
\end{aligned} \tag{4.3.5}$$

Putting Equation 4.3.4 and 4.3.5 together, z_i can be estimated by

$$E_{h(b,z)}(z_i) = \frac{p}{p + \sum_{t=1}^T (1 - p) w_t \exp(b_t^2) f(\mathbf{y}_i | b_t, z_i = 0) f(b_t | z_i = 0)}. \tag{4.3.6}$$

Note that $E_{h(b,z)}(z_i) = 0$ for subjects with non-zero observed \mathbf{y}_i values because by definition they cannot be part of the unsusceptible class. Furthermore, a dichotomous estimate of $E_{h(b,z)}(z_i)$ for each subject is created at each iteration such that $(\hat{z}_i)^{(s)} = 1$ if $E_{h(b,z)}(\hat{z}_i) > 0.5$ at iteration (s) and 0 otherwise.

Approximating Expectations using Gauss-Hermite Quadrature

Expectations of expected score functions and sufficient statistics necessary for the estimation of the MZIPO model parameters cannot be directly evaluated because the solutions are not available in closed form, so the Gauss-Hermite quadrature method (Abramowitz and Stegun, 1964) was used. This method is often used to solve complicated integrals involving the normal distribution. Though the continu-

ous latent variable in the MZIPO model is normally distributed, the Gauss-Hermite quadrature method will require a change of variables because the latent variable is not distributed as standard normal. Many of the expectations required for estimation contain the integral

$$\int_{-\infty}^{\infty} f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) db_i \quad (4.3.7)$$

where $f(b_i)$ is given in Equation 4.2.1.

Using a change of variables such that

$$x_i = \frac{b_i - \tau T_i}{\sqrt{2}}$$

which implies

$$b_i = \tau T_i + \sqrt{2}x_i,$$

Equation 4.3.7 can be written

$$\begin{aligned} & \int_{-\infty}^{\infty} f(\mathbf{y}_i | b_i = \tau T_i + \sqrt{2}x_i, z_i = 0) \frac{1}{\sqrt{2\pi}} \exp(-x_i^2) \frac{db_i}{dx_i} dx_i \\ &= \int_{-\infty}^{\infty} f(\mathbf{y}_i | b_i = \tau T_i + \sqrt{2}x_i, z_i = 0) \frac{1}{\sqrt{\pi}} \exp(-x_i^2) dx_i \end{aligned}$$

given $\frac{db_i}{dx_i} = \sqrt{2}$. This equation is in the form of integrals given in Liu and Pierce

(1994) and can be approximated by

$$\sum_{t=1}^T \frac{w_t}{\sqrt{\pi}} f(\mathbf{y}_i | \hat{\tau}T_i + \sqrt{2}x_t, z_i = 0)$$

where x_t is the t^{th} node of the Hermite polynomial and w_t is the corresponding weight. Gauss-Hermite approximation for Equation 4.3.6, the expected value of z_i , is given by

$$E_{h(b,z)} z_i \approx \frac{p}{p + (1-p) \sum_{t=1}^T \frac{w_t}{\sqrt{\pi}} f(\mathbf{y}_i | \hat{\tau}T_i + \sqrt{2}x_t, z_i = 0)}$$

and expected score functions (see Equation 4.3.2)

$$\begin{aligned} E_{h(b,z)} S_i(\Theta) &= \int_{-\infty}^{\infty} \sum_{z_i=0}^1 S_i(\Theta) h(b_i | \mathbf{y}_i) db_i \\ &= \frac{\int_{-\infty}^{\infty} S_i(\Theta) f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) f(z_i = 0) db_i}{\int_{-\infty}^{\infty} f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) f(z_i = 0) db_i} \\ &= \frac{\int_{-\infty}^{\infty} S_i(\Theta) f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) (1-p) db_i}{\int_{-\infty}^{\infty} f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) (1-p) db_i} \end{aligned}$$

can be approximated by

$$E_{h(b,z)} S_i(\Theta) \approx \frac{\sum_{t=1}^T \frac{w_t}{\sqrt{\pi}} S_i(\Theta) f(\mathbf{y}_i | \hat{\tau}T_i + \sqrt{2}x_t, z_i = 0)}{\sum_{t=1}^T \frac{w_t}{\sqrt{\pi}} f(\mathbf{y}_i | \hat{\tau}T_i + \sqrt{2}x_t, z_i = 0)}. \quad (4.3.8)$$

4.3.2 Maximization Step

Estimation of Parameters in Latent Class

The expected score function for p can be written as

$$\begin{aligned} E_{b,z}S_i(p) &= \int_{-\infty}^{\infty} \sum_z \frac{z_i - p}{p(1-p)} h(b_i, z_i | \mathbf{y}_i) db_i. \\ &= \frac{E_{b,z}(z_i) - p}{p(1-p)} \end{aligned}$$

Solving $\sum_{i=1}^n S_i(p) = 0$ results in a closed form solution

$$\hat{p} = \frac{\sum_{i=1}^n E_{b,z}(z_i)}{n},$$

the estimate of the probability of membership in the unsusceptible class.

Estimation of Class-Specific Parameters

As demonstrated above, parameters in the measurement model and structural model are only estimated in the susceptible class.

Estimation of parameters in the structural model

For the estimation of τ , define

$$\begin{aligned}
E_{h(b,z)}(S_i(\tau)) &= \int_{-\infty}^{\infty} \sum_{z_i=0}^1 S_i(\tau) h(b_i, z_i \mid \mathbf{y}_i, \boldsymbol{\theta}) db_i \\
&= \int_{-\infty}^{\infty} \sum_{z_i=0}^1 (1 - z_i) T_i(b_i - \tau T_i) h(b_i, z_i \mid \mathbf{y}_i, \boldsymbol{\theta}) db_i \\
&= \int_{-\infty}^{\infty} (1 - 0) T_i(b_i - \tau T_i) h(b_i, z_i = 0 \mid \mathbf{y}_i, \boldsymbol{\theta}) db_i \\
&\quad + \int_{-\infty}^{\infty} (1 - 1) T_i(b_i - \tau T_i) h(b_i, z_i = 1 \mid \mathbf{y}_i, \boldsymbol{\theta}) db_i \\
&= T_i \int_{-\infty}^{\infty} b_i h(b_i, z_i = 0 \mid \mathbf{y}_i, \boldsymbol{\theta}) db_i - \tau T_i^2 \int_{-\infty}^{\infty} h(b_i, z_i = 0 \mid \mathbf{y}_i, \boldsymbol{\theta}) db_i.
\end{aligned} \tag{4.3.9}$$

The first component of Equation 4.3.9 is

$$\begin{aligned}
\int_{-\infty}^{\infty} b_i h(b_i, z_i \mid \mathbf{y}_i, \boldsymbol{\theta}) db_i &= \frac{\int_{-\infty}^{\infty} b_i f(\mathbf{y}_i \mid b_i, z_i = 0) f(b_i \mid z_i = 0) f(z_i = 0) db_i}{\int_{-\infty}^{\infty} f(\mathbf{y}_i \mid b_i, z_i = 0) f(b_i \mid z_i = 0) f(z_i = 0) db_i} \\
&= \frac{(1 - p) \int_{-\infty}^{\infty} b_i f(\mathbf{y}_i \mid b_i, z_i = 0) f(b_i \mid z_i = 0) db_i}{(1 - p) \int_{-\infty}^{\infty} f(\mathbf{y}_i \mid b_i, z_i = 0) f(b_i \mid z_i = 0) db_i} \\
&= \frac{\int_{-\infty}^{\infty} b_i f(\mathbf{y}_i \mid b_i, z_i = 0) f(b_i \mid z_i = 0) db_i}{\int_{-\infty}^{\infty} f(\mathbf{y}_i \mid b_i, z_i = 0) f(b_i \mid z_i = 0) db_i}
\end{aligned}$$

and

$$\int_{-\infty}^{\infty} h(b_i, z_i \mid \mathbf{y}_i, \boldsymbol{\theta}) db_i = 1;$$

therefore

$$E_{h(b,z)}(S_i(\tau)) = \frac{T_i \int_{-\infty}^{\infty} b_i f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) db_i}{\int_{-\infty}^{\infty} f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) db_i} - \tau T_i^2.$$

Solving for τ we obtain

$$\hat{\tau} = \frac{\frac{\sum_{i=1}^n T_i \int_{-\infty}^{\infty} b_i f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) db_i}{\int_{-\infty}^{\infty} f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) db_i}}{\sum_{i=1}^n T_i^2}.$$

Estimation of parameters in the measurement model

Recall that the measurement model specifies the relationships between individual scale items and the continuous latent variable. The estimation of the $\alpha_k^{(j)}$ and β_j parameters, thresholds and factor loadings in the measurement model, involves the expectations of score functions and elements of the information matrix with respect to the posterior distribution. Expected score functions and second partial derivatives of class-specific parameters in the MZIPO model can be expressed in the form

$$S_i(\boldsymbol{\theta}_j) = (1 - z_i) S_i^*(\boldsymbol{\theta}_j).$$

For $z_i = 0$, the susceptible class,

$$E_{h(b_i, z_i = 0)}(S_i(\boldsymbol{\theta}_j)) = \frac{\int_{-\infty}^{\infty} S_i^*(\boldsymbol{\theta}_j) f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) db_i}{\int_{-\infty}^{\infty} f(\mathbf{y}_i | b_i, z_i = 0) f(b_i | z_i = 0) db_i}.$$

For example, the score function of $\alpha_k^{(j)}$ is given by

$$S_i(\alpha_k^{(j)}) = (1 - z_i) \left[I[y_{ij} = k] \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]} \right] - I[y_{ij} = k + 1] \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]} \right] \right].$$

If we define

$$S_i^*(\alpha_k^{(j)}) = \left[I[y_{ij} = k] \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]} \right] - I[y_{ij} = k + 1] \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]} \right] \right]$$

then $E_{h(b,z)}(S_i(\alpha_k^{(j)}))$ can be approximated using Equation 4.3.8. The remaining score functions and second partial derivatives (shown in Appendix C) for parameters in the measurement model can be approximated similarly.

Fisher Scoring Algorithm for Estimation of Parameters in Measurement Model

For parameters in the measurement model, the solution to $E_{b,z} S_i(\boldsymbol{\alpha}_j) = 0$ is not available in closed form. Therefore, a Fisher scoring algorithm is used. For the

estimation of β_j , we first consider a Taylor series expansion of the expected score

$$\begin{aligned} E_{b,z}S_i(\hat{\beta}_j) &\approx E_{b,z}S_i(\beta_j) + (\hat{\beta}_j + \beta_j)\frac{\partial}{\partial\beta_j}E_{b,z}S_i(\beta_j) \\ &\approx (\hat{\beta}_j + \beta_j)I_i(\beta_j) \end{aligned}$$

because $E_{b,z}S_i(\hat{\beta}_j) = 0$ and under regularity conditions expectation and differentiation can be interchanged.

Note that

$$I_i(\beta_j) = -\frac{\partial}{\partial\beta_j}E_{b,z}S_i(\beta_j) \tag{4.3.10}$$

is the i^{th} individual's contribution to the observed Fisher information. Taking the expectation of the observed Fisher information with respect to \mathbf{y}_i yields

$$\begin{aligned} J_i(\beta_j) &= -E_{y_i}\frac{\partial}{\partial\beta_j}E_bS_i(\beta_j) \\ &= -E_{y_i}\int_{-\infty}^{\infty}\sum_z\frac{\partial}{\partial\beta_j}S_i(\beta_j)h(b,z|\mathbf{y}_i)db_i \\ &= -E_{y_i}\int_{-\infty}^{\infty}\sum_z\left\{\frac{\partial}{\partial\beta_j}S_i(\beta_j)\right\}h(b,z|\mathbf{y}_i)db_i - E_{y_i}\int\sum_zS_i(\beta_j)\left\{\frac{\partial}{\partial\beta_j}h(b,z|\mathbf{y}_i)\right\}db_i \\ &= -\int_{-\infty}^{\infty}\sum_zE_{y_i}\left\{\frac{\partial}{\partial\beta_j}S_i(\beta_j)\right\}h(b,z|\mathbf{y}_i)db_i - \int\sum_zE_{y_i}S_i(\beta_j)\left\{\frac{\partial}{\partial\beta_j}h(b,z|\mathbf{y}_i)\right\}db_i \\ &= -\int_{-\infty}^{\infty}\sum_zE_{y_i}\left\{\frac{\partial}{\partial\beta_j}S_i(\beta_j)\right\}h(b,z|\mathbf{y}_i)db_i. \end{aligned}$$

Replacing $I_i(\beta_j)$ with $J_i(\beta_j)$ and solving Equation 4.3.10 for β_j yields

$$\beta_j^{(l+1)} = \beta_j^l + J^{-1}(\beta_j)E_{b,z}S(\beta_j). \quad (4.3.11)$$

The individual components in Equation 4.3.11 can be approximated using Gauss-Hermite quadrature. A similar Fisher scoring algorithm for the estimation of $\alpha_j^{(j)}$ is used. For categories $k = 0, \dots, K - 2$

$$\alpha_{k_j}^{(j)(l+1)} = \alpha_{k_j}^{(j)(l)} + J^{-1}(\alpha_{k_j}^{(j)})E_{b,z}S(\alpha_{k_j}^{(j)}) \quad (4.3.12)$$

and for category $k = K_j - 1$

$$\alpha_{K_j-1}^{(j)(l+1)} = \alpha_{K_j-1}^{(j)(l)} + J^{-1}(\alpha_{K_j-1}^{(j)})E_{b,z}S(\alpha_{K_j-1}^{(j)}). \quad (4.3.13)$$

The components of Equations 4.3.12 and 4.3.13 are approximated using Gauss-Hermite quadrature. Estimation of the MZIPO model was performed using SAS 9.2 IML.

4.3.3 Predictive validity

An essential step in model checking for any latent class model is the evaluation of latent class assignment. It is necessary to determine whether the model accurately classifies subjects into unobserved subgroups. Although the MZIPO model

is estimated using data from the baseline portion of the trial, assessing the predictive validity of class assignments in the model is performed by comparing the “gold standard” diagnostic measure for the two latent classes at subsequent time points in the study. Significant differences in the “gold standard” between the latent classes at time points after baseline indicate the predictive validity of the class assignment. The differences in the mean of the “gold standard” between the classes are evaluated with t-tests.

4.4 Example

The motivating example for this chapter comes from the PAL clinical trial. As described generally in Chapter 1 and more specifically in Chapter 3, a point of contention in lymphedema research is that incidence of lymphedema varies drastically in the literature, and one explanation is that a variety of diagnostic measures are used. In addition to swelling, self-reported symptom items such as those in the Norman Lymphedema Survey (Norman et al., 2001) could potentially be useful indicators of lymphedema incidence or exacerbation. However, in the PAL study the distribution of self-reported severity symptoms in the Norman Lymphedema survey was substantially zero-inflated. The PAL trial was designed such that half of the participants were previously diagnosed with lymphedema and half were not, which could partially explain the excess of zero responses. However, there is the potential for undiagnosed lymphedema at baseline as well as the potential for exercise inter-

vention to bring about lymphedema symptoms. The objective of the analysis was to use the MZIPO model to identify important Norman lymphedema symptoms and establish criterion validity compared to objective measures while simultaneously accounting for the zero-inflation that occurred in the Norman symptoms using a latent class component.

The sample comes from $N = 280$ women with complete data from the baseline portion of the PAL trial. Outcome measures included 7 severity items from the Norman Lymphedema Survey, as described in Chapter 3. Six of the items were chosen based on results of EFA models (clothing too tight, puffiness, skin feels leathery, arm feels tired, pain, difficulty writing) and a seventh item (swelling after exercise) was chosen from the scale as a potential “junk” item. Volume difference percentage between affected and unaffected arms was considered the “gold standard” diagnostic measure. In contrast to Chapter 3, the continuous volume difference measure was recoded into seven ordered categories with 5 percentage points in each category. This rescaling preserves the linear nature of the “gold standard” measure but eliminates an identifiability issue in modeling limb volume difference as continuous.

4.5 Results

Frequency distributions for the Norman Lymphedema Survey items are shown in Figure 4.1 for the full sample. Each item has an excess of zeros, with the percentage of zeros for each item ranging from 53.24% ($N=148$) for puffiness to 88.49% ($N=246$)

Table 4.1: Baseline characteristics for complete data sample of PAL data.

	Stratified by Latent Class						P value
	Full Sample (N=278)		Susceptible (N=217)		Unsusceptible (N=61)		
Age							
Mean, SD	56.13	8.87	56.50	8.92	54.78	8.63	0.1825
Education (Freq., Percent)							
High school or less	43	15.47	36	16.59	7	11.48	0.5994
Some college	95	34.17	74	34.10	21	34.43	
College or postgraduate	140	50.36	107	49.31	33	54.10	
Race							
White	183	65.83	136	62.67	47	77.05	0.0592
Black	83	29.86	72	33.18	11	18.03	
Other	12	4.32	9	4.15	3	4.92	
Marital Status							
Never married	31	11.27	24	11.21	7	11.48	0.2585
Currently married/living with partner	168	61.09	125	58.41	43	70.49	
Divorced/separated	49	17.82	41	19.16	8	13.11	
Widowed	27	9.82	24	11.21	3	4.92	
Occupation							
Professional	110	40.00	83	38.79	27	44.26	0.0632
Clerical or service	51	18.55	37	17.29	14	22.95	
Homemaker, student, or unemployed	25	9.09	16	7.48	9	14.75	
Other or unknown	24	8.73	21	9.81	3	4.92	
Retired	65	23.64	57	26.64	8	13.11	
Dominance							
Yes	140	50.36	111	51.15	29	47.54	0.6182
No	138	49.64	106	48.85	32	52.46	
Volume Difference							
Mean, SD	7.57	13.74	9.60	14.68	0.36	5.13	<0.0001
Lymphedema Diagnosis at Baseline							
Yes	133	47.84	126	58.06	7	11.48	<0.0001
No	145	52.16	91	41.94	54	88.52	

for difficulty writing. In general there were more moderate responses than slight responses, but fewer severe and very severe responses than moderate responses. The potential “junk” item, swelling after exercise, demonstrated a similar distribution although there were no very severe responses.

Table 4.1 summarizes baseline characteristics, presented for the complete data sample and stratified by latent class as predicted by the MZIPO model. Average age for the full sample was 56.13 (SD=8.87), and the majority of study participants were white (65.83%). About half of the sample (N=140, 50.36%) earned at

least a college degree, and the majority were currently married or living with a partner (N=168, 61.09%). Of the full sample, mean percent limb volume difference between affected and unaffected arms was 7.57 (SD=13.73). Almost half of the complete data sample was diagnosed with lymphedema at baseline (N=133, 47.84%), by trial design. Table 4.1 also presents baseline characteristics stratified by latent class assignment. The MZIPO model estimated that N=61 of the full sample could be classified as unsusceptible, while N=217 were classified as susceptible to lymphedema symptoms. Most baseline characteristics were not significantly different between the two groups with the exception of limb volume difference and lymphedema diagnosis. Mean volume difference in the susceptible class was 9.60 percent (SD=14.68) while mean volume difference in the unsusceptible class was 0.36 percent (SD=5.13; $p < 0.001$). Among subjects in the susceptible class, 58.06% (N=126) were diagnosed with lymphedema at baseline compared to only 11.48% (N=7) in the unsusceptible class ($p < 0.001$).

4.5.1 Ordered Categorical MIMIC Models vs. MZIPO Model

Model results are featured in Table 4.2. This table compares an ordered categorical MIMIC model from the full sample (Model 1), an ordered categorical MIMIC model where observed non-responders have been removed (Model 2), and finally the MZIPO model (Model 3). All models include the 7 symptom items chosen from previous exploratory factor analysis.

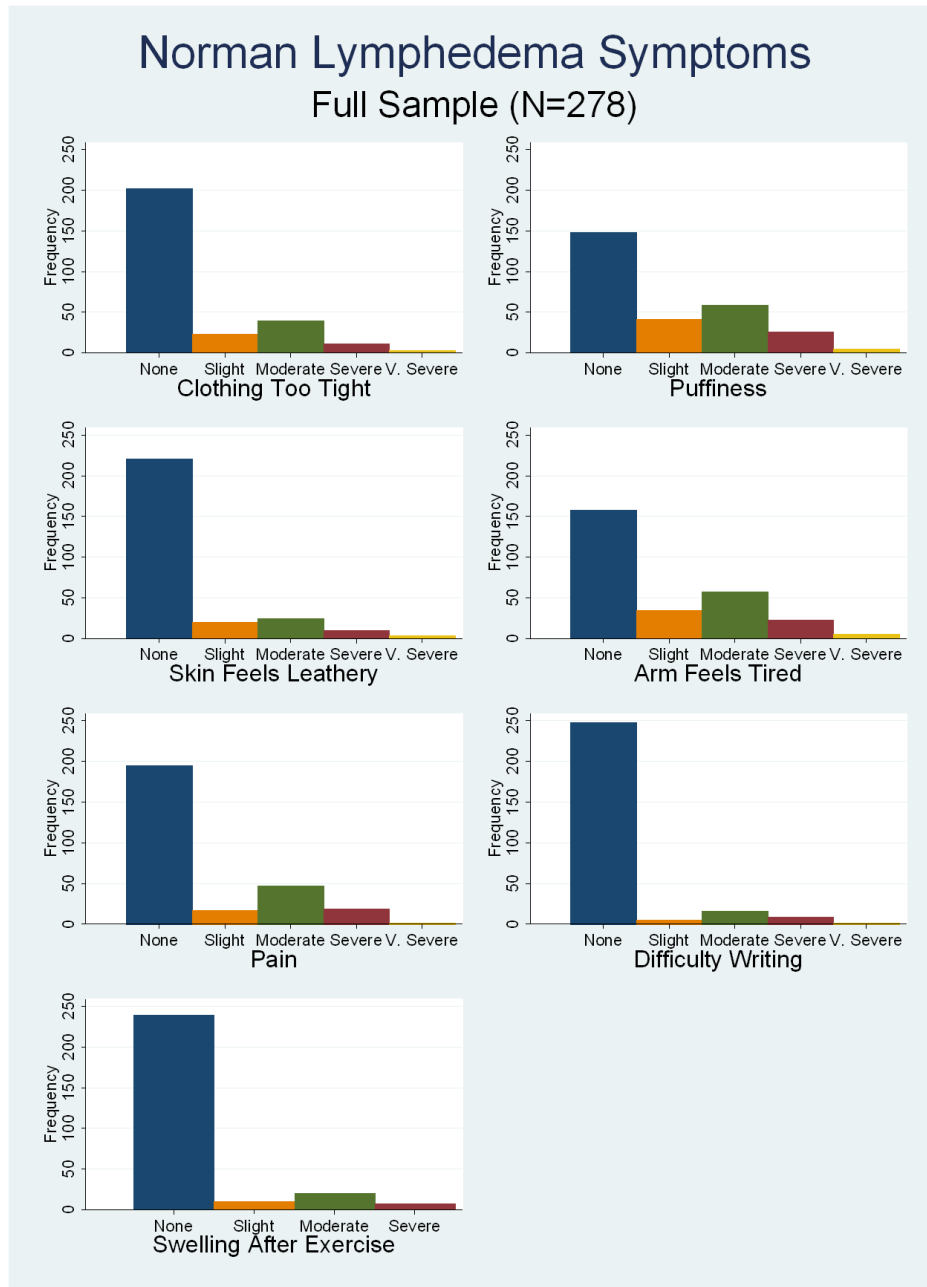


Figure 4.1: Histograms of Norman Lymphedema Survey items for N=278 complete data sample.

Model 1 yielded particularly high factor loading estimates for clothing too tight and puffiness ($\beta=6.4971$ and $\beta=4.6258$ respectively) and moderate factor loadings

Table 4.2: Ordinal MIMIC model vs. MZIPO model for Norman Lymphedema Survey items.

	Model 1: Ordinal MIMIC Full Sample (N=278)		Model 2: Ordinal MIMIC Observed Responders Only (N=170)		Model 3: MZIPO	
Proportion unsusceptible (p)	NA		NA		0.2194	
Assoc. with “gold standard” (τ)	0.0707		0.0300		0.0759	
Factor loadings (β)						
Clothing too tight	6.4971		3.0118		6.4895	
Puffiness	4.6258		3.4988		3.8520	
Skin feels leathery	2.9493		1.5945		1.9964	
Arm feels tired	2.5653		1.3328		1.7772	
Pain	1.5514		0.5488		0.9159	
Difficulty writing	1.8096		0.8798		1.3618	
Swelling after exercise	1.4454		0.2382		0.9583	
Thresholds (α)						
Clothing too tight						
α_0	2.3622		-0.6503		1.8536	
α_1	3.8102		0.5207		3.4778	
α_2	8.3107		3.9043		8.3723	
α_3	11.6256		6.7411		13.0676	
Puffiness						
α_0	-0.6084		-3.4946		-1.3867	
α_1	1.3499		-1.2159		0.6787	
α_2	4.9343		2.6984		4.0911	
α_3	8.2182		6.7802		7.5530	
Skin feels leathery						
α_0	2.0497		0.4721		1.4489	
α_1	2.8585		1.2356		2.1731	
α_2	4.5022		2.8257		3.7168	
α_3	6.2383		4.4958		5.2374	
Arm feels tired						
α_0	-0.0228		-1.5520		-0.5056	
α_1	0.9747		-0.4693		0.4815	
α_2	3.1981		1.7281		2.5479	
α_3	5.3502		3.8440		4.5403	
Pain						
α_0	0.8382		-0.1197		0.4623	
α_1	1.2258		0.2998		0.8297	
α_2	2.9070		1.9455		2.4726	
α_3	5.5618		4.4901		4.9728	
Difficulty writing						
α_0	2.3777		1.4241		1.9845	
α_1	2.5747		1.6366		2.1854	
α_2	3.7197		2.7146		3.2995	
α_3	5.6751		4.5897		5.2319	
Swelling after exercise						
α_0	2.0808		1.1561		1.6648	
α_1	2.5138		1.5614		2.0909	
α_2	4.0007		2.9965		3.5533	
Estimated latent lymphedema severity						
Mean, SD	0.1955	0.8264	0.2024	0.8635	0.0922	0.8706

for other items. The smallest factor loading in Model 1 was $\beta=1.4454$ for the potential “junk” item, swelling after exercise, but this loading was not much smaller than the factor loadings for several other items. Factor loading estimates measure

the strength of the relationship between an individual symptom and the continuous measure of latent lymphedema severity, where larger factor loadings indicate a stronger association between the item and the latent variable. The estimate of the coefficient of the regression of latent lymphedema severity on volume difference for Model 1 was $\tau=0.0707$. As in Chapter 3, this is a global measure of the “gold standard” on all Norman symptom items. The coefficient can be interpreted as the average increase in latent lymphedema severity for each 5% increase in volume difference. The regression coefficient allows us to evaluate criterion validity in that if there is a strong relationship between the “gold standard” and the set of symptoms, this is an indication that these two measures are highly associated with one another.

In Model 2 the ordered categorical MIMIC model was fit on the subset of the sample classified as observed responders. Factor loadings from Model 2 were all smaller than those in Model 1. Consistent with Model 1 results, clothing too tight and puffiness still produced the highest factor loadings at $\beta=3.0118$ and $\beta=3.4988$. One item whose factor loading changed dramatically from Model 1 to Model 2 was swelling after exercise. The smallest loading in Model 2, the factor loading for swelling after exercise, $\beta=0.2382$, was markedly smaller than the factor loading from Model 1. Furthermore, the estimate of the regression of latent lymphedema severity on the gold standard, $\tau=0.03$, decreased in Model 2 compared to Model 1.

Finally, Model 3 provides estimates from the MZIPO model. The predicted

probability of membership in the unsusceptible class was 21.94% of the original full sample. Conditional parameter estimates from the measurement model were fit only on the susceptible class, which comprised 78.06% of the sample. Generally, factor loading estimates for Model 3 were similar to those of Model 1, but most were slightly smaller. Factor loadings ranged from $\beta=0.915$ for pain to $\beta=6.490$ for clothing too tight. The estimate of the regression of latent lymphedema severity on the “gold standard” was $\tau=0.076$, similar to that of Model 1. One important difference in Model 3 compared to Model 1 is that swelling after exercise no longer had the smallest factor loading. This could be an indication that perhaps it is not truly a “junk” item.

4.5.2 Predictive Validity

In this example, it was clinically useful to determine at baseline subgroups that might or might not be susceptible to experiencing symptoms over time. Figure 4.2 presents a profile plot of mean volume difference at baseline, 3, 6, and 12 months for each of the latent classes. In each group, mean volume difference remains constant over time. However, means between the two classes differ at each time point. T-tests of volume difference at time points after baseline show that these differences were significant ($p < 0.0001$). These tests provided an early indication of the predictive validity of latent class assignment of the MZIPO model.

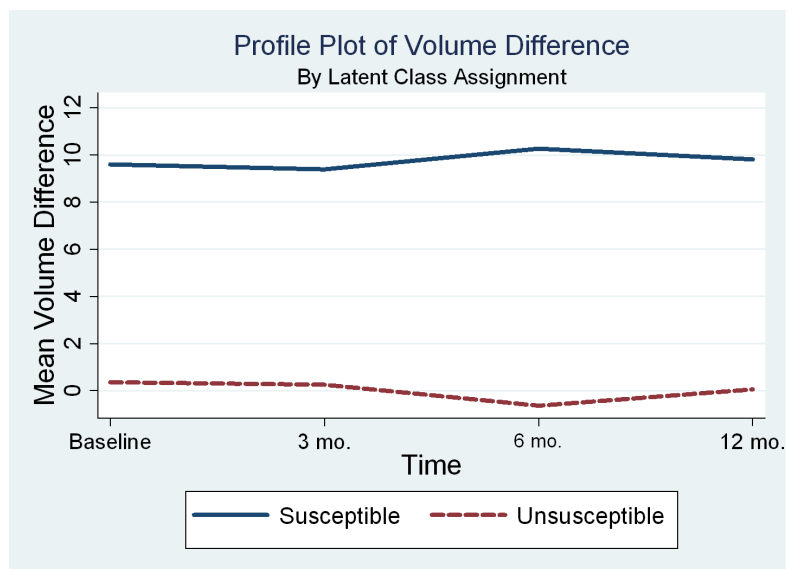


Figure 4.2: Profile plot of mean volume difference by predicted latent class assignment.

4.5.3 Evaluation of Observed Zeros and Latent Class Membership

Table 4.3 presents a cross-tabulation of latent class assignment as determined by the MZIPO model and observed symptom response. As described in Section 4.2, if a subject is in the observed response group, she cannot be assigned to the unsusceptible class, so all $N=170$ subjects with a response were classified as susceptible. Of the non-responders, 61 were classified as unsusceptible and 47 were classified as susceptible, despite zero responses on all 7 symptom items.

Table 4.4 presents several baseline characteristics for these three groups. It was expected that there would be differences among the three groups in terms of volume difference and baseline lymphedema diagnosis, and these differences pro-

Table 4.3: Comparison of latent class assignment and observed symptom response.

		Observed Symptom Responses		Total
		Non-Response	Response	
Latent Class	Unsusceptible	61	0	61
	Susceptible	47	170	217
Total		108	170	278

Table 4.4: Characteristics of predicted latent class assignment and observed symptom response.

	Stratified by Latent Class and Observed Response				P value	
	NR/S (N=47)		NR/UN (N=61)			R/S (N=170)
Volume Difference						
Mean (SD)	2.45	(7.22)	0.36	(5.13)	11.57 (15.61)	<0.0001
Baseline Lymph. Diagnosis						
Yes	8	17.02	7	11.48	118 69.41	<0.0001
No	39	82.98	54	88.52	52 30.59	

NR/UN=Non-response, Unsusceptible class; NR/S=Non-response, Susceptible class;
R/S=Response/Susceptible class

vide an indication that the MZIPO model does sufficiently well in distinguishing among subjects who are susceptible and unsusceptible. In particular, I expected the responder/susceptible group to have the highest average volume difference and greatest percentage of baseline diagnosis among the three groups, and that proved to be the case. The responder/susceptible group had an average volume difference of 11.57 (SD=15.61) and 69.41% of that group was diagnosed with lymphedema at baseline. Of particular interest was the comparison of the non-responder/susceptible group and the non-responder/unsusceptible group. The non-responder/susceptible group had an average volume difference of 2.45 (SD=7.22), compared to 0.36 (SD=5.13) for the non-responder/unsusceptible group. Furthermore, 17.02% of the non-responder/susceptible group and 11.48% of the non-

responder/susceptible group were diagnosed with lymphedema at baseline. That the non-responder/susceptible group had a higher mean volume difference and greater percentage of lymphedema diagnosis at baseline than the non-responder/unsusceptible provides an indication that latent class assignment is informative in distinguishing among non-responders.

4.6 Discussion

As in many other applications measuring symptom scales at baseline, data from the Norman Lymphedema Survey measured at baseline as part of the PAL trial exhibited zero-inflation in the distribution of scale items. To correct for this, I developed the MZIPO model, an intuitive model that exploits the use of two types of latent variables in a single model. Building from recent work in the latent variable literature as well as the univariate ZIPO model, the MZIPO model allows for the estimation of a traditional joint latent variable model while accounting for the presence of excess zeros with a latent class component. If I had employed a ZIPO model to assess each item separately, this approach would have separated out some of the observed non-responders. However, this method would not guarantee that the subgroup removed from the original sample would be the same as if all items were considered in a model together. In the case of multiple scale items, it was important to consider not only the relationship between the items and the “gold standard” diagnostic measure individually, but also the relationship among the items.

In this chapter, I considered several approaches for the evaluation of multiple items simultaneously and the evaluation of the relationship between those items and the “gold standard” in the presence of zero-inflation. The simplest approach was to fit an ordered categorical MIMIC model on the entire original sample (Model 1). While this approach incorporates all subject information, parameter estimates may not be entirely accurate because of excess of zeros in the item distributions. Another approach was to remove all observed non-responders from the sample and fit an ordered categorical MIMIC model on this subset of the original sample. This model had the advantage of less skewed item distributions, but in the example of the PAL trial data almost 40% of the original sample was removed. I advocate the use of the MZIPO model (Model 3) over a standard ordered categorical MIMIC model. In the MZIPO model, 21.9% of the original sample was classified as unsusceptible and therefore not included in the estimation of the conditional parameters in the measurement and structural components of the model. The MZIPO model allowed us to account for some of the zero-inflation in the item distributions while still including useful information on the “gold standard” from more subjects than if all observed non-responders were removed from the sample. Furthermore, the MZIPO model provided predicted latent class assignments that could be clinically useful in predicting lymphedema onset and flare-ups.

The use of the MZIPO model to define latent classes is particularly relevant to the PAL trial application. Many women do not experience symptoms of lym-

lymphedema at baseline, perhaps because they have not been diagnosed with lymphedema yet, are unable to identify lymphedema symptoms, or because they are truly not susceptible to experiencing symptoms. The ability to classify women who might be susceptible to symptoms at an early stage is essential because these women may require extra clinical evaluations and careful monitoring for lymphedema onset or exacerbation. Therefore, latent classes that exhibit predictive validity could have great clinical utility. The results from the MZIPO model and evaluation of the latent classes at subsequent time points in the PAL trial indicated predictive validity for latent class assignments. T-tests of mean volume difference at time points after baseline indicate predictive validity of the latent classes, which is significant because baseline classification into the susceptible or unsusceptible group could be useful in predicting who is likely to develop lymphedema exacerbations or flare-ups in the future.

The comparison of baseline characteristics among the 3 observed response and latent class groups also provides indication of the utility of the MZIPO model. A particularly interesting group is the non-response/susceptible class group because this group's information is included in Model 3 but not Model 2. As shown in the table, this group had a higher mean volume difference and a greater percentage of baseline lymphedema diagnosis than the non-response/unsusceptible group. These differences are an indication that this group was correctly classified into the susceptible class. Even though the subjects in this group did not experience symptoms at

baseline they may be susceptible to symptoms later on.

In sum, the MZIPO model is a complicated latent variable model and there are several issues to consider that will require future work. The two most pressing issues are statistical tests for parameter estimates and the evaluation of model assumptions. First, a means of determining statistical significance for the parameter estimates is necessary for model comparison. A likelihood ratio test could be a reasonable option to test the regression coefficient in the model without deriving the complicated standard error formulas necessary for the Wald or Score tests. Furthermore, testing the significance of factor loadings is an ultimate goal that will allow for added objectivity in the item selection process. Next, simulations are needed to assess the assumption that the latent class and the continuous latent variable are independent of one another. This assumption facilitated model estimation, but in some cases may not be valid. If simulations show that the MZIPO model is not robust to violations of this assumption, it is possible that the model will have to be reformulated to allow for the relationship between the latent class and the continuous latent variable.

In sum, I have shown that the MZIPO model can be a useful tool in determining the relationship among scale items and between scale items and a “gold standard” in the presence of unobserved heterogeneity. While more work is required to assess the practical utility of the MZIPO model, the work presented here illustrates a great potential for clinical usefulness because of the model’s ability to identify patient

subgroups that require extra attention.

Chapter 5

Conclusions

In this dissertation the aim was to investigate the use of joint latent variable models for multivariate outcomes in order to perform item selection and validation of symptom scales simultaneously. It has been shown that using joint latent variable models as an alternative to the typical two-stage process of item selection and validation is preferable because joint models eliminate the bias inherent in estimating the continuous latent variable. Joint latent variables provide estimates of the relationship between items in a scale as well as the relationship between the items and “gold standard” diagnostic measures. These models are particularly flexible as they accommodate a variety of response types and allow for covariates on the observed or latent variables.

In Chapter 2, I utilized a joint latent variable model, the MIMIC model, to simultaneously identify important indicators of PMS severity and to compare these

indicators to a clinician-rated “gold standard” measure. I developed a Score test for individual factor loadings in the MIMIC model subject to the model constraint that factor loadings be non-negative. Using this statistical test I determined that three items in the PMS symptom scale could be considered for removal from the scale, which could ultimately improve scale accuracy.

Chapter 3 featured an application of an extension to the traditional MIMIC model for ordered categorical scale items. This model was applied to data from the Norman Lymphedema Survey in the PAL clinical trial to evaluate symptom items at baseline and compare them to a current “gold standard” diagnostic measure. The ordered categorical formulation of the MIMIC model identified a potential “junk” item as not statistically significant, providing an indication that this item could indeed be “junk.” An investigation of the item distributions in the Norman Lymphedema Survey revealed a substantial amount of symptom non-response at baseline, which served as motivation for the development of the MZIPO model proposed in Chapter 4.

In order to account for zero-inflation in the item distributions of the Norman Lymphedema Scale, in Chapter 4 I proposed a new joint latent variable model that extended the ordered categorical MIMIC model to incorporate a latent class component to classify subjects as either susceptible or unsusceptible. The MZIPO model predicted that 21% of the original complete data sample from the PAL trial were in the unsusceptible class and provided estimates of factor loadings and the

coefficient of the regression on the latent measure of lymphedema severity for the susceptible class. This model was preferable to fitting the MIMIC model on the complete sample because it reduced the skewness of the item distributions, and it was also a preferable alternative to simply removing all observed non-responders from the sample because it allowed more subjects (i.e., all those with information on limb volume difference) to remain in the sample.

Overall, I have shown in this dissertation the utility of a variety of latent variable methods for scale development. These models have a number of advantages compared to traditional procedures. First, latent variable models in general can be considered dimension reduction techniques for multivariate outcomes. Many times in clinical research a biological process is measured by multiple physical responses, and latent variable models provide a means of summarizing these measures in a single model. Next, joint latent variable models in particular can yield greater scale accuracy and reduced participant burden by identifying the most important indicators of the latent variable and eliminating “junk” information. Finally, in terms of statistical modeling, using latent variable models for multiple observed outcomes is advantageous because of the conditional independence assumption that is often made. This assumption states that given the underlying constructs, the individual observed outcomes are independent of one another, an incredibly useful assumption that we took advantage of in order to ease estimation of the MZIPO model.

Although the general use of latent variable methods is highly recommended in

this dissertation, there are some drawbacks to these types of models that should be considered. First, the models are highly parameterized. Identifiability is a significant hurdle in many latent variable models, and most of the time it is necessary to place constraints on the model that may or may not be valid in order to get around this issue. Next, latent variable models require an in depth understanding of the confirmatory framework among variables. In models such as the MIMIC model or the MZIPO model, there are strong assumptions made about the nature and direction of the relationship between observed variables. Using these models appropriately in clinical research requires comprehensive knowledge of the substantive application. Finally, reproducibility is a critical concern in latent variable modeling. Factor structures, either in a factor analysis model or in the measurement model of a joint latent variable, are sometimes not easily reproduced given a different sample. For this reason, validation is extremely important when using these types of latent variable models.

Several issues presented in this dissertation warrant further consideration. First, it is of interest to further investigate the modified Score test for factor loadings in the MIMIC model presented in Chapter 2. There is a possibility that simulating data with an association between latent lymphedema severity and the “gold standard” diagnostic measure induces correlation among the items. If this is the case, we cannot separately assess the effect of correlation among items and association between items and the gold standard. In future work I will perform simulations to

look at the effect of both of these separately on the performance of the Score test for a potential “junk” item. Additionally, I will perform simulations to determine how the modified Score test performs in the presence of non-normally distributed scale items. If the modified Score test is robust to departures from normality, this will be another reason why it may be preferable to the traditional Wald test. Furthermore, in many clinical examples data are less than perfect, so it is important to determine how a statistical test will behave when scale item distributions are skewed. Overall, further simulations will provide a better idea of how useful the modified Score test will be in practice.

In addition to further evaluation of the modified Score test for factor loadings in the MIMIC model, I will also continue to build on the MZIPO model. After assessing the model assumptions and deriving standard errors for statistical tests, as described in Chapter 4, I would like to expand the MZIPO model to include covariates on the latent class assignment. Extending the model to include covariates will be particularly relevant to applications of the model to clinical research, where demographic characteristics or comorbidities may influence subgroup classification. The ultimate goal of this future work is to obtain latent class assignments that can accurately predict at baseline those subjects who are susceptible to future lymphedema exacerbations or onset.

In conclusion, the joint latent variable methods that I have proposed and implemented in this dissertation provide flexible approaches for scale development.

Future research will help to further elucidate the complexities of these methods with the ultimate goal of producing clinically relevant models that are useful in practice.

Appendix A

Derivations for Score Test of Factor Loadings in MIMIC Model

A.1 Score Function for λ

$$\frac{dl(y; \theta)}{d\lambda} = \frac{d}{d\lambda} \left[-\frac{n}{2} \ln(|\Sigma|) - \frac{1}{2} \sum_{i=1}^n \{(y_i - (\mu + \lambda\beta z_i))^\top \Sigma^{-1} (y_i - (\mu + \lambda\beta z_i))\} \right],$$

where

$$\frac{d \ln(|\Sigma|)}{d\lambda} = \frac{d \ln(|\Sigma|)}{d\Sigma} \star \frac{d\Sigma}{d\lambda}$$

and \star refers to the star product (MacRae, 1974).

$$\frac{d \ln(|\Sigma|)}{d\Sigma} = [2\Sigma^{-1} - \Sigma^{-1} \circ I_m]$$

where \circ refers to the Hadamard product.

$$\frac{d\Sigma}{d\lambda} = E_{(m,1)} \lambda^\top + I_m \otimes \lambda$$

so

$$\frac{d \ln(|\Sigma|)}{d\lambda} = [2\Sigma^{-1} - \Sigma^{-1} \circ I_m] \star [E_{(m,1)}\lambda^\top + I_m \otimes \lambda].$$

Next,

$$\begin{aligned} & \frac{d \sum_{i=1}^n \{(y_i - (\mu + \lambda \beta z_i))^T \Sigma^{-1} (y_i - (\mu + \lambda \beta z_i))\}}{d\lambda} \\ &= \sum_{i=1}^n \frac{d(y_i - (\mu + \lambda \beta z_i))^T \Sigma^{-1} (y_i - (\mu + \lambda \beta z_i))}{d\lambda} = \sum_{i=1}^n \frac{d(y_i - (\mu + \lambda \beta z_i))^T \Sigma^{-1} d(y_i - (\mu + \lambda \beta z_i))}{d\lambda}, \end{aligned}$$

where

$$\begin{aligned} \frac{d(y_i - (\mu + \lambda \beta z_i))^T}{d\lambda} &= \frac{d - \lambda^T \beta z_i}{d\lambda} \\ &= -\beta z_i I_m \end{aligned}$$

and

$$\begin{aligned} \frac{d \Sigma^{-1}}{d\lambda} &= -(\Sigma^{-1} \otimes I_m) \left(\frac{d \Sigma}{d\lambda} \right) (\Sigma^{-1}) \\ &= -(\Sigma^{-1} \otimes I_m) (E_{(m,1)} \lambda^T + I_m \otimes \lambda) (\Sigma^{-1}). \end{aligned}$$

Therefore,

$$\begin{aligned} & \frac{d \sum_{i=1}^n \{(y_i - (\mu + \lambda \beta z_i))^T \Sigma^{-1} (y_i - (\mu + \lambda \beta z_i))\}}{d\lambda} \\ &= \sum_{i=1}^n -2 \Sigma^{-1} \beta z_i (y_i - (\mu + \lambda \beta z_i)) \\ &\quad - ((y_i - (\mu + \lambda \beta z_i))^T \otimes I_m) (\Sigma^{-1} \otimes I_m) (E_{(m,1)} \lambda^T + I_m \otimes \lambda) \Sigma^{-1} ((y_i - (\mu + \lambda \beta z_i)), \end{aligned}$$

and

$$\begin{aligned}
\frac{dl(y; \theta)}{d\lambda} = & -\frac{n}{2} [2\Sigma^{-1} - \Sigma^{-1} \circ I_m] \star [E_{(m,1)} \lambda^\top + I_m \otimes \lambda] \\
& -\frac{1}{2} \sum_{i=1}^n 2\Sigma^{-1} \beta z_i (y_i - (\mu + \lambda \beta z_i)) \\
& -((y_i - (\mu + \lambda \beta z_i))^\top \otimes I_m)(\Sigma^{-1} \otimes I_m)(E_{(m,1)} \lambda^\top + I_m \otimes \lambda) \Sigma^{-1} ((y_i - (\mu + \lambda \beta z_i)).
\end{aligned}$$

A.2 Information Matrix

The information matrix for parameters in the MIMIC model is given by

$$I(\theta) = -E \frac{d^2 l}{d\theta d\theta^\top} = \begin{bmatrix} I_{\beta\beta} & I_{\beta\mu} & I_{\beta\Psi} & I_{\beta\lambda} \\ I_{\beta\mu}^\top & I_{\mu\mu} & I_{\mu\Psi} & I_{\mu\lambda} \\ I_{\beta\Psi}^\top & I_{\mu\Psi}^\top & I_{\Psi\Psi} & I_{\Psi\lambda} \\ I_{\beta\lambda}^\top & I_{\mu\lambda}^\top & I_{\mu\Psi}^\top & I_{\lambda\lambda} \end{bmatrix}$$

and we are most interested in the element $I_{\lambda\lambda}$. The second partial derivative with respect to λ is given by

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \lambda \partial \lambda^\top} = & \frac{\partial}{\partial \lambda^\top} \left(-\frac{n}{2} [2\Sigma^{-1} - (\Sigma^{-1} \circ I_m)] \star [E_{(m,1)} \lambda^\top + I_m \otimes \lambda] \right. \\ & + \frac{1}{2} \sum_{i=1}^n 2\Sigma^{-1} \beta z_i (\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda} \beta z_i]) + ((\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda} \beta z_i])^\top \otimes I_m) (\Sigma^{-1} \otimes I_m) \\ & \left. \times (E_{(m,1)} \lambda^\top + I_m \otimes \lambda) \Sigma^\top (\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda} \beta z_i]) \right), \end{aligned}$$

where

$$\begin{aligned}
& \frac{\partial}{\partial \lambda^\top} [2\Sigma^{-1} - (\Sigma^{-1} \circ I_m)] \star [E_{(m,1)} \lambda^\top + I_m \otimes \lambda] \\
&= [2\Sigma^{-1} - (\Sigma^{-1} \circ I_m)] \star \frac{\partial [E_{(m,1)} \lambda^\top + I_m \otimes \lambda]}{\partial \lambda^\top} \\
&+ [E_{(m,1)} \lambda^\top + I_m \otimes \lambda] \star (I_m \otimes \text{vec}(I_m) \otimes I_1) \left(\frac{\partial [2\Sigma^{-1} - (\Sigma^{-1} \circ I_m)]}{\partial \lambda^\top} \right) \\
&\times (I_m \otimes \text{vec}^\top(I_1) \otimes I_m) \\
&= [2\Sigma^{-1} - (\Sigma^{-1} \circ I_m)] \star (E_{(m,1)} E_{(1,m)} + (I_M \otimes I_m)) \\
&+ [E_{(m,1)} \lambda^\top + I_m \otimes \lambda] \star (I_m \otimes E_{(m,1)}) \left(-2\Sigma^\top [E_{(m,1)} \lambda^\top + I_m \otimes \lambda]^\top \right. \\
&\quad \left. \times (\Sigma^{-1} \otimes I_m) + (\Sigma^{-1} \circ I_m) [E_{(m,1)} \lambda^\top + I_m \otimes \lambda]^\top ((\Sigma^{-1} \circ I_m) \times I_m) \right) (I_m \otimes I_m)
\end{aligned}$$

and

$$\begin{aligned}
& \frac{\partial 2\Sigma^{-1} \beta z_i (\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda} \beta z_i])}{\partial \lambda^\top} = \frac{\partial 2\Sigma^{-1} \beta z_i}{\partial \lambda^\top} (\mathbf{y}_i - [\boldsymbol{\mu} \\
&\quad + \boldsymbol{\lambda} \beta z_i]) \otimes I_m + (2\Sigma^{-1} \beta z_i) \frac{\partial (\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda} \beta z_i])}{\partial \lambda^\top} \\
&= -2 \left((\Sigma^{-1} \otimes I_1) (E_{(m,1)} \lambda^\top + I_m \otimes \lambda)^\top (\Sigma^{-1} \otimes I_m) \right) \beta z_i \\
&\quad \times ((\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda} \beta z_i]) \otimes I_m) - 2\Sigma^{-1} \beta^2 z_i^2.
\end{aligned}$$

Next,

$$\begin{aligned}
& \frac{\partial((\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i])^\top \otimes I_m)(\Sigma^{-1} \otimes I_m)(E_{(m,1)}\boldsymbol{\lambda}^\top + I_m \otimes \lambda)\Sigma^\top(\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i])}{\partial\boldsymbol{\lambda}^\top} \\
&= \frac{\partial((\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i])^\top \otimes I_m)}{\partial\boldsymbol{\lambda}^\top} \\
&\quad \times \left((\Sigma^{-1} \otimes I_m)(E_{(m,1)}\boldsymbol{\lambda}^\top + I_m \otimes \lambda)\Sigma^\top(\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i]) \otimes I_m \right) \\
&\quad + ((\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i])^\top \otimes I_m) \\
&\quad \times \frac{\partial(\Sigma^{-1} \otimes I_m)(E_{(m,1)}\boldsymbol{\lambda}^\top + I_m \otimes \lambda)\Sigma^{-1}(\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i])}{\partial\boldsymbol{\lambda}^\top},
\end{aligned}$$

where

$$\begin{aligned}
& \frac{\partial((\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i])^\top \otimes I_m)}{\partial\boldsymbol{\lambda}^\top} = I_m(I_m \otimes (-E_{(1,m)}\beta z_i))(I_{(m,m)} \otimes I_m), \\
& \frac{\partial(\Sigma^{-1} \otimes I_m)(E_{(m,1)}\boldsymbol{\lambda}^\top + I_m \otimes \lambda)\Sigma^{-1}(\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i])}{\partial\boldsymbol{\lambda}^\top} \\
&= \frac{\partial(\Sigma^{-1} \otimes I_m)(E_{(m,1)}\boldsymbol{\lambda}^\top + I_m \otimes \lambda)\Sigma^{-1}}{\partial\boldsymbol{\lambda}^\top}((\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i]) \otimes I_m) \\
&\quad + (\Sigma^{-1} \otimes I_m)(E_{(m,1)}\boldsymbol{\lambda}^\top + I_m \otimes \lambda)\Sigma^{-1} \frac{\partial(\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i])}{\partial\boldsymbol{\lambda}^\top} \\
&= \frac{\partial(\Sigma^{-1} \otimes I_m)(E_{(m,1)}\boldsymbol{\lambda}^\top + I_m \otimes \lambda)\Sigma^{-1}}{\partial\boldsymbol{\lambda}^\top}((\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i]) \otimes I_m) \\
&\quad + (\Sigma^{-1} \otimes I_m)(E_{(m,1)}\boldsymbol{\lambda}^\top + I_m \otimes \lambda)\Sigma^{-1}(-I_m\beta z_i),
\end{aligned}$$

$$\begin{aligned}
& \frac{\partial(\Sigma^{-1} \otimes I_m)(E_{(m,1)}\lambda^\top + I_m \otimes \lambda)\Sigma^{-1}}{\partial\lambda^\top} \\
&= \frac{\partial(\Sigma^{-1} \otimes I_m)(E_{(m,1)}\lambda^\top + I_m \otimes \lambda)}{\partial\lambda^\top} (\Sigma^{-1} \otimes I_m) \\
&+ (\Sigma^{-1} \otimes I_m)(E_{(m,1)}\lambda^\top + I_m \otimes \lambda) \frac{\partial\Sigma^{-1}}{\partial\lambda^\top} \\
&= \frac{\partial(\Sigma^{-1} \otimes I_m)(E_{(m,1)}\lambda^\top + I_m \otimes \lambda)}{\partial\lambda^\top} (\Sigma^{-1} \otimes I_m) \\
&+ (\Sigma^{-1} \otimes I_m)(E_{(m,1)}\lambda^\top + I_m \otimes \lambda) \left(-\Sigma^{-1}((E_{(m,1)}\lambda^\top + I_m \otimes \lambda))^\top (\Sigma^{-1} \otimes I_m) \right),
\end{aligned}$$

and

$$\begin{aligned}
& \frac{\partial(\Sigma^{-1} \otimes I_m)(E_{(m,1)}\lambda^\top + I_m \otimes \lambda)}{\partial\lambda^\top} = \frac{\partial(\Sigma^{-1} \otimes I_m)}{\partial\lambda^\top} ((E_{(m,1)}\lambda^\top + I_m \otimes \lambda) \otimes I_m) \\
&+ (\Sigma^{-1} \otimes I_m) \frac{\partial(E_{(m,1)}\lambda^\top + I_m \otimes \lambda)}{\partial\lambda^\top} \\
&= \left(I_{(m,m)} \left(I_m \otimes (-\Sigma^{-1}(E_{(m,1)}\lambda^\top + I_m \otimes \lambda))^\top (\Sigma^{-1} \otimes I_m) \right) \right) (I_{(m,m)} \otimes I_m) \\
&\times ((E_{(m,1)}\lambda^\top + I_m \otimes \lambda) \otimes I_m) \\
&+ (\Sigma^{-1} \otimes I_m) \left(E_{(m,1)}E_{(1,m)} + (I_m \otimes I_m) \right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \lambda \partial \lambda^\top} &= -\frac{n}{2} [2\Sigma^{-1} - (\Sigma^{-1} \circ I_m)] \star (E_{(m,1)} E_{(1,m)} + (I_m \otimes I_m)) + [E_{(m,1)} \lambda^\top + I_m \otimes \lambda] \star (I_m \otimes E_{(m,1)}) \\
&\quad \times \left(-2\Sigma^\top [E_{(m,1)} \lambda^\top + I_m \otimes \lambda]^\top (\Sigma^{-1} \otimes I_m) + \right. \\
&\quad \left. (\Sigma^{-1} \circ I_m) [E_{(m,1)} \lambda^\top + I_m \otimes \lambda]^\top ((\Sigma^{-1} \circ I_m) \times I_m) \right) (I_m \otimes I_m) \\
&\quad + \frac{1}{2} \sum_{i=1}^n -2 \left((\Sigma^{-1} \otimes I_1) (E_{(m,1)} \lambda^\top + I_m \otimes \lambda)^\top (\Sigma^{-1} \otimes I_m) \right) \beta z_i ((\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda} \beta z_i]) \otimes I_m) - 2\Sigma^{-1} \beta^2 z_i^2 \\
&\quad + I_m (I_m \otimes (-E_{(1,m)} \beta z_i)) (I_{(m,m)} \otimes I_m) \left((\Sigma^{-1} \otimes I_m) (E_{(m,1)} \lambda^\top + I_m \otimes \lambda) \Sigma^\top (\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda} \beta z_i]) \otimes I_m \right) \\
&\quad + ((\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda} \beta z_i])^\top \otimes I_m) \left(\left(\left(\left(I_{(m,m)} (I_m \otimes (-\Sigma^{-1} (E_{(m,1)} \lambda^\top + I_m \otimes \lambda)^\top (\Sigma^{-1} \otimes I_m)) \right) (I_{(m,m)} \otimes I_m) \right) \right. \right. \\
&\quad \times \left. \left. ((E_{(m,1)} \lambda^\top + I_m \otimes \lambda) \otimes I_m) \right) \right) (\Sigma^{-1} \otimes I_m) \\
&\quad + (\Sigma^{-1} \otimes I_m) (E_{(m,1)} E_{(1,m)} + (I_m \otimes I_m)) (\Sigma^{-1} \otimes I_m) \\
&\quad + (\Sigma^{-1} \otimes I_m) (E_{(m,1)} \lambda^\top + I_m \otimes \lambda) \left(-\Sigma^{-1} ((E_{(m,1)} \lambda^\top + I_m \otimes \lambda)^\top (\Sigma^{-1} \otimes I_m)) \right) ((\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda} \beta z_i]) \otimes I_m) \\
&\quad + (\Sigma^{-1} \otimes I_m) (E_{(m,1)} \lambda^\top + I_m \otimes \lambda) \Sigma^{-1} (-I_m \beta z_i).
\end{aligned}$$

Negative Expectation of Second Partial Derivative

From Weller we know that

$$ABC = B \star \{(\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i])(\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i])^\top \otimes I_{(m,m)}\},$$

so

$$\begin{aligned} E(ABC) &= \mathbf{B}^\top \star \{(\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i])(\mathbf{y}_i - [\boldsymbol{\mu} + \boldsymbol{\lambda}\beta z_i])^\top \otimes I_{(m,m)}\} \\ &= \mathbf{B}^\top \star \{\Sigma \otimes I_{(m,m)}\}. \end{aligned}$$

The negative expectation of the second partial derivative can be written as

$$\begin{aligned}
-E\left(\frac{\partial^2 \ell}{\partial \lambda \partial \lambda^\top}\right) &= \frac{n}{2} [2\Sigma^{-1} - (\Sigma^{-1} \circ I_m)] \star (E_{(m,1)} E_{(1,m)} + (I_M \otimes I_m)) \\
&+ [E_{(m,1)} \lambda^\top + I_m \otimes \lambda] \star (I_m \otimes E_{(m,1)}) \left(-2\Sigma^\top [E_{(m,1)} \lambda^\top + I_m \otimes \lambda]^\top (\Sigma^{-1} \otimes I_m) \right. \\
&+ (\Sigma^{-1} \circ I_m) [E_{(m,1)} \lambda^\top + I_m \otimes \lambda]^\top ((\Sigma^{-1} \circ I_m) \times I_m) \left. \right) (I_m \otimes I_m) \\
&+ \sum_{i=1}^n \Sigma^{-1} \beta^2 z_i^2 + \frac{n}{2} \left(\left(\left(\left(I_{(m,m)} (I_m \otimes (-\Sigma^{-1} (E_{(m,1)} \lambda^\top + I_m \otimes \lambda))^\top (\Sigma^{-1} \otimes I_m)) \right) \right) \right. \right. \\
&\times (I_{(m,m)} \otimes I_m) \left. \left. \right) \right) ((E_{(m,1)} \lambda^\top + I_m \otimes \lambda) \otimes I_m) \\
&+ (\Sigma^{-1} \otimes I_m) (E_{(m,1)} E_{(1,m)} + (I_m \otimes I_m)) \left. \right) (\Sigma^{-1} \otimes I_m) \\
&+ (\Sigma^{-1} \otimes I_m) (E_{(m,1)} \lambda^\top + I_m \otimes \lambda) \\
&\times \left(-\Sigma^{-1} ((E_{(m,1)} \lambda^\top + I_m \otimes \lambda))^\top (\Sigma^{-1} \otimes I_m) \right) \left. \right)^\top \star (\Sigma \otimes I_{(m,m)}).
\end{aligned}$$

Appendix B

Example Mplus Code for Ordered Categorical MIMIC Model

TITLE: MIMIC Model for Categorical Data;

DATA: FILE IS example.dat;

VARIABLE:

NAMES ARE item1 item2 item3 item4 item5 goldstandard;

USEVARIABLES ARE item1 item2 item3 item4 item5 goldstandard;

CATEGORICAL ARE item1 item2 item3 item4 item5;

ANALYSIS:

ESTIMATOR=WLSMV;

PARAMETERIZATION=DELTA;

MODEL:

*This specifies the measurement model;

FACTOR1 BY item1* item2 item3 item4 item5;

*This specifies the structural model;

FACTOR1 ON goldstandard;

*This specifies factor variance at 1;

FACTOR1@1;

Appendix C

Derivations for MZIPO Model

C.1 Score Function

The score function for τ can be written as

$$\begin{aligned}\frac{\partial \ell_\tau}{\partial \tau} &= \frac{\partial}{\partial \tau} \sum_{i=1}^n (1 - z_i) \left[-\ln(\sqrt{2\pi}) - \frac{1}{2}(b_i - \tau T_i)^2 \right] \\ &= -\frac{1}{2} \sum_{i=1}^n (1 - z_i) \frac{\partial}{\partial \tau} (b_i - \tau T_i)^2 \\ &= \sum_{i=1}^n (1 - z_i) T_i (b_i - \tau T_i),\end{aligned}$$

and the score function for p can be written as

$$\begin{aligned}\frac{\partial \ell_p}{\partial p} &= \frac{\partial}{\partial p} \sum_{i=1}^n z_i \ln p + \sum_{i=1}^n (1 - z_i) \ln(1 - p) \\ &= \sum_{i=1}^n \frac{z_i}{p} - \sum_{i=1}^n \frac{(1 - z_i)}{(1 - p)} \\ &= \sum_{i=1}^n \frac{z_i - p}{p(1 - p)}.\end{aligned}$$

For item j , the score function for β_j is

$$\begin{aligned} \frac{\partial \ell}{\partial \beta_j} = & \sum_{i=1}^n (1 - z_i) \left[I[y_{ij} = k] \left[\frac{-b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_k^{(j)} - \beta_j b_i] + b_i \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right] \right]}{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]} \right] \right. \\ & \left. + I[y_{ij} = K] \left[\frac{b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]}{1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]} \right] \right]. \end{aligned}$$

For categories $k = 0, \dots, K - 2$ and a item j , the score function for $\alpha_k^{(j)}$ is

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha_k^{(j)}} = & \sum_{i=1}^n (1 - z_i) \left[I[y_{ij} = k] \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]}{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]} \right] \right. \\ & \left. - I[y_{ij} = k + 1] \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]}{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]} \right] \right], \end{aligned}$$

and for category $k = K - 1$,

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha_{K-1}^{(j)}} = & \sum_{i=1}^n (1 - z_i) \left[I[y_{ij} = K - 1] \left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]}{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]} \right] \right. \\ & \left. - I[y_{ij} = K] \left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]}{1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]} \right] \right]. \end{aligned}$$

C.2 Information Matrix

Notation and Definitions

Important Derivatives

$$\frac{\partial}{\partial \alpha_k^{(j)}} \Psi[\alpha_k^{(j)} - \beta_j b_i] = \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]$$

$$\begin{aligned} \frac{\partial}{\partial \alpha_k^{(j)}} \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] &= \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]^2 \\ &\quad - \Psi[\alpha_k^{(j)} - \beta_j b_i]^2 [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] \end{aligned}$$

$$\frac{\partial}{\partial \beta_j} \Psi[\alpha_k^{(j)} - \beta_j b_i] = -b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]$$

$$\begin{aligned} \frac{\partial}{\partial \beta_j} \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] &= b_i \Psi[\alpha_k^{(j)} - \beta_j b_i]^2 [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] \\ &\quad - b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]^2 \end{aligned}$$

$$\begin{aligned}
& \frac{\partial}{\partial \alpha_k^{(j)}} \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]} \right] \\
&= \frac{\left[\frac{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]}{\partial \alpha_k^{(j)}} \frac{\partial \Psi[\alpha_k^{(j)} - \beta_j b_i]}{\partial \alpha_k^{(j)}} [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] \right]}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2} \\
&= \frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right] \frac{\partial \Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]}{\partial \alpha_k^{(j)}}}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2} \\
&= \frac{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2} \left[\Psi[\alpha_k^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2 - \Psi[\alpha_k^{(j)} - \beta_j b_i]^2 \left[1 - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right] \right] \\
&+ \frac{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right] \right]^2}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \alpha_k^{(j)} \partial \alpha_k^{(j)}} &= (1 - z_i) \left[I[y_{ij} = k] \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]}{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^2} \right. \right. \\
&- \left. \frac{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] \right]^2}{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^2} \right. \\
&- \left. I[y_{ij} = k + 1] \left[\frac{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2} \right. \right. \\
&+ \left. \left. \frac{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] \right]^2}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2} \right] \right].
\end{aligned}$$

For category $K - 1$,

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \alpha_{K-1}^2} &= (1 - z_i) \left[I[y_{ij} = K - 1] \frac{\partial}{\partial \alpha_{K-1}^{(j)}} \left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]} \right] \right. \\ &\quad \left. - I[y_{ij} = K] \frac{\partial}{\partial \alpha_{K-1}^{(j)}} \left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]} \right] \right], \end{aligned}$$

where

$$\begin{aligned} \frac{\partial}{\partial \alpha_{K-1}^{(j)}} &\left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]} \right] \\ &= \frac{\left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]}{\partial \alpha_{K-1}^{(j)}} \frac{\partial \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{\partial \alpha_{K-1}^{(j)}} \right]}{\left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]}{\partial \alpha_{K-1}^{(j)}} \right]^2} \\ &= \frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{\left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]}{\partial \alpha_{K-1}^{(j)}} \right]^2} \\ &= \frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] \left[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]^2 \right]}{\left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]}{\partial \alpha_{K-1}^{(j)}} \right]^2} \\ &= \frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2}{\left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]}{\partial \alpha_{K-1}^{(j)}} \right]^2} \end{aligned}$$

and

$$\begin{aligned}
& \frac{\partial}{\partial \alpha_{K-1}^{(j)}} \left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]} \right] \\
&= \frac{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] \frac{\partial \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]}{\partial \alpha_{K-1}^{(j)}} [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2} \\
&= \frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] \frac{\partial [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{\partial \alpha_{K-1}^{(j)}}}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2} \\
&= \frac{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] \left[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]^2 [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] \right]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2} \\
&+ \frac{[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \alpha_{K-1}^{(j)} \partial \alpha_{K-1}^{(j)}} &= (1 - z_i) \left[I[y_{ij} = K-1] \left[\frac{[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]] [\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]^2 [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]]^2} \right. \right. \\
&- \left. \frac{[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2}{[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]]^2} \right. \\
&- \left. I[y_{ij} = K] \left[\frac{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] [\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]^2 [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2} \right. \right. \\
&+ \left. \left. \frac{[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2} \right] \right].
\end{aligned}$$

For $k = 0$ to $K - 2$,

$$\begin{aligned}
& \frac{\partial^2 \ell}{\partial \beta_j \partial \alpha_k^{(j)}} \\
&= (1 - z_i) \left[I[y_{ij} = k] \frac{\partial}{\partial \beta_j} \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i][1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]} \right] \right. \\
&\quad \left. - I[y_{ij} = k + 1] \frac{\partial}{\partial \beta_j} \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i][1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]} \right] \right],
\end{aligned}$$

where

$$\begin{aligned}
& \frac{\partial}{\partial \beta_j} \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]} \right] \\
&= \frac{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right] \frac{\partial \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\partial \beta_j}}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2} \\
&\quad - \frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] \frac{\partial [\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\partial \beta_j}}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2} \\
&= \frac{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right] \frac{\partial \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\partial \beta_j}}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2} \\
&\quad - \frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] \frac{\partial [\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\partial \beta_j}}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \beta_j \partial \alpha_k^{(j)}} &= (1 - z_i) \left[I[y_{ij} = k] \frac{\left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]}{\left[b_i \Psi[\alpha_k^{(j)} - \beta_j b_i]^2 [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] - b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]^2 \right]}{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^2} \right] \\
&\quad - \frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] \left[-b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] + b_i \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]] \right]}{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^2} \\
&\quad - I[y_{ij} = k + 1] \frac{\left[\frac{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]}{\left[b_i \Psi[\alpha_{k+1}^{(j)} - \beta_j b_i]^2 [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] - b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]^2 \right]}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2} \right] \\
&\quad - \frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{k+1}^{(j)} - \beta_j b_i]] \left[-b_i \Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{k+1}^{(j)} - \beta_j b_i]] + b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] \right]}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2}
\end{aligned}$$

For $k = K - 1$,

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \beta_j \partial \alpha_{K-1}^{(j)}} &= (1 - z_i) \left[I[y_{ij} = K - 1] \frac{\partial}{\partial \beta_j} \left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]} \right] \right. \\
&\quad \left. - I[y_{ij} = K] \frac{\partial}{\partial \beta_j} \left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]} \right] \right],
\end{aligned}$$

where

$$\begin{aligned}
& \frac{\partial}{\partial \beta_j} \left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi(\alpha_{K-1}^{(j)})]}{\Psi(\alpha_{K-1}^{(j)} - \beta_j b_i) - \Psi(\alpha_{K-2}^{(j)} - \beta_j b_i)} \right] \\
&= \frac{\left[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i] \right] \frac{\partial \Psi(\alpha_{K-1}^{(j)})}{\partial \beta_j} [1 - \Psi(\alpha_{K-1}^{(j)})]}{\left[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i] \right]^2} \\
& \quad - \frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi(\alpha_{K-1}^{(j)})] - \beta_j b_i \frac{\partial \Psi(\alpha_{K-1}^{(j)})}{\partial \beta_j} [1 - \Psi(\alpha_{K-1}^{(j)})]}{\left[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i] \right]^2} \\
&= \frac{\left[\Psi(\alpha_{K-1}^{(j)} - \beta_j b_i) - \Psi(\alpha_{K-2}^{(j)} - \beta_j b_i) \right] \left[b_i \Psi(\alpha_{K-1}^{(j)} - \beta_j b_i)^2 [1 - \Psi(\alpha_{K-1}^{(j)}) - \beta_j b_i] - b_i \Psi(\alpha_{K-1}^{(j)} - \beta_j b_i) [1 - \Psi(\alpha_{K-1}^{(j)})]^2 \right]}{\left[\Psi(\alpha_{K-1}^{(j)} - \beta_j b_i) - \Psi(\alpha_{K-2}^{(j)} - \beta_j b_i) \right]^2} \\
& \quad - \frac{\Psi(\alpha_{K-1}^{(j)} - \beta_j b_i) [1 - \Psi(\alpha_{K-1}^{(j)})] - \beta_j b_i \left[-b_i \Psi(\alpha_{K-1}^{(j)} - \beta_j b_i) [1 - \Psi(\alpha_{K-1}^{(j)})] + b_i \Psi(\alpha_{K-2}^{(j)} - \beta_j b_i) [1 - \Psi(\alpha_{K-2}^{(j)}) - \beta_j b_i] \right]}{\left[\Psi(\alpha_{K-1}^{(j)} - \beta_j b_i) - \Psi(\alpha_{K-2}^{(j)} - \beta_j b_i) \right]^2}
\end{aligned}$$

and

$$\begin{aligned}
& \frac{\partial}{\partial \beta_j} \left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]} \right] \\
&= \frac{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] \frac{\partial \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]}{\partial \beta_j} [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2} \\
&\quad - \frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] \frac{\partial [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{\partial \beta_j}}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2} \\
&= \frac{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] \left[b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]^2 [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] - b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2 \right]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2} \\
&\quad - \frac{b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \frac{\partial^2 \ell}{\partial \beta_j \partial \alpha_{K-1}^{(j)}} = (1-z_i) \left[I[y_{ij}=K-1] \frac{\left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]}{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]} \right]^2 \left[b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]^2 \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right] - b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]^2 \right]}{\left[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i] \right]^2} \right] \\
& - \frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right] \left[-b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i] \right] + b_i \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i] \right] \right]}{\left[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i] \right]^2} \\
& - I[y_{ij} = K] \frac{\left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]}{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]} \right]^2 \left[b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]^2 \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right] - b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]^2 \right]}{\left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]^2} \\
& - \frac{b_i \left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]}{\left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]^2} \right]^2}{\left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]^2}.
\end{aligned}$$

121

For each item, the second partial derivative for threshold parameters one category apart from one another can be written as

$$\frac{\partial^2 \ell}{\partial \alpha_k^{(j)} \partial \alpha_{k+1}^{(j)}} = (1-z_i) \left[I[y_{ij} = k] \frac{\partial}{\partial \alpha_{k+1}^{(j)}} \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]}{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]} \right] - I[y_{ij} = k+1] \frac{\partial}{\partial \alpha_{k+1}^{(j)}} \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]}{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]} \right] \right],$$

where

$$\frac{\partial}{\partial \alpha_{k+1}^{(j)}} \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]} \right] = 0,$$

and

$$\begin{aligned} & \frac{\partial}{\partial \alpha_{k+1}^{(j)}} \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]} \right] \\ &= \frac{\left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]} \right]}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2} \frac{\partial \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\partial \alpha_{k+1}^{(j)}} \\ &= \frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2} \frac{\partial [\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i]]}{\partial \alpha_{k+1}^{(j)}} \\ &= \frac{-\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] \Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{k+1}^{(j)} - \beta_j b_i]]}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2}. \end{aligned}$$

Therefore,

$$\frac{\partial^2 \ell}{\partial \alpha_k^{(j)} \partial \alpha_{k+1}^{(j)}} = (1 - z_i) \left[I[y_{ij} = k + 1] \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] \Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{k+1}^{(j)} - \beta_j b_i]]}{\left[\Psi[\alpha_{k+1}^{(j)} - \beta_j b_i] - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2} \right] \right].$$

Similarly

$$\frac{\partial^2 \ell}{\partial \alpha_{K-1}^{(j)} \partial \alpha_{K-2}^{(j)}} = (1 - z_i) \left[I[y_{ij} = K-1] \left[\frac{\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i][1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i][1 - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i]]}{\left[\Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] - \Psi[\alpha_{K-2}^{(j)} - \beta_j b_i] \right]^2} \right] \right].$$

The second partial derivative with respect to beta can be written as

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \beta_j \beta_j} &= (1 - z_i) \left[\sum_{k=0}^{K-1} I[y_{ij} = k] \frac{\partial}{\partial \beta_j} \left[\frac{-b_i \Psi[\alpha_k^{(j)} - \beta_j b_i][1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]] + b_i \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i][1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]]}{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]} \right] \right] \\ &\quad + I[y_{ij} = K] \frac{\partial}{\partial \beta_j} \left[\frac{b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i][1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]} \right], \end{aligned}$$

where

$$\begin{aligned} \frac{\partial}{\partial \beta_j} \left[\frac{-b_i \Psi[\alpha_k^{(j)} - \beta_j b_i][1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] + b_i \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i][1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]]}{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]} \right] &= \\ \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]}{\partial - b_i \Psi[\alpha_k^{(j)} - \beta_j b_i][1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] + b_i \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i][1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]]} \right] \frac{\partial \Psi[\alpha_k^{(j)} - \beta_j b_i]}{\partial \beta_j} & \\ \left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]}{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^2} \right] & \\ \left[\frac{-b_i \Psi[\alpha_k^{(j)} - \beta_j b_i][1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] + b_i \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i][1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]]}{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^2} \right] \frac{\partial \Psi[\alpha_k^{(j)} - \beta_j b_i]}{\partial \beta_j} & \\ \left[\frac{-b_i \Psi[\alpha_k^{(j)} - \beta_j b_i][1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] + b_i \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i][1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]]}{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^2} \right] & \end{aligned}$$

$$\begin{aligned}
& \frac{\partial \left[-b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] + b_i \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]] \right]}{\partial \beta_j} \\
&= -b_i \left[b_i \Psi[\alpha_k^{(j)} - \beta_j b_i]^2 [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] - b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]]^2 \right] \\
&+ b_i \left[b_i \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]^2 [1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]] - b_i \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]]^2 \right],
\end{aligned}$$

and

$$\frac{\partial \Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]}{\partial \beta_j} = -b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_k^{(j)} - \beta_j b_i]] + b_i \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]]$$

Next,

$$\begin{aligned}
& \left[\frac{\frac{\partial}{\partial \beta_j} \left[\frac{b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]} \right]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]} \right] \\
&= \frac{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] \frac{\partial b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{\partial \beta_j}}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2} \\
&= \frac{\left[\frac{b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]} \right] \frac{\partial [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{\partial \beta_j}}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2} \\
&= \frac{\left[\frac{b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]} \right] \left[\frac{b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]^2 [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]] - b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] [1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2} \right]}{[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i]]^2},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_j} &= (1-z_i) \left[\frac{\sum_{k=0}^{K-1} I[y_{ij}=k]}{\left[\frac{\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]}{b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i]} \right]^2} \right. \\
&\quad \left. - b_i \frac{\left[b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] \right]^2 \left[1 - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right] - b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^2}{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^2} \right] \\
&\quad + \frac{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^2}{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^2} \\
&\quad - \frac{\left[-b_i \Psi[\alpha_k^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_k^{(j)} - \beta_j b_i] \right] + b_i \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right] \right]^2}{\left[\Psi[\alpha_k^{(j)} - \beta_j b_i] - \Psi[\alpha_{k-1}^{(j)} - \beta_j b_i] \right]^2} \\
&\quad + I[y_{ij} = K] \left[\frac{b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]^2 \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right] - b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]^2}{\left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]^2} \right] \\
&\quad - \frac{\left[b_i \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right] \right]^2}{\left[1 - \Psi[\alpha_{K-1}^{(j)} - \beta_j b_i] \right]^2} \cdot
\end{aligned}$$

Bibliography

- Abramowitz, M. and Stegun, I. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York: Dover Publications.
- Ahmed, R., Prizment, A., Lazovich, D., Schmitz, K., and Folsom, A. (2008). Lymphedema and quality of life in breast cancer survivors: The Iowa Women's Health Study. *Journal of Clinical Oncology* **26**, 5689–5696.
- Archer, C. and Jennrich, R. (1973). Standard errors for rotated factor loadings. *Psychometrika* **38**, 581–592.
- Arminger, G., Stein, P., and Wittenberg, J. (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika* **64**, 475–494.
- Campbell, D. and Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* **56**, 81–105.
- Clark, L. and Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment* **7**, 309–319.

- Comrey, A. and Lee, H. (1992). *A first course in factor analysis*. Lawrence Erlbaum.
- Cormier, J., Xing, Y., Zaniletti, I., Askew, R., Stewart, B., and Armer, J. (2009). Minimal limb volume change has a significant impact on breast cancer survivors. *Lymphology* **42**, 161–175.
- Cudeck, R. and O’Dell, L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin* **115**, 475–487.
- Dolan, C. and van der Maas, H. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika* **63**, 227–253.
- Freeman, E., DeRubeis, R., and Rickels, K. (1996). Reliability and validity of a daily diary for premenstrual syndrome. *Psychiatry Research* **65**, 97–106.
- Freeman, E., Rickels, K., Sondheimer, S., and Polansky, M. (1999). Differential response to antidepressants in women with premenstrual syndrome/premenstrual dysphoric disorder: a randomized controlled trial. *Archives of General Psychiatry* **56**, 932–939.
- Freeman, E., Rickels, K., Sondheimer, S., Polansky, M., and Xiao, S. (2004). Continuous or intermittent dosing with sertraline for patients with severe premenstrual syndrome or premenstrual dysphoric disorder. *American Journal of Psychiatry* **161**, 343–351.

- Freeman, E., Rickels, K., Yonkers, K., Kunz, N., McPherson, M., and Upton, G. (2001). Venlafaxine in the treatment of premenstrual dysphoric disorder. *Obstetrics & Gynecology* **98**, 737–744.
- Guilford, J. (1954). *Psychometric methods*. McGraw-Hill New York.
- Guy, W. (1976). Early Clinical Drug Evaluation UNIT (ECDEU) Assessment Manual for Psychopharmacology. *National Institute of Mental Health publication DHEW* pages 218–222.
- Hauser, R. and Goldberger, A. (1971). The treatment of unobservable variables in path analysis. *Sociological Methodology* **3**, 81–117.
- Hayashi, K. and Kumar Sen, P. (1998). On covariance estimators of factor loadings in factor analysis. *Journal of Multivariate Analysis* **66**, 38–45.
- Hayashi, K. and Yung, Y. (1999). Standard errors for the class of orthomax-rotated factor loadings: Some matrix results. *Psychometrika* **64**, 451–460.
- Jennrich, R. (1974). Simplified formulae for standard errors in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology* **27**, 122–131.
- Jennrich, R. and Thayer, D. (1973). A note on Lawley's formulas for standard errors in maximum likelihood factor analysis. *Psychometrika* **38**, 571–580.

- Jóreskog, K. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika* **59**, 381–389.
- Joreskog, K. and Goldberger, A. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association* **70**, 631–639.
- Jóreskog, K. and Moustaki, I. (2000). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research* **36**, 347–387.
- Kelley, M. and Anderson, S. (2008). Zero inflation in ordinal data: Incorporating susceptibility to response through the use of a mixture model. *Statistics in Medicine* **27**, 3674–3688.
- Lawley, D. and Maxwell, A. (1962). Factor analysis as a statistical method. *The Statistician* **12**, 209–229.
- Little, R. and Rubin, D. (1987). *Statistical analysis with missing data*. Wiley New York.
- Liu, Q. and Pierce, D. (1994). A note on gauss-hermite quadrature. *Biometrika* **81**, 624–629.
- Lubke, G. and Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods* **10**, 21–39.

- Lubke, G. and Muthén, B. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling: A Multidisciplinary Journal* **14**, 26–47.
- Lubke, G. and Neale, M. (2006). Distinguishing Between Latent Classes and Continuous Factors: Resolution by Maximum Likelihood? *Multivariate Behavioral Research* **41**, 499–532.
- Lubke, G. and Neale, M. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: class invariance of parameters of factor mixture models. *Multivariate Behavioral Research* **43**, 592–620.
- MacRae, E. (1974). Matrix derivatives with an application to an adaptive linear decision problem. *The Annals of Statistics* **2**, 337–346.
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)* **42**, 109–142.
- Moustaki, I. (2000). A latent variable model for ordinal variables. *Applied Psychological Measurement* **24**, 211–223.
- Muthén, B. (1984a). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* **49**, 115–132.
- Muthén, B. (1984b). A general structural equation model with dichotomous, or-

- dered categorical, and continuous latent variable indicators. *Psychometrika* **49**, 115–132.
- Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. *Advances in latent variable mixture models* **1**, 1–24.
- Muthén, B. and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* **55**, 463–469.
- Norman, S., Miller, L., Erikson, H., Norman, M., and McCorkle, R. (2001). Development and validation of a telephone questionnaire to characterize lymphedema in women treated for breast cancer. *Physical Therapy* **81**, 1192–1205.
- Nunnally, J., Bernstein, I., and Berge, J. (1967). *Psychometric theory*. New York: McGraw-Hill.
- Ogasawara, H. (1996). Standard errors for rotated factor loadings by normalized orthomax method. *Japanese Journal of Behaviormetrics* **23**, 122–129.
- Ogasawara, H. (1998). Standard errors for rotation matrices with an application to the promax solution. *British journal of mathematical & statistical psychology* **51**, 163–178.
- Ogasawara, H. (1999). Standard errors for procrustes solutions. *Japanese Psychological Research* **41**, 121–130.

- Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Oxford: Nielsen & Lydiche.
- Rogers, G. (1980). *Matrix derivatives*. New York-Basel: Dekker.
- Samejima, F. (1970). Estimation of latent ability using a response pattern of graded scores. *Psychometrika* **35**, 139–139.
- Sammel, M. and Ryan, L. (1996). Latent variable models with fixed effects. *Biometrics* **52**, 650–663.
- Sammel, M. and Ryan, L. (2002). Effects of covariance misspecification in a latent variable model for multiple outcomes. *Statistica Sinica* **12**, 1207–1222.
- Sammel, M., Ryan, L., and Legler, J. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**, 667–678.
- Schmitz, K., Ahmed, R., Troxel, A., Cheville, A., Lewis-Grant, L., Smith, R., Bryan, C., Williams-Smith, C., and Chittams, J. (2010). Weight lifting for women at risk for breast cancer-related lymphedema. *The Journal of the American Medical Association* **304**, 2699–2705.
- Schmitz, K., Ahmed, R., Troxel, A., Cheville, A., Smith, R., Lewis-Grant, L., Bryan, C., Williams-Smith, C., and Greene, Q. (2009). Weight lifting in women

- with breast-cancer-related lymphedema. *New England Journal of Medicine* **361**, 664–673.
- Schmitz, K., Courneya, K., Matthews, C., Demark-Wahnefried, W., Galvão, D., Pinto, B., Irwin, M., Wolin, K., Segal, R., Lucia, A., et al. (2010). American College of Sports Medicine roundtable on exercise guidelines for cancer survivors. *Medicine & Science in Sports & Exercise* **42**, 1409–1426.
- Schmitz, K., Troxel, A., Cheville, A., Grant, L., Bryan, C., Gross, C., Lytle, L., and Ahmed, R. (2009). Physical activity and lymphedema (the PAL trial): Assessing the safety of progressive strength training in breast cancer survivors. *Contemporary clinical trials* **30**, 233–245.
- Shih, Y., Xu, Y., Cormier, J., Giordano, S., Ridner, S., Buchholz, T., Perkins, G., and Elting, L. (2009). Incidence, treatment costs, and complications of lymphedema after breast cancer among women of working age: A 2-year follow-up study. *Journal of Clinical Oncology* **27**, 2007–2014.
- Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology* **15**, 201–293.
- Streiner, D. and Norman, G. (1994). *Health measurement scales*. Oxford: Oxford University Press.
- Styan, G. (1973). Hadamard products and multivariate statistical analysis. *Linear Algebra and its Applications* **6**, 217–240.

- Tabachnick, B., Fidell, L., and Osterlind, S. (2001). Using multivariate statistics.
- Thurstone, L. (1938). Primary mental abilities. *Psychometric monographs* .
- Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics* **59**, 254–262.
- Werts, C. and Linn, R. (1970). Path analysis: Psychological examples. *Psychological Bulletin* **74**, 193–212.
- Yuan, K., Marshall, L., and Bentler, P. (2002). A unified approach to exploratory factor analysis with missing data, nonnormal data, and in the presence of outliers. *Psychometrika* **67**, 95–121.
- Yung, Y. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika* **62**, 297–330.
- Zellner, A. (1970). Estimation of regression relationships containing unobservable independent variables. *International Economic Review* **11**, 441–454.
- Zhang, D. and Lin, X. (2008). Variance component testing in generalized linear mixed models for longitudinal/clustered data and other related topics. *Random effect and latent variable model selection* pages 19–36.