University of Alberta

Statistical Issues With Item Response Data Analysis

by

Xiaoming Sheng ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**

in

Statistics

Department of Mathematical and Statistical Sciences

Edmonton, Alberta

Fall 2002

National Library of Canada
Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services
Acquisitions et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

0-612-81265-0

Canada

University of Alberta

Library Release Form

Name of Author: Xiaoming Sheng

Title of Thesis: Statistical Issues With Item Response Data Analysis

Degree: Doctor of Philosophy

Year this Degree Granted: 2002

Xiaoming Sheng

Department of Mathematical
and Statistical Sciences

University of Alberta

Edmonton, Alberta

Canada T6G 2G1

Date: July 11, 2002

In God we trust. All others must bring data.
– Robert Hayden

Statistics are like a bikini.
What is revealed is interesting,
what is concealed is crucial.
– Aaron Levenstein

The purpose of models is not to fit the data but to sharpen the questions.
– Samuel Karlin

UNIVERSITY OF ALBERTA

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Statistical Issues With Item Response Data Analysis** submitted by **Xiaoming Sheng** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in Statistics.

Dr. Narasimha Prasad (Chair)

K. C. Carriere

Dr. K. C. Carriere (Supervisor)

SRLele

Dr. Subhash Lele

Dr. Jeffrey Johnson (Public Health Sciences)

Dr. Chairmaine Dean (Simon Fraser University)

April 18, 2002

To my parents, my wife, and my daughter.

Abstract

This thesis concerns with the modelling of ordinal categorical data. Ordinal response data are commonly observed in health and medical investigations that include several items. The qualitative nature of such data, which are ordinal and not equally spaced, has generated much concern when a quantitative analytic tool is employed to analyze them. The Rasch model approach and its generalizations have been given much attention and are widely used in modelling item response data.

We first developed an improved conditional maximum likelihood estimation procedure for the Rasch model with item response data. Based on the conditional maximum likelihood method, we implemented a simultaneous estimation method that can estimate the Rasch parameters more efficiently. We then obtained the asymptotic properties of these estimators and developed the conditional likelihood ratio test for the goodness-of-fit of the model. The improved performance of our estimators was demonstrated by comparing it to that of the current conditional method known as the CON procedure, where it outperformed CON in both model fit and the precision of the Rasch estimators.

Second, we developed a method to account for the inter-item correlations

in the Rasch Models. The polychoric correlation coefficient uses the concept of latent variables. With the assumption of a common polychoric correlation coefficient among items, we advocated the theory of generalized estimation equations approach to obtain consistent estimators of the parameters in the Rasch model.

Then, we discussed techniques used for missing data analysis. In item response data, we often encounter missing observations due to various reasons. We are particularly interested in the implications of various missing data strategies and those of various missing data mechanisms in Rasch models. The bootstrap under hot deck imputation was suggested to be the best approach in producing consistent estimators and variances in various situations.

The utility of our methods developed in this thesis was demonstrated by using the data on a study of families of lung cancer patients. We considered fitting the Rasch model to each family health and care variable to obtain Rasch scores of person ability parameters adjusted by item difficulty for each subject. Comparisons are then made to show the advantage of using our methods over other approaches such as the average scoring method, and the traditional Rasch method.

Acknowledgements

First of all, I would like to thank Dr. K. C. Carrière for her supervision and financial support during my research and study at the University of Alberta. I would like to express my sincere thanks for her excellent guidance and generous advice, constant encouragement, very careful reading and critical discussions, great help and kind patience at all stages in the preparation of the thesis. Without these, the completion of this thesis would never have been possible.

Thanks to Dr. Atanu Biswas for his valuable discussions and suggestions on contents in *Chapter 4* of the thesis.

I thank Dr. Charmaine Dean, Dr. Jeff Johnson, Dr. Subhash Lele, and Dr. N. G. N. Prasad, the members of my supervisory and examining committee, for carefully reading my thesis with many valuable suggestions and comments. I also thank the Statistics Center, and the Department of Mathematical and Statistical Sciences, for granting me teaching assistantship during my study. Thanks also go to Dr. Rhonda Rosychuk, for providing financial support so that I have time to focus on writing my thesis. The faculty members, my colleagues and friends deserve my thanks for their kind help. Special thanks go to the computing staff and the administrative staff for their warm help and

cooperation.

Finally, but most importantly, my greatest appreciation is presented to my parents, my wife and my daughter. Their constant love, support, and encouragement always accompany me, anytime and anywhere, till forever.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background of the Thesis

This thesis deals with the modelling of ordinal categorical data. Ordinal response data are commonly observed in health and medical investigations that include several items. They include binary data ("yes" or "no"), Poisson counts (number of successes), rating scales ("disagree", "neutral", and "agree" etc.), partial credit (awarding partial credit on intermediate levels of performance), and other ordinal categorical data. For example, a symptom distress scale (SDS) has been used to obtain information about the health of lung cancer patients (Kristjanson et. al., 1997). This scale includes 13 relevant symptom distress questions and uses 5-point scale to obtain qualitative data (1=normal, 2=occasional, 3=frequent, 4=usual, or 5=constant distress). The qualitative nature of such data, which are ordinal and not equally spaced, has generated much concern when a quantitative analytic tool is employed to analyze them.

The primary goal in the modelling of item response data is to find a unique measurement of the person's abilities and of the item difficulties that satisfies the properties of the fundamental measurement, which include sufficiency, divisibility, unidimensionality, linearity, abstraction, invariance, sample-free calibration, test-free measurement, and conjoint transitivity (Wright, 1997).

The common practice has been to simply obtain an average score for each person or item, and analyze it using standard analytic tools. However, the item or person scores obtained from Likert scale data often do not fulfill these properties of fundamental measurement. This fact casts doubt on the validity of the findings and conclusions based on the assumption that the simple average scores are accurate reflections of the person.

One analytic method in item response theory that does not require much of these unverifiable assumptions is the Rasch measurement, which is a way to convert ordinal observations into linear measures while satisfying the properties mentioned above. Rasch (1960) proposed the analytic method, known as the Rasch model, which avoids the limitations of averaging scores and satisfies the requirement for basic measurement properties. The Rasch model approach and its generalizations have been given much attention and are widely used in modelling item response data.

In this thesis, we will discuss the estimation and inference issues that arise from the Rasch analysis. First, we consider an improved estimation method by removing the independent assumption on the model parameters in the current approach. Then we discuss a method for the Rasch model to account for the inter-item correlations. We will also consider missing data analysis techniques under various missing data mechanisms. The lung cancer study on family satisfaction is used to illustrate the methods proposed in the thesis, and to compare with other methods currently being used.

1.2 Thesis Overview

This thesis is organized in the following manner:

Chapter 2 gives a general review of the literature on the Rasch model. We give an overview of the Rasch Model to review its history and current methodologies and to discuss current estimation and inference tools on the model. Some issues on missing data analysis are also reviewed.

Chapter 3 develops an improved conditional maximum likelihood estimation procedure for the Rasch model with item response data (Sheng and Carrière, 2002). Based on the conditional maximum likelihood method, we implement a simultaneous estimation method that can estimate the Rasch parameters more efficiently. We then obtain the asymptotic properties of these estimators and develop the conditional likelihood ratio test for the goodness-of-fit of the model. The improved performance of our estimators is demonstrated by comparing it to that of the current conditional method known as the CON procedure (Wright and Masters, 1982). We conclude that our estimation method outperforms CON in both model fit and the precision of the Rasch estimators.

Chapter 4 recognizes the inter-item correlation, known as the polychoric correlation, that may not be negligible and develop a method to account for them in the Rasch Models. The polychoric correlation coefficient uses the concept of latent variables, which are usually continuous although unobservable, giving rise to the apparent complexity of the data. With the assumption of a common polychoric correlation coefficient among items, we advocated the theory of generalized estimation equations approach to obtain consistent estimators of the parameters in the Rasch model.

Chapter 5 discusses techniques used for missing data analysis in Rasch model. In item response data, we often encounter missing observations due to various reasons. We are particularly interested in the implications of various missing data strategies (a method deleting missing observations, a simple hot deck imputation, and the bootstrap technique) and those of various missing data mechanisms (missing completely at random, data missing at random, and nonignorable missing data). The bootstrap under hot deck imputation was suggested to be the best approach in producing consistent estimators and variances in various situations (Shao and Sitter, 1996). Large sample behaviors of the estimators are examined in a simulation study to evaluate the comparative performances under various settings.

Chapter 6 demonstrates the utility of our methods proposed in this thesis using the data on a study of families of lung cancer patients. The data set was collected on an ordinal scale for relevant family health and care questions, and presents many analytic challenges stated above, including missing data problems. We consider fitting the Rasch model to each family health and care variable to obtain Rasch scores in a linear scale, i.e. person ability parameters adjusted by item difficulty, for each subject. Comparisons are then made to show the advantage of using our proposed methods over conventional approaches such as average scoring method, or the traditional Rasch method.

Finally, *Chapter 7* summaries the main contributions of the thesis, and suggests possible future research to further improve on the methodologies suggested in this thesis.

Chapter 2

Review of Literature

2.1 The Data Structure

Suppose there are I multiple-choice questions given to N subjects. For question i, subject j has to choose X_{ij} from one out of several possible choices, divided into $m+1$ mutually exclusive categories. We denote these categories to have integer values so that X_{ij} takes a data point $0, 1, \ldots, m$. Denote the person response vectors of length I

$$\mathbf{x_j} = (X_{1j}, X_{2j}, \ldots, X_{Ij})', \qquad (j = 1, \ldots, N),$$

and the item response vectors of length N

$$\mathbf{x_i} = (X_{i1}, X_{i2}, \ldots, X_{iN})', \qquad (i = 1, \ldots, I).$$

The data matrix $\mathbf{X}$ is obtained as in *Table 2.1.*

The scores for subject j and item i are defined as

$$0 \leq r_j = X_{.j} \leq Im, \qquad 0 \leq s_i = X_{i.} \leq Nm,$$

where $X_{.j} = \sum_i^I X_{ij}$ and $X_{i.} = \sum_j^N X_{ij}$.

2.2 Overview of Rasch Models

2.2.1 The polytomous Rasch model

The general term we deal with here is the *fundamental measurement*, which includes the following properties: sufficiency, divisibility, unidimensionality, linearity, abstraction, invariance, sample-free calibration, test-free measurement, and conjoint transitivity (Wright, 1997).

The usual method of analysis has been to simply take the average among the items to obtain a score for each person, or average among the persons to obtain a score for each item. The item or person scores obtained from Likert scale data often do not fulfill these properties of fundamental measurement. This fact casts doubt on the validity of the findings and conclusions based on assumptions that the scores are accurate reflections of the study subject.

Solutions to these problems did not emerge until George Rasch implemented a measurement model known as the Rasch model to convert the Likert scale data to a continuous measurement data. Originally Rasch (1960) developed the *multidimensional dichotomous Rasch model* which is designed to measure separate latent traits for each response category k, $k = 0, \ldots, m$. The function specifies the multiplicative definition of fundamental measurement for each dichotomous observations as:

$$P_{ij} = \frac{\exp(\beta_j - \delta_i)}{1 + \exp(\beta_j - \delta_i)} \tag{2.1}$$

where P_{ij} is the probability of a correct solution for subject j on item i, to be determined by a measure of person ability β_j and a calibration of item difficulty δ_i.

Over the last four decades, several generalizations have been developed. These models commonly use the *threshold approach* (Andrich, 1978d) in that the dichotomous Rasch model is assumed to hold with the probability of passing the thresholds between two neighboring response categories $k - 1$ and k, given as:

$$P(X_{ij} = k | X_{ij} \in \{k-1, k\}) = \frac{\exp(\beta_j - \delta_{ik})}{1 + \exp(\beta_j - \delta_{ik})} \tag{2.2}$$

where β_j is the person ability of person j, δ_{ik} is the item difficulty of step $k+1$ in item i.

The models differ in the way they treat the threshold parameters δ_{ik}. All models decompose them into linear components, but the basic difference lies in the assumption about threshold distances on a continuum.

Among these threshold approaches, the least restrictive one is the *partial credit model* (Masters, 1982):

$$\pi_{ijk} = P(X_{ij} = k) = \frac{\exp \sum_{k'=0}^{k} (\beta_j - \delta_{ik'})}{\sum_{k''=0}^{m} \exp \sum_{k'=0}^{k''} (\beta_j - \delta_{ik'})} = \frac{\exp(k\beta_j - \sigma_{ik})}{\sum_{k''=0}^{m} \exp(k''\beta_j - \sigma_{ik''})} \tag{2.3}$$

where σ_{ik} is a cumulative threshold parameter defined as $\sigma_{ik} = \sum_{k'=0}^{k} \delta_{ik'}$. Because of its generality, this model is referred to as the **polytomous Rasch model.**

2.2.2 Special Cases and Extensions of the Rasch Model

There are several generalizations of the Rasch Model proposed and adopted by various investigators since Rasch proposed it in 1960.

1. Dichotomous Rasch model (Rasch, 1960).

   Under this model, the response is binary, for example, "true" or "false", "fail" or "pass", "dead" or "alive". This is simply a special case of the polytomous Rasch Model. In these examples, the response variable only takes value 0 or 1:

$$P(X_{ij} = 1) = \frac{\exp(\beta_j - \delta_i)}{1 + \exp(\beta_j - \delta_i)}$$

2. Polytomous Rasch model.

   This is the general Rasch model with polytomous responses, any nominal or ordinal unidimensional responses from 0 to m. Unidimensionality means that the measurement of any object or entity describes only one attribute of the object measured. This is a universal characteristic of all measurement (Thurstone, 1931). The model is presented as follows:

$$P(X_{ij} = k) = \frac{\exp(k\beta_j - \sigma_{ik})}{\sum_{k''=0}^{m} \exp(k''\beta_j - \sigma_{ik''})}$$

   There are various special models proposed, depending upon the data types.

   (a) Partial Credit model (PCM, Masters, 1982).

This is a simple extension of dichotomous response to more intermediate levels of responses. In consequence, partial credit is given to get these intermediate levels. For example, students are given partial credits when they choose a reasonable but not perfect answer. The model is

$$P(X_{ij} = k) = \frac{\exp(k\beta_j - \sum_{k'=0}^{k} \delta_{ik'})}{\sum_{k''=0}^{m} \exp(k''\beta_j - \sum_{k'=0}^{k''} \delta_{ik'})}$$

(b) Rating Scale model (Andrich, 1978c, 1978d, 1979).

This is the most commonly seen in survey questionnaire data. The study participant is asked to choose the attitude level among a fixed set of rating points, ranging from, for example, "strongly disagree", "disagree", "neutral", "agree", and "strongly agree". The 5-point scale data on symptom distress questions studying the health of lung cancer patients fits into this category of data type. Because the same set of rating points is used with every item, it is usually assumed that the relative difficulties of the steps in each item should not vary from item to item. This then can be incorporated into the Partial Credit model by resolving each item step into two components so that $\delta_{ik} = \delta_i + \tau_k$, and therefore

$$P(X_{ij} = k) = \frac{\exp(k\beta_j - k\delta_i - \sum_{k'=0}^{k} \tau_{k'})}{\sum_{k''=0}^{m} \exp(k''\beta_j - k''\delta_i - \sum_{k'=0}^{k''} \tau_{k'})}$$

(c) Binomial Trials (Rasch, 1972, Andrich, 1978a, 1978b).

This model is appropriate when the outcome is the number of successes in an independent trial. Such data are special cases of rating scale data with relative difficulties between two neighboring categories being fixed for all items $\delta_{ik} = \delta_i + \log \frac{k}{m-k+1}$. The model becomes

$$P(X_{ij} = k) = \frac{\exp(k\beta_j - k\delta_i)/\frac{k!(m-k)!}{m!}}{\sum_{k''=0}^{m} \exp\left((k''\beta_j - k''\delta_i)/\frac{k''!(m-k'')!}{m!}\right)}$$

(d) Poisson Counts (Rasch, 1960).

In some testing situations, there is no clear upper limit on number of trials that might be observed and counted. For example, the number of successes in a fixed period of time is the data of this type. Here $\delta_{ik} = \delta_i + \log k$, and

$$P(X_{ij} = k) = \frac{\exp(k\beta_j - k\delta_i)/k!}{\sum_{k''=0}^{m} \exp((k''\beta_j - k''\delta_i)/k''!)}$$

3. Two- and Three- Parameter Logistic Models.

The above polytomous Rasch models are known as the one-parameter logistic (1-PL) model. It contains only one parameter δ_{ik} per category per item. Birnbaum (1968) extended it to two- and three-parameter logistic models.

Birnbaum's main contribution was his proposal of the two-parameter logistic (2-PL) model or *Birnbaum Model*:

$$p(X_{ij} = k | X_{ij} \in \{k-1, k\}) = \frac{\exp\{a_i(\beta_j - b_{ik})\}}{1 + \exp\{a_i(\beta_j - b_{ik})\}} \tag{2.4}$$

where b_{ik} is the item difficulty parameter, the point on the person ability scale where a person has a probability of success 0.50 on step k in item i; whereas the value of a_i is proportional to the slope of the tangent to the response function at this point. If b_{ik} increases, the response function moves to the right, and a higher ability is needed to produce the same probability of success on this step of the item. Also, the larger the value of a_i, the better the item discriminates between the probabilities of success of persons with abilities below and above $\beta_j = b_{ik}$. For this reason, a_i is called the *discrimination* parameter (Birnbaum, 1968).

Birnbaum (1968) also proposed a third parameter for inclusion in the model to account for the non-zero performance of persons with low-ability on multiple choices. This non-zero performance is due to the probability of guessing correct answers to multiple-choice items. The model takes the form

$$p(X_{ij} = k | X_{ij} \in \{k-1, k\}) = c_i + (1 - c_i)\frac{\exp\{a_i(\beta_j - b_{ik})\}}{1 + \exp\{a_i(\beta_j - b_{ik})\}} \quad (2.5)$$

Equation (2.5) follows immediately from the assumption that the person either knows the correct response with a probability described by *Equation (2.4)* or guesses with a probability of success equal to the value of c_i. It is clear that the parameter c_i is the height of the lower asymptote of the response function. Although *Equation (2.5)* no longer defines a logistic function, the model is called as the three-parameter logistic (3-PL) model. The c-parameter is sometimes referred to as the *guessing parameter*, since its function is to account for the performance of persons with low-ability in responding to the test items (Birnbaum, 1968).

4. Other Extensions of the Rasch model.

   There are various extensions of the model presented in previous sections. We give some of them here.

   (a) Multidimensional Rasch model.

Stegelmann (1983) expanded the Rasch model to a general model that person ability parameters have more than one dimension in each category k.

$$P(X_{ij} = k) = \frac{\exp(\beta_{jk} - \sigma_{ik})}{\sum_{k''=0}^{m} \exp(\beta_{jk''} - \sigma_{ik''})}$$

(b) Linear Rating Scale model.

Fischer and Parzer (1991) extended the rating scale model in such a way that the item parameters are linearly decomposed into certain basic parameters, thus allowing incorporation of covariates. This extended model is denoted as the Linear Rating Scale model. Suppose we have P-dimensional covariates α_p, $p = 1, \ldots, P$, then

$$P(X_{ij} = k) = \frac{\exp(k\beta_j - k\delta_i - \sum_{k'=0}^{k} \tau_{k'})}{\sum_{k''=0}^{m} \exp(k''\beta_j - k''\delta_i - \sum_{k'=0}^{k''} \tau_{k'})}$$

where $\delta_i = \sum_p w_{ip}\alpha_p + c$, w_{ip} are weights for the item parameter δ_i associated with covariates α_p.

(c) Linear Partial Credit model.

Fischer and Ponocny (1994) extended the PCM under the assumption of a certain linear decomposition of the item-category parameters into basic parameters. This model is referred to as the Linear

Partial Credit model. Similar to that in the Linear Rating Scale model, given the P-dimensional covariates α_p, $p = 1, \ldots, P$, we have

$$P(X_{ij} = k) = \frac{\exp(k\beta_j - \sum_{k'=0}^{k} \delta_{ik'})}{\sum_{k''=0}^{m} \exp(k''\beta_j - \sum_{k'=0}^{k''} \delta_{ik'})}$$

where $\sigma_{ik} = \sum_{k'=0}^{k} \delta_{ik'} = \sum_p w_{ikp}\alpha_p + kc$, w_{ikp} are weights for δ_{ik} associated with covariates α_p.

(d) Multidimensional Polytomous Latent Trait (MPLT) model.

Kelderman and Rijkes (1994) further generalized the multidimensional Rasch model to a Multidimensional Polytomous Latent Trait model for polytomously scored items.

$$P(X_{ij} = k) = \frac{\exp \sum_{j=1}^{N} \alpha_{ijk}(\beta_j - \delta_{ik})}{\sum_{k''=0}^{m} \exp \sum_{j=1}^{N} \alpha_{ijk''}(\beta_j - \delta_{ik''})}$$

Table 2.2 summarizes the special cases of the polytomous Rasch model as well as some of their extensions.

2.3 Estimation and Inference for the Rasch Model

Various estimation strategies for the Rasch model have been proposed. Linacre (1999) provided an overview of currently available estimation procedures and software. Some are non-iterative, while many others are iterative methods. Non-iterative methods include graphical methods and normal approximation. Iterative methods include datum-by-datum method, marginal estimation methods with and without distributional assumptions. There are common software that incorporate these iterative methods. For example, PAIR is based on datum-by-datum method, JMLE, CMLE and log-linear models are marginal methods without distributional assumptions about the model parameters, while MMLE and PROX are marginal methods with distributional assumptions about the model parameters. Also, there is CON, which is based on the conditional likelihood method. After examining these popular software, Linacre (1999) concluded that the various Rasch estimation methods produced, in general, statistically equivalent results, although each has its own strengths and shortcomings.

Fisher (1934) showed that parameter separability is the necessary and sufficient condition for sufficient statistics. This leads to two main methods of estimation for the parameters: conditional maximum likelihood (CML) method and unconditional maximum likelihood (UML) method (Lehmann and Casella, 1996). The CML method takes full advantage of parameter separability by keeping the person parameters out of the calibration procedure entirely, while the UML method estimates the person parameters simultaneously with the item parameters.

2.3.1 Unconditional Maximum Likelihood Estimation

Wright and Masters (1982) used unconditional maximum likelihood method to estimate the Rasch parameters. The likelihood function based on the full data set is

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\delta}; \mathbf{X}) = \prod_{j=1}^{N} \prod_{i=1}^{I} P(X_{ij} = k_{ij}) \quad = \quad \frac{\exp \sum_{j=1}^{N} \sum_{i=1}^{I} \sum_{k=0}^{k_{ij}} (\beta_j - \delta_{ik})}{\prod_{j=1}^{N} \prod_{i=1}^{I} (\sum_{k''=0}^{m} \exp \sum_{k'=0}^{k''} (\beta_j - \delta_{ik'}))}$$

The log-likelihood function is

$$
\begin{aligned}
l(\boldsymbol{\beta},\boldsymbol{\delta};\mathbf{X}) &= \sum_{j=1}^{N}\sum_{i=1}^{I} k_{ij}\beta_j - \sum_{j=1}^{N}\sum_{i=1}^{I}\sum_{k=0}^{k_{ij}}\delta_{ik} \\
&\quad - \sum_{j=1}^{N}\sum_{i=1}^{I}\log(\sum_{k''=0}^{m}\exp\sum_{k'=0}^{k''}(\beta_j-\delta_{ik'})) \\
&= \sum_{j=1}^{N} r_j\beta_j - \sum_{i=1}^{I}\sum_{k=0}^{m} N_{ik}\delta_{ik} \\
&\quad - \sum_{j=1}^{N}\sum_{i=1}^{I}\log(\sum_{k''=0}^{m}\exp\sum_{k'=0}^{k''}(\beta_j-\delta_{ik'})) \qquad (2.6)
\end{aligned}
$$

where N_{ik} is the number of persons responding *on or above* category $k+1$ to item i so that $\sum_{j=1}^{N}\sum_{k=0}^{k_{ij}}\delta_{ik} = \sum_{k=0}^{m} N_{ik}\delta_{ik}$.

It can be shown that

$$
\begin{aligned}
\frac{\partial\log(\sum_{k''=0}^{m}\exp\sum_{k'=0}^{k''}(\beta_j-\delta_{ik'}))}{\partial\beta_j} &= \frac{\sum_{k''=0}^{m} k''\exp\sum_{k'=0}^{k''}(\beta_j-\delta_{ik'})}{\sum_{k''=0}^{m}\exp\sum_{k'=0}^{k''}(\beta_j-\delta_{ik'})} \\
&= \sum_{k=0}^{m} k\pi_{ijk} \\
\frac{\partial\log(\sum_{k''=0}^{m}\exp\sum_{k'=0}^{k''}(\beta_j-\delta_{ik'}))}{\partial\delta_{ik}} &= \frac{-\sum_{k''=k}^{m}\exp\sum_{k'=0}^{k''}(\beta_j-\delta_{ik'})}{\sum_{k''=0}^{m}\exp\sum_{k'=0}^{k''}(\beta_j-\delta_{ik'})} \\
&= -\sum_{k'=k}^{m}\pi_{ijk'}
\end{aligned}
$$

in which the item difficulty parameter δ_{ik} of category $k+1$ appears only in those terms for which $k' \geq k$, so that the derivative of $\sum_{k'=0}^{m} \delta_{ik'}$ with respect to δ_{ik} truncates the summation from $\sum_{k'=0}^{m}$ to $\sum_{k'=k}^{m}$.

Therefore, the first derivatives of the log-likelihood with respect to β_j and δ_{ik} are

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \beta_j} &= r_j - \sum_{i=1}^{I} \sum_{k=0}^{m} k\pi_{ijk} \\ \frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \delta_{ik}} &= -N_{ik} + \sum_{j=1}^{N} \sum_{k'=k}^{m} \pi_{ijk'} \end{aligned} \tag{2.7}$$

for $j = 1, \ldots, N; i = 1, \ldots, I; k = 1, \ldots, m$. This is the estimation equation for β_j and δ_{ik}.

The second derivative are

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \beta_j^2} &= -\sum_{i=1}^{I} (\sum_{k=0}^{m} k^2 \pi_{ijk} - (\sum_{k=0}^{m} k\pi_{ijk})^2) \\ \frac{\partial^2 l(\boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \delta_{ik}^2} &= -\sum_{j=0}^{N} (\sum_{k'=k}^{m} \pi_{ijk'} - (\sum_{k'=k}^{m} \pi_{ijk'})^2) \end{aligned}$$

respectively for β_j and δ_{ik}.

Estimators of these parameters may be obtained iteratively using, for example, the Newton-Raphson method. The iterative procedure gives the following formula:

$$\beta_r^{(t+1)} = \beta_r^{(t)} - \frac{r - \sum_{i=1}^{I}\sum_{k=0}^{m} k\hat{\pi}_{ik}^{(t)}(r)}{-\sum_{i=1}^{I}[\sum_{k=0}^{m} k^2\hat{\pi}_{ik}^{(t)}(r) - (\sum_{k=0}^{m} k\hat{\pi}_{ik}^{(t)}(r))^2]}$$

$$\delta_{ik}^{(t+1)} = \delta_{ik}^{(t)} - \frac{-N_{ik} + \sum_{r=1}^{Im-1} N(r)\sum_{k'=k}^{m}\hat{\pi}_{ik'}^{(t)}(r)}{-\sum_{r=1}^{Im-1} N(r)[\sum_{k'=k}^{m}\hat{\pi}_{ik'}^{(t)}(r) - (\sum_{k'=k}^{m}\hat{\pi}_{ik'}^{(t)}(r))^2]}$$

where $\hat{\pi}_{ik}^{(t)}(r)$ is the estimated probability of a person with a score r responding in category $k+1$ to item i after t iterations, and $N(r)$ is the number of persons with score r.

Asymptotic standard errors can be estimated from the denominator of the last iteration:

$$SE(\beta_r) = [\sum_{i=1}^{I}(\sum_{k=0}^{m} k^2\hat{\pi}_{ik}^{(t)}(r) - (\sum_{k=0}^{m} k\hat{\pi}_{ik}^{(t)}(r))^2)]^{-1/2}$$

$$SE(\delta_{ik}) = [\sum_{r=1}^{Im-1} N(r)(\sum_{k'=k}^{m}\hat{\pi}_{ik}^{(t)}(r) - (\sum_{k'=k}^{m}\hat{\pi}_{ik}^{(t)}(r))^2)]^{-1/2}$$

The UML method is relatively easy to program, and for this reason it is widely used in practice. However, the estimates are inconsistent for $N \to \infty$ with I and m fixed, although consistency does hold in some cases, for example $N, I, N/I \to \infty$ for dichotomous responses (Anderson, 1973a). In some cases, e.g. fixed I, the estimators may be biased, although the bias can be effectively removed in the dichotomous case by multiplying the estimates by $(I-1)/I$

(Wright and Douglas, 1977). It also suggests that the same correction may be appropriate for reducing the bias even in the polytomous response cases.

2.3.2 Conditional Maximum Likelihood Estimation

Among the estimation methods for the Rasch models, the conditional maximum likelihood method, despite its computational difficulty, has enjoyed popularity due to such desirable features as consistency and well-defined standard errors. It can be shown (Masters, 1982) that the conditional probability of response vector $\mathbf{x_j}$ given the score $r_j = r$ for person j is a function of the item parameters $\boldsymbol{\delta}$ only:

$$P(X_{1j} = k_1, \ldots, X_{Ij} = k_I | r_j = r) = \frac{\exp(-\sum_{i=1}^{I}\sum_{k=0}^{k_i}\delta_{ik})}{\sum_{\mathbf{x_j}\in\mathcal{R}}\exp(-\sum_{i=1}^{I}\sum_{k=0}^{k_i}\delta_{ik})} \tag{2.8}$$

where $\mathcal{R} = \{\mathbf{x_j} : r_j = r\}$.

Similarly the conditional probability of response vector $\mathbf{x_i}$ given the item score for $m + 1$ categories $\mathbf{s_i} = \mathbf{s} = (s_{i0}, \ldots, s_{im})'$ is a function of the person parameters $\boldsymbol{\beta}$ only:

$$P(X_{i1} = k_1, \ldots, X_{iN} = k_N | \mathbf{s_i} = \mathbf{s}) = \frac{\exp(-\sum_{j=1}^{N} k_j\beta_j)}{\sum_{\mathbf{x_i}\in\mathcal{S}}\exp(-\sum_{j=1}^{N} k_j\beta_j)} \tag{2.9}$$

where $\mathcal{S} = \{\mathrm{x_i} : \mathrm{s_i} = \mathrm{s}\}$.

The above two equations demonstrate the separability of the parameters in the Rasch model. This leads to the conditional maximum likelihood (CML) estimation method for the parameters.

It follows that the conditional probability of responding to category $k+1$ in item i given score $r_j = r$ is

$$\begin{aligned} \pi_{ijk}(r) &= P(X_{ij} = k | r_j = r) \\ &= \frac{\exp(-\sum_{k'=0}^{k} \delta_{ik'}) \sum_{\mathbf{X}_{-i} \in \mathcal{T}} \exp(-\sum_{i' \neq i}^{I} \sum_{k'=0}^{X_{i'j}} \delta_{i'k'})}{\sum_{k''=0}^{m} \{\exp(-\sum_{k'=0}^{k''} \delta_{ik'}) \sum_{\mathbf{X}_{-i} \in \mathcal{T}} \exp(-\sum_{i' \neq i}^{I} \sum_{k'=0}^{X_{i'j}} \delta_{i'k'})\}} \end{aligned} \tag{2.10}$$

where $\sum_{\mathbf{X}_{-i} \in \mathcal{T}}$ is the sum over all response vectors $\mathbf{X}_i$ which *exclude item* i and produce the score $r - k$.

We can then write the conditional likelihood over all N persons. On differentiating it with respective to δ_{ik}, we have the estimation equation for δ_{ik}. The ability parameters β_j then can be estimated by maximizing the likelihood with estimators of item difficulty parameters inserted. Estimators may be obtained iteratively using, for example, Newton-Raphson method. Asymptotic standard errors can be estimated from the denominator of the last iteration. This method is known as the CON procedure (Wright and Masters, 1982).

Fischer (1981) and Jacobsen (1989) gave necessary and sufficient conditions for the existence and uniqueness of the CML estimators in the dichotomous and polytomous Rasch models, respectively. Generally, CML estimators are consistent under regularity conditions (Rao, 1965, p. 299). Anderson (1973b) also assumed that the distribution of the CML estimators in the dichotomous Rasch model is asymptotically normal.

2.3.3 Goodness-of-fit Test

Tests of how well the data fit the Rasch model were conducted at several different levels. Tests of how well the items fit the Rasch model can identify problematic items: these may be items that differ qualitatively from others, or that are ambiguous or flawed. Tests of how well the person measures fit the model can identify persons with responses that do not follow the general pattern. Example of these tests can be found in Wright and Masters (1982).

2.3.4 Conditional Likelihood Ratio Test

Tests of global goodness-of-fit indicate the overall fit of a set of data to the model. Based on the CML method of parameter estimation, we have the conditional likelihood ratio test. The ratio statistic λ is defined as the ratio of the over-all conditional likelihood over the restricted likelihood function for all persons. Small values of λ indicates deviations from the model. It can be shown that in dichotomous case the statistic $G^2 = -2\log\lambda$ has a limiting $\chi^2_{(Im-1)(Im-2)}$ distribution as $N(r) \to \infty$ from the asymptotic properties of CML estimates (Anderson, 1973b).

2.4 Missing Data Analysis

Missing values are common in many experiments or surveys that involve human subjects. In most such studies, the unobserved data are treated as missing, in the sense that there are true underlying values that would have been observed if the data collection techniques had been better. However, treating the unobserved data as missing can sometimes lead to a serious estimation bias, depending on how the data have become unobservable and missing. We review several possible types of missing data situation and strategies to deal with them in this section. Our discussion is limited to the case of missing data on the outcome variable.

2.4.1 Types of Missing Data

The most appropriate way to handle missing or incomplete data depends on how data points became missing. Little and Rubin (1987) define three unique types of missing data mechanisms.

1. Data Missing Completely at Random (MCAR).

   This is when cases with complete data are indistinguishable from cases with incomplete data. Heitjan (1997) provides an example of data with MCAR, when a research associate shuffles raw data sheets and arbitrarily discards some of the sheets. In another example, data with MCAR can arise when an investigator randomly assigns research participants to complete only some portion of a survey instrument for some particular reasons, such as to save time and cost. Graham, Hofer, and MacKinnon (1996) illustrate the use of planned missing data patterns of this type to gather responses on more survey items from fewer research participants than those ordinarily obtainable from the standard survey study paradigm, in which every research participant receives and answers every survey question.

2. Data Missing at Random (MAR).

   This occurs when the cases with incomplete data differ from the cases with complete data, but the pattern of missing data is traceable or predictable from other variables in the database rather than the specific variable on which the data are missing. For example, if research participants with low self-esteem are less likely to return for follow-up sessions in a study that examines anxiety level over time as a function of self-esteem, and the researcher has self-esteem at the initial session available, then self-esteem is related to the values of the incomplete data at follow-up sessions. Another example is a test on reading comprehension: Investigators administers a reading comprehension test at the beginning of a survey administration session, and finds that research participants with low reading comprehension scores may be less likely to complete the entire survey. In both of these examples, we recognize that the actual variables where data are missing are not the cause of the incomplete data. Rather, the cause of the missing data is in some other external influence, clearly distinguishable from the case of data with MCAR.

3. Nonignorable Missing Data.

There are situations where the pattern of missing data is not completely random nor is predictable from other variables in the database. If a participant in a weight-loss study does not attend a weigh-in due to concerns about his weight loss, the response on this subject is missing due to nonignorable factors. In contrast to the MAR situation outlined above where the missing data pattern is explainable by other measured variables in a study, nonignorable missing data arise when the reason for the missing data is explainable, but unmeasurable, because the very variable(s) on which the data are missing are unobservable.

In practice, it is often difficult to make the MCAR assumption. There is often some degree of relationship between the missing values and other observed data. Therefore, the pattern of MAR is an assumption that is most often, but not always, tenable.

2.4.2 Methods of Handling Missing Data

In this section, we review how various researchers developed strategies to deal with missing data (Little and Rubin, 1987). Some commonly used methods for handling missing data are listed below, covering those widely recognized approaches to handling data with incomplete cases.

1. Deletion method

   Here, the missing data will be simply discarded. There are two data deletion methods used often, typically when using standard statistical software.

   (a) List-wise or case-wise data deletion.

   In this case, when a record for a subject contains missing data for any variable, his/her entire record is removed before the analysis.

   (b) Pairwise data deletion.

   This is possible for certain computations. For example, in calculating bivariate correlations or covariances, the available pairwise data are used, discarding only the incomplete pairs.

2. Imputation method

   In this approach, each missing value will be estimated and treated as if it is an actual observation in the analysis.

   (a) Mean substitution.

   For a variable with missing data in some records, that variable's mean value, computed from complete cases, is substituted for all missing values.

   (b) Regression methods.

   In this approach, a regression equation is constructed based on complete subset data. Then, for cases where Y is missing, their missing values are imputed using the predicted values from the regression equation and they are used in the subsequent analysis. An improvement to this method involves adding uncertainty to the imputed values of Y so that the mean response value is not always the one to be imputed, given by the multiple imputation method, described below.

(c) Multiple imputation.

This is similar to the maximum likelihood estimation method for the missing values. However, the multiple imputation generates several values suitable for filling in for each of the missing data. Typically, 2 to 5 data sets are created in this fashion with 2 to 5 different imputed values for each of the missing data. The investigator then analyzes these multiple data sets using an appropriate statistical analysis method, treating them as if they were actual complete data. The results from these analysis are then combined into a single summary report. Huang (2001) discussed that the extra resources required for analyzing multiple data sets appeared unnecessary if simple imputation is done informatively.

(d) Hot deck imputation.

Here, the investigator will identify the most similar case to the case with a missing value and substitute the case's Y value for the missing case's Y value. Disadvantage in this approach is that the distribution of the data may not be preserved by repeated use of the available data.

3. Expectation Maximization (EM) approach.

   This is an iterative procedure that proceeds in two discrete steps. First, in the expectation (E) step, the expected value of the complete data log likelihood is computed. In the maximization (M) step, the expected values obtained from the E step is substituted for the missing data, followed by maximizing the likelihood function to obtain new parameter estimates as if no data were missing. The procedure iterates through these two steps until convergence is reached.

4. Maximum likelihood method.

   This method use all available data to generate maximum likelihood-based sufficient statistics. Usually these consist of a covariance matrix of the variables and a vector of means. This technique is also known as Full Information Maximum Likelihood (Wothke, 1998). Carrière (1994) discussed that a simple and non-iterative but valid technique is possible for small sample repeated measures data. Carrière (1999) also reported two important findings relevant to all small sample studies that (1) non-iterative, simple, efficient and valid technique is possible for missing data; and (2) even with small sample data, for which various covariance models may be indistinguishable, the empirical size and power are sensitive to misspecified assumptions about the covariance structure.

5. Re-sampling Method: Jackknife and bootstrap method.

   Jackknife is an estimating method by deleting one datum each time from the original data set and re-calculating the estimator based on the rest of the data. Bootstrap utilizes all $2^n - 1$ nonempty subsets of a data set of size n. Both of these re-sampling methods can be used for missing data analysis.

Little and Rubin (1987) and Wothke (1998) review these methods and conclude that list-wise, pairwise, and mean substitution missing data handling methods are inferior when compared with likelihood based methods such as incomplete maximum likelihood or multiple imputation. Regression methods are somewhat better, but not as good as hot deck imputation or maximum likelihood approaches. The EM method falls somewhere in between; it is generally superior to list-wise, pairwise, and mean substitution approaches, but it lacks the ability to accommodate the uncertainty that is possible with the incomplete maximum likelihood and multiple imputation methods.

Table 2.1: The data matrix

Item	*Subject*					
	1	$\dots$	j	$\dots$	N	*item score*
1	X_{11}	$\dots$	X_{1j}	$\dots$	X_{1N}	s_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
i	X_{i1}	$\dots$	X_{ij}	$\dots$	X_{iN}	s_i
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
I	X_{I1}	$\dots$	X_{Ij}	$\dots$	X_{IN}	s_I
subject score	r_1	$\dots$	r_j	$\dots$	r_N	

Table 2.2: Various Rasch models and their extensions

Response data type	$P(X_{ij} = k) \propto$
Dichotomous	0 or $\exp(\beta_j - \delta_i)$
Unidimensional Rasch	$\exp(k\beta_j - \sigma_{ik})$
Partial Credit	$\exp(k\beta_j - \sum_{k'=0}^{k} \delta_{ik'})$
Rating Scale	$\exp(k\beta_j - k\delta_i - \sum_{k'=0}^{k} \tau_{k'})$
Binomial Trials	$\exp(k\beta_j - k\delta_i)/\frac{k!(m-k)!}{m!}$
Poisson Counts	$\exp(k\beta_j - k\delta_i)/k!$
2-PL Model	$\exp(k\alpha_i\beta_j - \sigma_{ik})$
Linear Rating Scale	$\exp(k\beta_j - (k\sum_p w_{ip}\alpha_p + kc + \sum_{k'=0}^{k} \tau_{k'}))$
Linear Partial Credit	$\exp(k\beta_j - (\sum_p w_{ikp}\alpha_p + kc))$
Multidimensional Rasch	$\exp(\beta_{jk} - \sigma_{ik})$
MPLT	$\exp \sum_{j=1}^{N} \alpha_{ijk}(\beta_j - \delta_{ik})$

Chapter 3

An Improved CML Estimation Procedure

3.1 Introduction

Consider the response data in *Chapter 2*, where I multiple-choice questions are given to N subjects. For question i ($i = 1, \ldots, I$) and subject j ($j = 1, \ldots, N$), we obtain X_{ij}, which takes a value k ($k = 0, \ldots, m$); there are $m+1$ possible values (categories) for each question.

In the Rasch estimation method used today, the parameters are thought to be independent and unrelated; they are estimated separately and independently of one another. Now we consider improving the CON procedure (Wright and Masters, 1982) to implement a simultaneous estimation method for the polytomous Rasch model, based on the conditional maximum likelihood. We also derive the asymptotic properties of the CML estimators and then develop a conditional likelihood ratio test for the goodness-of-fit of the model. Finally, simulation studies were used to assess the performance of our approach, as compared to the CON procedure.

3.2 Modified CML Estimation

Consider the unidimensional polytomous Rasch model (2.3) in *Chapter 2*:

$$\begin{aligned}\pi_{ijk} = P(X_{ij} = k) &= \frac{\exp\sum_{k'=0}^{k}(\beta_j - \delta_{ik'})}{\sum_{k''=0}^{m}\exp\sum_{k'=0}^{k''}(\beta_j - \delta_{ik'})} \\ &= \frac{\exp(k\beta_j - \sigma_{ik})}{\sum_{k''=0}^{m}\exp(k''\beta_j - \sigma_{ik''})}\end{aligned}$$

We will now discuss how we propose to modify the estimation method used in the CON procedure, which is based on the conditional likelihood function (Sheng and Carrière, 2002). To define the conditional likelihood function, we will first consider the elementary symmetric function, known as the γ-function (Wright and Masters, 1982), which is used in almost every step of the numerical calculation. Define a subject score $r_j = \sum_{i=1}^{I} X_{ij}$ and a γ-function with score

r as:

$$\gamma(r) = \sum_{\mathbf{x_j} \in \mathcal{R}} \exp(-\sum_{i=1}^{I} \sum_{k'=0}^{X_{ij}} \delta_{ik'}) \tag{3.1}$$

where $\sum_{\mathbf{x_j} \in \mathcal{R}}$ is the sum over the permutation of response vectors $\mathbf{X}_j$ with score r. Note that $\gamma(r)$ is one component in the model (2.3), based on which the conditional likelihood estimation will be formulated as in (2.10). We then extend some properties of the γ-functions to polytomous data as given in *Lemma 3.1.*

Lemma 3.1 *The first and second derivatives of the γ-functions are:*

$$\begin{aligned}
\frac{\partial \gamma(r)}{\partial \delta_{ik}} &= -\sum_{k'=k}^{m} (\exp(-\sigma_{ik'})\gamma_{-i}(r-k')) \\
\frac{\partial^2 \gamma(r)}{\partial \delta_{ik}^2} &= \sum_{k'=k}^{m} (\exp(-\sigma_{ik'})\gamma_{-i}(r-k')) \\
\frac{\partial^2 \gamma(r)}{\partial \delta_{ik} \partial \delta_{pq}} &= \sum_{k'=max(k,q)}^{m} (\exp(-\sigma_{ik'}) \exp(-\sigma_{pk'})\gamma_{(-i,-p)}(r-k')) \\
& \qquad (i \neq p \text{ or } k \neq q)
\end{aligned}$$

where $\gamma_{-i}(r-k)$ is the "reduced" γ-function that excludes item i and produces score $r-k$.

Proof:

The first order derivative of a γ-function with respect to δ_{ik} is

$$
\begin{aligned}
\frac{\partial\gamma(r)}{\partial\delta_{ik}} &= \frac{\partial\sum_{\mathbf{x_j}\in\mathcal{R}}\exp(-\sum_{i=1}^{I}\sum_{k'=0}^{X_{ij}}\delta_{ik})}{\partial\delta_{ik}} \\
&= \sum_{\mathbf{x_j}\in\mathcal{R}}\frac{\partial\exp(-\sum_{i=1}^{I}\sum_{k'=0}^{X_{ij}}\delta_{ik})}{\partial\delta_{ik}}
\end{aligned}
$$

Note that only the terms $(X_{ij} \geq k)$ contain δ_{ik} in their expression, and therefore we have

$$
\begin{aligned}
\frac{\partial\gamma(r)}{\partial\delta_{ik}} &= \sum_{\mathbf{x_j}\in\mathcal{R}}\left(-I_{(X_{ij}\geq k)}\exp(-\sum_{k'=0}^{X_{ij}}\delta_{ik'} - \sum_{i'\neq i}\sum_{k'=0}^{X_{i'j}}\delta_{i'k'})\right) \\
&= -\sum_{k'=k}^{m}(\exp(-\sigma_{ik'})\gamma_{-i}(r-k'))
\end{aligned}
$$

for an indicator function $I(.)$, because the second part of the exponent is independent of δ_{ik}.

The second order derivative with respect to δ_{ik} is

$$
\frac{\partial^2\gamma(r)}{\partial\delta_{ik}^2} = \partial\left(\frac{\partial\gamma(r)}{\partial\delta_{ik}}\right)\Big/\partial\delta_{ik}
$$

Using the first derivative obtained above, we find

$$
\begin{aligned}
\frac{\partial^2\gamma(r)}{\partial\delta_{ik}^2} &= \frac{\partial \sum_{\mathbf{x_j}\in\mathcal{R}} \left(-I_{(X_{ij}\geq k)} \exp(-\sum_{k'=0}^{X_{ij}} \delta_{ik'} - \sum_{i'\neq i}\sum_{k'=0}^{X_{i'j}} \delta_{i'k'})\right)}{\partial\delta_{ik}} \\
&= \sum_{\mathbf{x_j}\in\mathcal{R}} \left(I_{(X_{ij}\geq k)} \exp(-\sum_{k'=0}^{X_{ij}} \delta_{ik'} - \sum_{i'\neq i}\sum_{k'=0}^{X_{i'j}} \delta_{i'k'}) \right) \\
&= \sum_{k'=k}^{m} (\exp(-\sigma_{ik'})\gamma_{-i}(r-k'))
\end{aligned}
$$

The second order derivative with respect to the parameters δ_{ik} and δ_{pq} is

$$
\begin{aligned}
\frac{\partial^2\gamma(r)}{\partial\delta_{ik}\partial\delta_{pq}} &= \partial\left(\frac{\partial\gamma(r)}{\partial\delta_{ik}}\right) \Big/ \partial\delta_{pq} \\
&= \frac{\partial \sum_{\mathbf{x_j}\in\mathcal{R}} \left(-I_{(X_{ij}\geq k)} \exp(-\sum_{k'=0}^{X_{ij}} \delta_{ik'} - \sum_{i'\neq i}\sum_{k'=0}^{X_{i'j}} \delta_{i'k'})\right)}{\partial\delta_{pq}}
\end{aligned}
$$

Since only those terms ($X_{ij} \geq k$ & $X_{pj} \geq q$) contain both δ_{ik} and δ_{pq}, we therefore have

$$
\begin{aligned}
\frac{\partial^2\gamma(r)}{\partial\delta_{ik}\partial\delta_{pq}} &= \sum_{\mathbf{x_j}\in\mathcal{R}} \left(I_{(X_{ij}\geq k, X_{pj}\geq q)} \exp(-\sum_{k'=0}^{X_{ij}} \delta_{ik'} - \sum_{k'=0}^{X_{pj}} \delta_{pk'} - \sum_{i'\neq i,p}\sum_{k'=0}^{X_{i'j}} \delta_{i'k'}) \right) \\
&= \sum_{k'=max(k,q)}^{m} (\exp(-\sigma_{ik'})\exp(-\sigma_{pk'})\gamma_{(-i,-p)}(r-k'))
\end{aligned}
$$

for $i \neq p$ *or* $k \neq q$.

□

A simple combinational argument gives the recursive formula (Wright and Masters, 1982):

$$\gamma(r) = \sum_{k=0}^{m} \exp(-\sigma_{ik})\gamma_{-i}(r-k) \tag{3.2}$$

for $i = 1, \ldots, I$, which also applies to the reduced γ-function, $\gamma_{-i}(r)$. That is, for $i = 1, \ldots, I$, we can obtain the recursive formulae used in the calculation of the CML method:

$$\gamma_{(-1,-2,\ldots,-(i-1))}(r) = \sum_{k=0}^{m} \exp(-\sigma_{ik})\gamma_{(-1,-2,\ldots,-i)}(r-k) \tag{3.3}$$

These recursive formulae given in (3.2) and (3.3) play a very important role, as they are used to reduce the heavy loading of computing time in iteration. Using these functions, we obtain the following conditional probability, as summarized in *Lemma 3.2*.

Lemma 3.2 *The conditional probability of subject j responding to category $k+1$ on item i, given score $r_j = r$, is*

$$\pi_{ijk}(r_j = r) = \frac{\exp(-\sigma_{ik})\gamma_{-i}(r-k)}{\gamma(r)}$$

Proof:

Let $\sum_{\mathbf{X}_{-i} \in \mathcal{T}}$ denote the sum over the permutations of response vectors $\mathbf{X}_j$ that exclude item i and produce the score $r - k$. Then, by definition,

$$\pi_{ijk}(r_j = r) = P(X_{ij} = k | r_j = r)$$

which is equivalent to

$$\pi_{ijk}(r_j = r) = P(X_{ij} = k, \sum_{i' \neq i} X_{i'j} = r - k)$$

This is just one case in which the response vector $\mathbf{X}_j$ produces score r. By including all permutations of such cases, it follows that the probability

$$\begin{aligned}\pi_{ijk}(r_j = r) &= \frac{\exp(-\sum_{k'=0}^{k} \delta_{ik'}) \sum_{\mathbf{X}_{-i} \in \mathcal{T}} \exp(-\sum_{i' \neq i} \sum_{k'=0}^{X_{i'j}} \delta_{i'k'})}{\sum_{k''=0}^{m} \{\exp(-\sum_{k'=0}^{k''} \delta_{ik'}) \sum_{\mathbf{X}_{-i} \in \mathcal{T}} \exp(-\sum_{i' \neq i} \sum_{k'=0}^{X_{i'j}} \delta_{i'k'})\}} \\ &= \frac{\exp(-\sigma_{ik}) \gamma_{-i}(r - k)}{\sum_{k=0}^{m} \exp(-\sigma_{ik}) \gamma_{-i}(r - k)}\end{aligned}$$

by the definition of γ-functions. Then, from the recursive formula in Equations 3 and 4, we have

$$\pi_{ijk}(r_j = r) = \frac{\exp(-\sigma_{ik}) \gamma_{-i}(r - k)}{\gamma(r)}$$

$\square$

Therefore, using *Lemma 3.2*, we can write the conditional likelihood of the item difficulty parameters $\boldsymbol{\delta}$ over all N subjects as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\delta};\mathbf{r}) &= \prod_{j=1}^{N}\frac{\exp(-\sum_{i=1}^{I}\sum_{k'=0}^{k_{ij}}\delta_{ik'})}{\gamma(r)} \\ &= \frac{\exp(-\sum_{j=1}^{N}\sum_{i=1}^{I}\sum_{k'=0}^{k_{ij}}\delta_{ik'})}{\prod_{r=0}^{Im}(\gamma(r))^{N(r)}} \qquad (3.4)\end{aligned}$$

where $N(r)$ is the number of subjects with score r. The log-likelihood is then

$$l(\boldsymbol{\delta};\mathbf{r}) = -\sum_{i=1}^{I}\sum_{k=0}^{m}N_{ik}\delta_{ik} - \sum_{r=1}^{Im-1}N(r)\log\gamma(r), \qquad (3.5)$$

because $\sum_{j=1}^{N}\sum_{k'=0}^{k_{ij}}\delta_{ik'} = \sum_{k=0}^{m}N_{ik}\delta_{ik}$, where N_{ik} is the number of subjects responding on or above category $k+1$ to item i (i.e. $X_{ij}=k$).

The first derivative of the log-likelihood with respect to δ_{ik} is

$$\begin{aligned}\frac{\partial l(\boldsymbol{\delta};\mathbf{r})}{\partial\delta_{ik}} &= -N_{ik} - \sum_{r=1}^{Im-1}N(r)\frac{\partial\log\gamma(r)}{\partial\delta_{ik}} \\ &= -N_{ik} + \sum_{r=1}^{Im-1}N(r)\sum_{k'=k}^{m}\frac{\exp(-\sigma_{ik'})\gamma_{-i}(r-k')}{\gamma(r)}, \qquad (3.6)\end{aligned}$$

which gives the estimation equation for δ_{ik}. Note that the summation of the latter quantity is just $\sum_{k'=k}^{m}\pi_{ik'}(r)$, the probability of responding to k or larger, given a score of r.

The Rasch model estimators can now be obtained via an iterative technique, for example, the Newton-Raphson method. Note that the CON procedure does not consider the second order partial derivatives with respect to two different parameters. It assumes them to be simply zero, which is to assume that the Rasch parameters are mutually independent. In contrast, we allow interdependence among the parameters, a condition that is common and rather realistic. The Hessian matrix given in *Theorem 3.3* below is the results upon applying *Lemma 3.1* to $\log\gamma(r)$.

Theorem 3.3 *Let the* $(I*m)\times(I*m)$ *matrix* $H(r,\boldsymbol{\delta})$ *be given by the Hessian matrix, the second derivatives of* $\log\gamma(r)$*, with its elements given by*

$$\frac{\partial^2\log\gamma(r)}{\partial\delta_{ik}^2} = \sum_{k'=k}^{m}\left(\frac{\exp(-\sigma_{ik'})\gamma_{-i}(r-k')}{\gamma(r)} - \left(\frac{\exp(-\sigma_{ik'})\gamma_{-i}(r-k')}{\gamma(r)}\right)^2\right)$$

and

$$\begin{aligned}\frac{\partial^2\log\gamma(r)}{\partial\delta_{ik}\partial\delta_{pq}} &= \frac{\sum_{k'=max(k,q)}^{m}\exp(-\sigma_{ik'})\exp(-\sigma_{pk'})\gamma_{(-i,-p)}(r-k')}{\gamma(r)} - \\ &\quad \frac{\sum_{k'=k}^{m}(\exp(-\sigma_{ik'})\gamma_{-i}(r-k'))\sum_{q'=q}^{m}(\exp(-\sigma_{pq'})\gamma_{-p}(r-q'))}{(\gamma(r))^2} \\ &\quad (i\neq p \text{ or } k\neq q). \qquad (3.7)\end{aligned}$$

Proof:

First, note that

$$\frac{\partial \log \gamma(r)}{\partial \delta_{ik}} = \frac{\partial \gamma(r)/\partial \delta_{ik}}{\gamma(r)}$$

Now, the diagonal elements of the Hessian matrix are

$$\begin{aligned}\frac{\partial^2 \log \gamma(r)}{\partial \delta_{ik}^2} &= \partial \left(\frac{\partial \log \gamma(r)}{\partial \delta_{ik}} \right) \Big/ \partial \delta_{ik} \\ &= \partial \left(\frac{\partial \gamma(r)/\partial \delta_{ik}}{\gamma(r)} \right) \Big/ \partial \delta_{ik}\end{aligned}$$

Using the basic rules for derivatives, we have

$$\begin{aligned}\frac{\partial^2 \log \gamma(r)}{\partial \delta_{ik}^2} &= \frac{\partial^2 \gamma(r)/\partial \delta_{ik}^2}{\gamma(r)} - \left(\frac{\partial \gamma(r)/\partial \delta_{ik}}{\gamma(r)} \right)^2 \\ &= \sum_{k'=k}^{m} \left(\frac{\exp(-\sigma_{ik'})\gamma_{-i}(r-k')}{\gamma(r)} - \left(\frac{\exp(-\sigma_{ik'})\gamma_{-i}(r-k')}{\gamma(r)} \right)^2 \right)\end{aligned}$$

The last step follows from *Lemma 3.1.*

Similarly, the off-diagonal elements of the Hessian matrix are

$$\begin{aligned}
\frac{\partial^2 \log \gamma(r)}{\partial \delta_{ik} \partial \delta_{pq}} &= \partial \left(\frac{\partial \gamma(r)/\partial \delta_{ik}}{\gamma(r)} \right) \Big/ \partial \delta_{pq} \\
&= \left(\frac{\partial^2 \gamma(r)}{\partial \delta_{ik} \partial \delta_{pq}} \Big/ \gamma(r) \right) - \left(\frac{\partial \gamma(r)}{\partial \delta_{ik}} \frac{\partial \gamma(r)}{\partial \delta_{pq}} \Big/ (\gamma(r))^2 \right) \\
&= \frac{\sum_{k'=max(k,q)}^{m} \exp(-\sigma_{ik'}) \exp(-\sigma_{pk'}) \gamma_{(-i,-p)}(r-k')}{\gamma(r)} - \\
&\quad \frac{\sum_{k'=k}^{m} (\exp(-\sigma_{ik'}) \gamma_{-i}(r-k')) \sum_{q'=q}^{m} (\exp(-\sigma_{pq'}) \gamma_{-p}(r-q'))}{(\gamma(r))^2}
\end{aligned}$$

$$(i \neq p \; or \; k \neq q)$$

□

Further, define $H(\boldsymbol{\delta}) = \sum_{r=1}^{Im-1} N(r) H(r, \boldsymbol{\delta})$. Then, the iterative estimation procedure updates the estimates at each iteration as

$$\hat{\boldsymbol{\delta}}^{(t)} = \hat{\boldsymbol{\delta}}^{(t-1)} + \left(H^{-1}(\boldsymbol{\delta}) \frac{\partial l(\boldsymbol{\delta}; \mathbf{r})}{\partial \boldsymbol{\delta}} \right) \Big|_{\hat{\boldsymbol{\delta}}^{(t-1)}}, \qquad t = 1, \ldots, T. \tag{3.8}$$

The asymptotic covariance matrix of $\hat{\boldsymbol{\delta}}$ can be estimated as $H^{-1}(\hat{\boldsymbol{\delta}}^{(T)})$ after the last iteration T. The ability parameters $\boldsymbol{\beta}$ can then be estimated. using the usual unconditional maximum likelihood method, by maximizing the complete likelihood as a function of the estimated item difficulty parameters.

3.3 Properties of the New Estimators

CML estimators are consistent under regularity conditions (Rao, 1965, p. 299) when a minimal sufficient statistic exists (Anderson, 1970). In the Rasch model, the minimal sufficient statistic is the subject score r_j for the ability parameter β_j for subject j. We will now consider asymptotic properties of the CML estimators of the item difficulty parameters $\boldsymbol{\delta}$. Let $\hat{\boldsymbol{\delta}}_{\mathrm{x_j} \in \mathcal{R}}$ be the restricted CML estimators, and let $\boldsymbol{\delta}_0$ be the true value of the parameter.

Theorem 3.4 *(Consistency of CML estimators)*

1. *When $N(r) \to \infty$, then $\hat{\boldsymbol{\delta}}_{\mathrm{x_j} \in \mathcal{R}} \xrightarrow{P} \boldsymbol{\delta}_0$, and furthermore $\hat{\boldsymbol{\delta}}_{\mathrm{x_j} \in \mathcal{R}}$ is asymptotically normally distributed, with mean $\boldsymbol{\delta}_0$ and asymptotic covariance matrix $H^{-1}(r, \boldsymbol{\delta}_0)/N(r)$, i.e. $\hat{\boldsymbol{\delta}}_{\mathrm{x_j} \in \mathcal{R}} \sim AN(\boldsymbol{\delta}_0, H^{-1}(r, \boldsymbol{\delta}_0)/N(r))$.*

2. *When $N(r) \to \infty$, $N(r)/N$ exists for all r, then $\hat{\boldsymbol{\delta}} \xrightarrow{P} \boldsymbol{\delta}_0$, that is, the overall CML estimators for $\boldsymbol{\delta}$ are consistent, and $\hat{\boldsymbol{\delta}}$ is also asymptotically normally distributed, with mean $\boldsymbol{\delta}_0$ and asymptotical covariance matrix $H^{-1}(\boldsymbol{\delta}_0)$, i.e. $\hat{\boldsymbol{\delta}} \sim AN(\boldsymbol{\delta}_0, H^{-1}(\boldsymbol{\delta}_0))$.*

Proof:

For the dichotomous case, see Anderson (1973b). We will now sketch the proof for the polytomous case.

1. The first part of the proof follows immediately from the estimation results of maximizing the conditional likelihood function because the response vector $\mathbf{x_j}$'s are i.i.d. random vectors.

2. The overall CML estimation equation and the restricted CML estimation equation are connected by

$$\frac{\partial l(\boldsymbol{\delta};\mathbf{r})}{\partial \boldsymbol{\delta}} = \sum_{r=1}^{Im-1} \frac{\partial l_{\mathbf{x_j}\in\mathcal{R}}(\boldsymbol{\delta};r)}{\partial \boldsymbol{\delta}}$$

It follows from the first part above that the unique solutions to each term on the right side converge in probability to the same vector $\boldsymbol{\delta}_0$. Continuity arguments show that the solutions to the left side must converge in probability to $\boldsymbol{\delta}_0$ as well.

To prove the asymptotic distribution, we will now consider a Taylor expansion of $\partial l(\boldsymbol{\delta};\mathbf{r})/\partial\boldsymbol{\delta}$ around $\hat{\boldsymbol{\delta}}$. Since $\hat{\boldsymbol{\delta}}$ satisfies $\partial l(\boldsymbol{\delta};\mathbf{r})/\partial\boldsymbol{\delta} = \mathbf{0}$, we get

$$\partial l(\boldsymbol{\delta};\mathbf{r})/\partial\boldsymbol{\delta} = \left(\sum_{r=1}^{Im-1} N(r)H(r,\boldsymbol{\delta}^*)\right)(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0)$$

where $||\boldsymbol{\delta}^* - \boldsymbol{\delta}_0|| \leq ||\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0||$.

From the first part of the second proof, $\hat{\boldsymbol{\delta}} \xrightarrow{P} \boldsymbol{\delta}_0$, hence, $\boldsymbol{\delta}^* \xrightarrow{P} \boldsymbol{\delta}_0$ as

$N(r) \to \infty$ for all r. Then the right side has the same limiting distribution as

$$\left(\sum_{r=1}^{Im-1} N(r)H(r, \boldsymbol{\delta}_0) \right) (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0),$$

i.e.

$$(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \sim H^{-1}(\boldsymbol{\delta}_0)\mathbf{D}(\boldsymbol{\delta}_0)$$

where $\mathbf{D}(\boldsymbol{\delta})$ denotes the vector $\partial l(\boldsymbol{\delta}; \mathbf{r})/\partial \boldsymbol{\delta}$.

For each r, $l_{\mathbf{x_j} \in \mathcal{R}}(\boldsymbol{\delta}; r)$ is the likelihood function of $N(r)$ i.i.d. random vectors. Hence, from the maximum likelihood estimation result, we get

$$\mathbf{D}_{\mathbf{x_j} \in \mathcal{R}}(r, \boldsymbol{\delta}_0) \sim AN(\mathbf{0}, N(r)H(r, \boldsymbol{\delta}_0)).$$

However, since $\mathbf{D}(\boldsymbol{\delta}_0) = \sum_{r=1}^{Im-1} \mathbf{D}_{\mathbf{x_j} \in \mathcal{R}}(r, \boldsymbol{\delta}_0)$, it follows immediately that

$$\mathbf{D}(\boldsymbol{\delta}_0) \sim AN(\mathbf{0}, H(\boldsymbol{\delta}_0)).$$

Therefore, $(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0)$ has a limiting normal distribution with mean $\mathbf{0}$ and covariance matrix

$$H^{-1}(\boldsymbol{\delta}_0)H(\boldsymbol{\delta}_0)H^{-1}(\boldsymbol{\delta}_0) = H^{-1}(\boldsymbol{\delta}_0).$$

$\square$

In consequence of *Theorem 3.4*, we can make more accurate inferences about these parameters than it was possible using the current procedure. For example, suppose we want to test whether two items are of the same "difficulty". The test statistic is obtained as a linear combination of $\hat{\delta}$'s. Using *Theorem 3.3* and *Theorem 3.4.2*, we can obtain the estimated asymptotic variance of the test statistic, which is asymptotically normal under the null hypothesis. Therefore, we can use standard tools to test the hypotheses concerning the difficulty of items. Similarly, we can make more accurate inferences about $\boldsymbol{\beta}$ as a function of $\boldsymbol{\delta}$.

Since the CML method maximizes the conditional likelihood, not the full likelihood, this implies loss of information, thus, in general the estimates are not efficient. Anderson (1970), however, shows the efficiency holds if the sufficient statistic r is "weakly" ancillary with respect to the ability parameter $\boldsymbol{\beta}$. Later Anderson (1973a) shows this indeed holds for the Rasch model. The conclusion is that when $N(r) \to \infty$ for all r, the proposed CML estimates are asymptotically efficient, and that loss of information becomes negligible.

3.4 Conditional Likelihood Ratio Test

The conditional likelihood ratio test is intimately related to the CML method of parameter estimation. The conditional likelihood ratio is defined as

$$\lambda = \frac{\mathcal{L}(\hat{\boldsymbol{\delta}}; \mathbf{r})}{\prod_{r=1}^{Im-1} \mathcal{L}_{\mathbf{x_j} \in \mathcal{R}}(\hat{\boldsymbol{\delta}}_{\mathbf{x_j} \in \mathcal{R}}; r)} \tag{3.9}$$

where $\mathcal{L}(\boldsymbol{\delta}; \mathbf{r})$ is the full conditional likelihood and $\mathcal{L}_{\mathbf{x_j} \in \mathcal{R}}(\boldsymbol{\delta}; r)$ is the restricted likelihood function for subjects with score r. Since

$$\mathcal{L}(\boldsymbol{\delta}; \mathbf{r}) = \prod_{r=1}^{Im-1} \mathcal{L}_{\mathbf{x_j} \in \mathcal{R}}(\boldsymbol{\delta}; r),$$

it is obvious that $\lambda \leq 1$. Small values of λ indicate deviations from the model. Consider the test statistic

$$G^2 = -2 \log \lambda = 2 \sum_{r=1}^{Im-1} l_{\mathbf{x_j} \in \mathcal{R}}(\hat{\boldsymbol{\delta}}_{\mathbf{x_j} \in \mathcal{R}}; r) - 2l(\hat{\boldsymbol{\delta}}; \mathbf{r}) \tag{3.10}$$

and we have the following *Theorem 3.5* for the limiting distribution of G^2.

Theorem 3.5 *When $N(r) \to \infty$ for all r, the statistic $G^2 = -2 \log \lambda$ has a limiting χ^2 distribution with degrees of freedom $(Im - 1)(Im - 2)$.*

Proof: Anderson (1973b) gave a proof of *Theorem 3.5* for the dichotomous case, a proof that can be easily extended to polytomous cases as following.

Since $l(\boldsymbol{\delta}) = \sum_{r=1}^{Im-1} l_{\mathrm{x_j}\in\mathcal{R}}(\boldsymbol{\delta})$, we may rewrite G^2 as

$$G^2 = 2\left(l(\boldsymbol{\delta}_0) - l(\hat{\boldsymbol{\delta}})\right) - 2\left(\sum_{r=1}^{Im-1} l_{\mathrm{x_j}\in\mathcal{R}}(\boldsymbol{\delta}_0) - \sum_{r=1}^{Im-1} l_{\mathrm{x_j}\in\mathcal{R}}(\hat{\boldsymbol{\delta}}_{\mathrm{x_j}\in\mathcal{R}})\right)$$

Now let us expand $l(\boldsymbol{\delta}_0)$ around $\hat{\boldsymbol{\delta}}$, and $l_{\mathrm{x_j}\in\mathcal{R}}(\boldsymbol{\delta}_0)$ around $\hat{\boldsymbol{\delta}}_{\mathrm{x_j}\in\mathcal{R}}$, we get

$$\begin{aligned} G^2 &= -(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0)^T H(\boldsymbol{\delta}^*)(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \\ &\quad + \sum_{r=1}^{Im-1} (\hat{\boldsymbol{\delta}}_{\mathrm{x_j}\in\mathcal{R}} - \boldsymbol{\delta}_0)^T N(r) H(r, \boldsymbol{\delta}^*_{\mathrm{x_j}\in\mathcal{R}})(\hat{\boldsymbol{\delta}}_{\mathrm{x_j}\in\mathcal{R}} - \boldsymbol{\delta}_0) \end{aligned}$$

where $||\boldsymbol{\delta}^* - \boldsymbol{\delta}_0|| \le ||\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0||$ and $||\boldsymbol{\delta}^*_{\mathrm{x_j}\in\mathcal{R}} - \boldsymbol{\delta}_0|| \le ||\hat{\boldsymbol{\delta}}_{\mathrm{x_j}\in\mathcal{R}} - \boldsymbol{\delta}_0||$.

When $N(r) \to \infty$ for all r, from *Theorem 3.4* we know that $\hat{\boldsymbol{\delta}} \xrightarrow{P} \boldsymbol{\delta}_0$ and $\hat{\boldsymbol{\delta}}_{\mathrm{x_j}\in\mathcal{R}} \xrightarrow{P} \boldsymbol{\delta}_0$, so $\boldsymbol{\delta}^* \xrightarrow{P} \boldsymbol{\delta}_0$ and $\boldsymbol{\delta}^*_{\mathrm{x_j}\in\mathcal{R}} \xrightarrow{P} \boldsymbol{\delta}_0$. Therefore, G^2 has the same limiting distribution as

$$-(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0)^T H(\boldsymbol{\delta}_0)(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) + \sum_{r=1}^{Im-1} (\hat{\boldsymbol{\delta}}_{\mathrm{x_j}\in\mathcal{R}} - \boldsymbol{\delta}_0)^T N(r) H(r, \boldsymbol{\delta}_0)(\hat{\boldsymbol{\delta}}_{\mathrm{x_j}\in\mathcal{R}} - \boldsymbol{\delta}_0)$$

From the proof of *Theorem 3.4*, we know that

$$(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \sim H^{-1}(\boldsymbol{\delta}_0)\mathbf{D}(\boldsymbol{\delta}_0),$$

and similarly

$$(\hat{\boldsymbol{\delta}}_{\mathrm{x_j}\in\mathcal{R}} - \boldsymbol{\delta}_0) \sim (N(r)H(r, \boldsymbol{\delta}_0))^{-1}\mathbf{D}_{\mathrm{x_j}\in\mathcal{R}}(\boldsymbol{\delta}_0).$$

Hence,

$$\begin{aligned} G^2 \sim & -\mathbf{D}^T(\boldsymbol{\delta}_0)H(\boldsymbol{\delta}_0)^{-1}\mathbf{D}(\boldsymbol{\delta}_0) \\ & + \sum_{r=1}^{Im-1} \mathbf{D}^T_{\mathbf{x_j}\in\mathcal{R}}(\boldsymbol{\delta}_0)(N(r)H(r,\boldsymbol{\delta}_0))^{-1}\mathbf{D}_{\mathbf{x_j}\in\mathcal{R}}(\boldsymbol{\delta}_0) \end{aligned}$$

Now define vector $D^*(\boldsymbol{\delta}) = (\mathbf{D}^T_{\mathbf{x_j}\in 1}(\boldsymbol{\delta}), \ldots, \mathbf{D}^T_{\mathbf{x_j}\in(Im-1)}(\boldsymbol{\delta}))^T$, two $(Im-1)^2 \times (Im-1)^2$ matrices $H^*(\boldsymbol{\delta}) = diag\{(N(1)H(1,\boldsymbol{\delta}))^{-1}, \ldots, (N(Im-1)H(Im-1,\boldsymbol{\delta}))^{-1}\}$ and $H^{**}(\boldsymbol{\delta}) = \mathbf{1}\ diag\{H^{-1}(\boldsymbol{\delta}), \ldots, H^{-1}(\boldsymbol{\delta})\}\ \mathbf{1}^T$, then

$$G^2 \sim \mathbf{D}^*(\boldsymbol{\delta}_0)^T(H^*(\boldsymbol{\delta}_0) - H^{**}(\boldsymbol{\delta}_0))\mathbf{D}^*(\boldsymbol{\delta}_0)$$

Since $\mathbf{D}^T_{\mathbf{x_j}\in 1}(\boldsymbol{\delta}), \ldots, \mathbf{D}^T_{\mathbf{x_j}\in(Im-1)}(\boldsymbol{\delta})$ are independent to each other and asymptotically normal, applying a condition given by Rao (1965, P.152, Eq(3b.4.6)), G^2 is asymptotically χ^2 if

$$(H^*(\boldsymbol{\delta}_0) - H^{**}(\boldsymbol{\delta}_0))(H^*(\boldsymbol{\delta}_0))^{-1}(H^*(\boldsymbol{\delta}_0) - H^{**}(\boldsymbol{\delta}_0)) = H^*(\boldsymbol{\delta}_0) - H^{**}(\boldsymbol{\delta}_0),$$

or equivalently,

$$H^{**}(\boldsymbol{\delta}_0)(H^*(\boldsymbol{\delta}_0))^{-1}H^{**}(\boldsymbol{\delta}_0) = H^{**}(\boldsymbol{\delta}_0),$$

which is obvious from the definition of $H^{**}(\boldsymbol{\delta})$ and $H^*(\boldsymbol{\delta})$. The degrees of freedom of the χ^2 distribution is equal to

$$\begin{aligned}
rank(H^*(\boldsymbol{\delta}_0) - H^{**}(\boldsymbol{\delta}_0)) &= rank((H^*(\boldsymbol{\delta}_0) - H^{**}(\boldsymbol{\delta}_0))(H^*(\boldsymbol{\delta}_0))^{-1}) \\
&= trace((H^*(\boldsymbol{\delta}_0) - H^{**}(\boldsymbol{\delta}_0))(H^*(\boldsymbol{\delta}_0))^{-1}) \\
&= trace(\mathbb{I} - H^{**}(\boldsymbol{\delta}_0)(H^*(\boldsymbol{\delta}_0))^{-1}) \\
&= (Im-1)^2 - (Im-1) \\
&= (Im-1)(Im-2)
\end{aligned}$$

$\square$

3.5 Simulation Results

We shall now assess, via computer simulation, the performance of our estimation method in dichotomous and polytomous Rasch models, thus verifying the theoretical results that we obtained in the previous sections. We will also show that our simultaneous approach outperforms that based on the CON procedure in the Rasch models. We replicated the simulation 1000 times to obtain the empirical distribution of the CML estimators and that of the conditional likelihood ratio test statistic G^2. We then compared our results with those obtained from the CON procedure.

3.5.1 Dichotomous Rasch Model

In this case, the response data are binary ($m = 1$). First, we generated data for $N = 100$ subjects for $I = 4$ items from a given distinct set of parameters for the ability of persons

$$\beta' = (0, 0.25, 0.5, 0.75, 1)$$

and for the difficulty of items

$$\delta = \begin{pmatrix} 0 & -0.3 \\ 0 & -0.1 \\ 0 & 0.1 \\ 0 & 0.3 \end{pmatrix}$$

with $\mathbf{1}'\boldsymbol{\delta}\mathbf{1} = 0$ for identifiability (see Wright and Masters, 1982).

Figure 3.1 shows the Q-Q plot of G^2 versus quantiles from the χ^2 distribution with 6 degrees of freedom, using our modified approach and the conventional CML approach, respectively. Here, we used the empirical distribution of the parameters to estimate the true distribution upon verifying the normality assumption. The plot shows that our approach using G^2 is indeed approximately distributed as χ^2 with 6 degrees of freedom, with a few exceptions at large quantiles, whereas results from the CON procedure deviate from the assumed distribution. Consistent with the mean and variance of the χ^2_6 distribution, our approach had a mean of 6.368 and a variance of 11.74. The results for the CON procedure were 10.25 and 21.59, respectively, which almost double the numbers expected under the assumed distribution.

Table 3.1 shows the ratios of variances and MSEs of estimators of $\boldsymbol{\beta}$ using our procedure, divided by those using CON. The inestimable parameters (Linacre, 1999) of $\beta = 0$ and $\beta = 1$ were not included. Our approach produced smaller biases, variances, and MSEs than those of CON. In both approaches, the biases of the estimators proved to be negligible. Comparison of $\hat{\boldsymbol{\delta}}$ (available from the authors) resulted in much the same conclusion.

3.5.2 Polytomous Rasch Model

We also considered the polytomous Rasch model with multinomial response data, with $m = 2$. As in the previous simulation, we generated data for $N = 100$ subjects using $I = 4$ items for a distinct set of parameters for the ability of persons

$$\boldsymbol{\beta}' = (0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2)$$

and for the difficulty of items

$$\boldsymbol{\delta} = \begin{pmatrix} 0 & -0.2962 & 0.2185 \\ 0 & -0.4270 & 0.4717 \\ 0 & -0.3885 & 0.6263 \\ 0 & -0.3852 & 0.1803 \end{pmatrix}$$

with $\mathbf{1}'\boldsymbol{\delta}\mathbf{1} = 0$ for identifiability (see Wright and Masters, 1982).

Again, we carried out 1000 simulations to observe the empirical performance of the estimators. *Figure 3.2* shows the Q-Q plot of G^2 versus quantiles from the χ^2_{42}. The plot clearly shows that G^2 is approximately distributed as χ^2_{42}. Our approach had a mean of 45.53 and a variance of 91.05, again consistent with those of the χ^2_{42} distribution. The CON procedure resulted in 54.30 and 191.4, respectively.

Table 3.2 shows the ratios of variances and MSEs of the estimators of $\boldsymbol{\beta}$ using our procedure, divided by those using CON. In general, the results were unbiased or manifested very little bias. The variances and MSEs obtained using our approach were smaller than those from the CON procedure, with some exceptions at extreme values of r's.

3.5.3 Generalization

To draw an overall conclusion and to appreciate the effect of increasing the size of items, the number of categories, and the number of subjects, we expanded the scope of the simulation studies. *Table 3.3* shows the summary of these further analysis. When the number of subjects increases, more precise estimators are obtained, with smaller variances and MSEs, as was anticipated. Further, the more items or categories there are, the less precisely we can estimate the parameters using our approach as compared to the CON procedure. One unusually large ratio presented in the case of 20 items, 3 categories and 2000 subjects was caused by an estimate in one of the simulations from the CON procedure that produced extremely large bias, and thus large MSE.

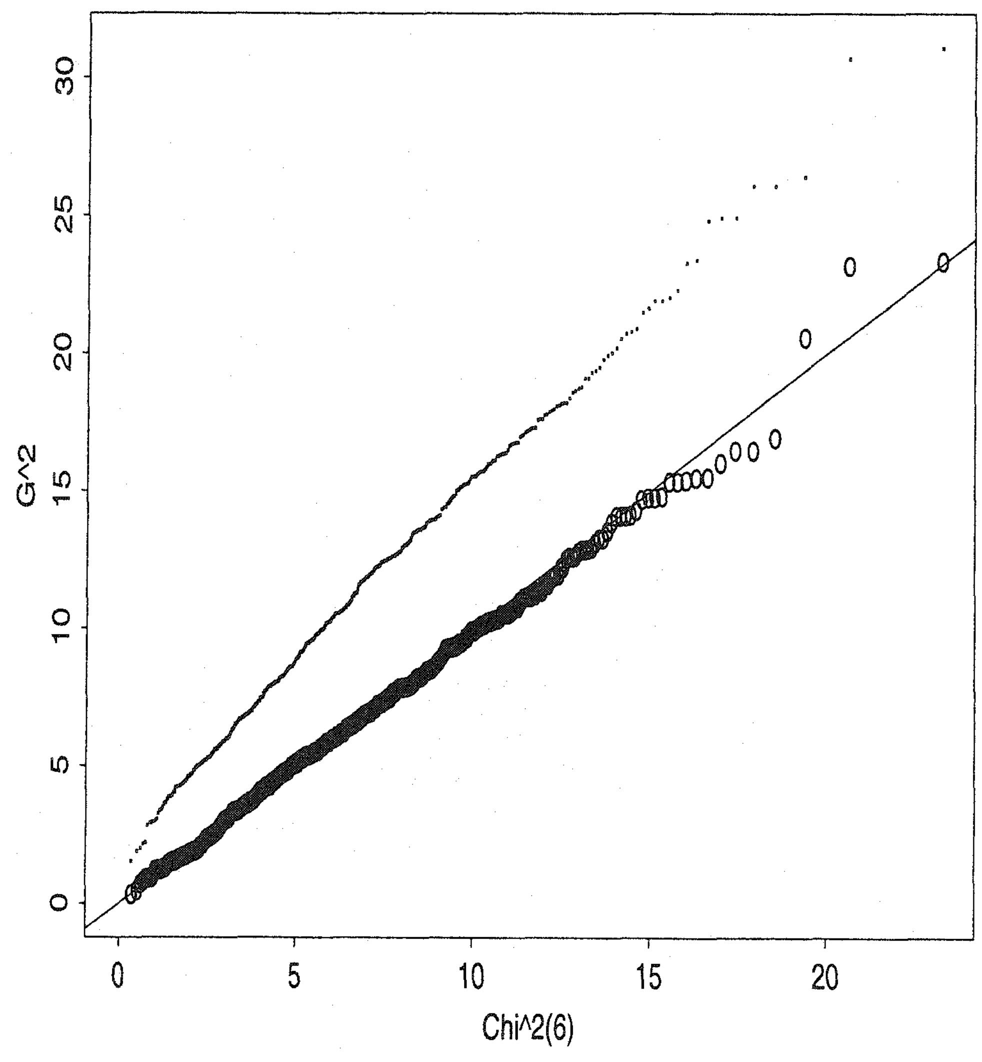
3.6 Summary

In this study, we implemented a simultaneous CML estimation method for dichotomous and polytomous Rasch models and derived their asymptotic properties. The advantage to this approach of carrying out the simultaneous estimation method was rather substantial in both the polytomous and dichotomous Rasch models. The improvement in results, as compared to those of the currently used conditional approach, were especially apparent for intermediate values of the ability parameters. In our approach, we did not assume the independence of the Rasch model parameters. This approach is also useful when testing hypotheses concerning some functions of these parameters, for example, the linear contrast of several selected parameters of interest.

The simultaneous estimation method also had a huge impact on the model fit. We constructed the conditional likelihood ratio test for the goodness-of-fit of the model. The test statistic was shown to be distributed according to the assumed asymptotic χ^2 distribution. This was as expected. On the other hand, the corresponding result based on the current conditional approach deviates significantly from the expected distribution. The asymptotic distribution of the statistic using the current method should be a χ^2, if the Rasch parameters are indeed independent to each other. However, because the Rasch parameters are naturally correlated, we were not able to verify its asymptotic distribution in simulation.

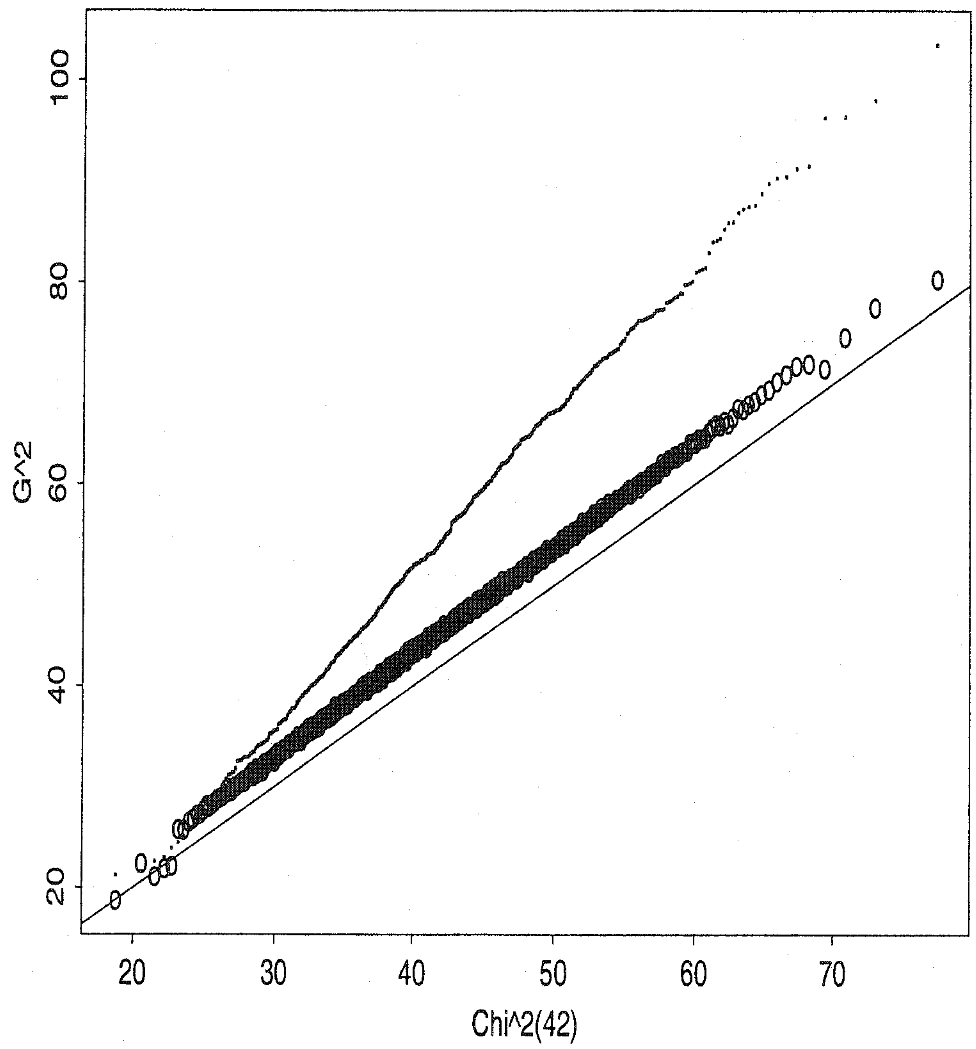
In summary, our conclusion is that the current approach has shortcomings in not considering correlations implicit in the Rasch model parameters. Our implementation, based on a conditional likelihood function, improved the fit of the data, as well as the precision of the estimators in comparison with those of the other CML method known as CON. In light of the recent review (Linacre, 1999) of various estimation methods for the Rasch model, which found all to be statistically identical, we suggest that our method is superior to all others currently available.

Figure 3.1: Q-Q plot of G^2 vs quantiles of the χ^2 distribution in the dichotomous model ($I = 4, m = 1, N = 100$)



Note: Shown are the quantiles from the respective χ^2 distributions (solid line), the G^2 statistic from the CON procedure (indicated with dots), and the G^2 statistic from the proposed procedure (indicated with small circles).

Figure 3.2: Q-Q plot of G^2 vs quantiles of the χ^2 distribution in the polytomous model ($I = 4, m = 2, N = 100$)



Note: Shown are the quantiles from the respective χ^2 distributions (solid line), the G^2 statistic from the CON procedure (indicated with dots), and the G^2 statistic from the proposed procedure (indicated with small circles).

Table 3.1: Comparison of the two estimation procedures for β in the dichotomous model $(I = 4, m = 1, N = 100)$

β	$\widehat{Var(\hat{\beta})}$	$\widehat{MSE(\hat{\beta})}$
0.25	0.3431	0.3428
0.50	0.0073	0.0073
0.75	0.3028	0.3282
overall	0.2092	0.2087

Note: The entries are ratios of variances and MSEs based on the CON procedure divided by those of the proposed method in this chapter. The row for 'overall' corresponds to that of the trace of the variance and MSE matrices. Ratios smaller than 1 indicate improvement via our method.

Table 3.2: Comparison of the two estimation procedures for $\boldsymbol{\beta}$ in the polytomous model ($I = 4, m = 2, N = 100$)

β	$\widehat{Var(\hat{\beta})}$	$\widehat{MSE(\hat{\beta})}$
0.25	3.7175	3.7594
0.50	0.8163	1.0320
0.75	0.1639	0.2734
1.00	0.0166	0.0195
1.25	0.1657	0.2248
1.50	0.8842	1.0858
1.75	3.7879	3.8760
overall	0.8078	0.9058

Note: The entries are ratios of variances and MSEs based on the current method compared to those of the proposed method in this chapter. The row for 'overall' corresponds to that of the trace of the variance and MSE matrices. Ratios smaller than 1 indicate improvement via our method.

Table 3.3: Summary comparisons of the two estimation procedures

(I, m, N)	$\widehat{Var(\hat{\beta})}$	$\widehat{MSE(\hat{\beta})}$
(4, 1, 100)	0.2092	0.2087
(4, 1, 2000)	0.1152	0.1303
(4, 2, 100)	0.8078	0.9058
(4, 2, 2000)	0.6803	0.7077
(20, 1, 100)	0.6083	0.5721
(20, 1, 2000)	0.9183	0.3744
(20, 2, 2000)	0.4072	0.0094

Note: The entries are ratios of total variances and total MSEs of the estimated β (i.e., trace of the variance and MSE matrix), based on the CON procedure divided by those of the proposed method in this chapter. Ratios smaller than 1 indicate improvement via our method.

Chapter 4

Accounting for the Polychoric Correlation Among Items in the Rasch Model

4.1 Introduction

In the previous chapter, we presented some important improvements on the current Rasch analysis method. In this chapter, we consider another improvement, recognizing that in these investigations, the questionnaire items are inter-correlated. For example, in a study using items that measure symptom distress on a scale, many of the items are closely correlated. If one complains about frequency of chest pains, he/she will most likely also complain about their intensity. As another example, consider a clinical trial of patients in a hospital. If one complains severe cough problems, he/she usually develops sore throat. The correlation between these items has not been taken into consideration in the development of methodology for Rasch models. As evident in the literature (for example, Liang and Zeger, 1986), ignoring the presence of significant correlations can lead to a loss of efficiency and serious bias in the study conclusions.

In this chapter, we develop a method of accounting for inter-item correlation. Correlation among items that are measured on an ordinal scale is called the polychoric correlation (Olsson, 1979). The polychoric correlation coefficient is formulated using the concept of latent variables, variables that are usually continuous, although unobservable, giving rise to the apparent complexity of the data (Miller et al. 1962). We use the generalized estimation equations approach (GEE) to obtain consistent estimates of the parameters of the Rasch model. Simulation study demonstrates the relative of these estimates efficiency over those obtained without considering the inter-item correlation.

4.2 Polychoric Correlation for Ordinal Random Variables

As this thesis deals with ordinal data, we first discuss how to estimate the dependency in these data. Suppose we observed two ordinal random variables x and y, and consider the threshold approach to the analysis. That is, we assume that there exist corresponding latent variables ξ and η so that $x = i$ if and only if $a_{i-1} < \xi \leq a_i$, and $y = j$ if and only if $b_{j-1} < \eta \leq b_j$, for $(i, j) \in \{0, \ldots, m\}$ where m is the number of categories x and y take, and a_i and b_j are threshold parameters.

When considering dependency among ordinal categorical data, use of the ordinary Pearson correlation is not recommended. Olsson (1979) showed that using the Pearson correlation leads to biased estimates. Instead, one should use the polychoric correlation, i.e. the correlation among the underlying latent random variables ξ and η.

Poon and Lee (1987) developed the most general model for estimating the polychoric correlation. The full maximum likelihood estimates of the polychoric correlation coefficient and threshold parameters were obtained via the Fletcher-Powell algorithm. A computationally more efficient approach, called "the partition maximum likelihood method," was also proposed by Poon and

Lee (1987). However, the asymptotic properties of the estimates obtained using this method were not investigated. Poon, Lee, and Bentler (1990) used a pseudo maximum likelihood approach, which is computationally more efficient than the full maximum likelihood approach, and discussed the asymptotic distribution of the estimates.

Ronning and Kukuk (1996) compared the efficiency of the estimators from a joint likelihood against those of a conditional likelihood for measuring the association of two ordinal variables. The maximum likelihood estimators of the correlation and threshold parameters are consistent in both approaches; however, estimators from the conditional model are less efficient.

In the Rasch analysis, the threshold parameters are related to the model specification. We used a method similar to the full maximum likelihood method to obtain an estimate of the polychoric correlation coefficient in the context of Rasch models, and we used the GEE method to obtain estimates of the Rasch model parameters.

4.3 Generalized Estimation Equations

In this section, we review the GEE method, as we will be using this technique to estimate the parameters in the Rasch model. Consider a response variable y_{ij}, not necessarily measured independently, which belongs to an exponential family where the marginal distribution is

$$f(y_{ij}) = \exp((y_{ij}\theta_{ij} - a(\theta_{ij}) + b(y_{ij}))\phi)$$

where ϕ is a common scale parameter. This property allows the analyst to use the usual maximum likelihood approach to estimate the parameters θ_{ij}.

Liang and Zeger (1986) proposed the Generalized Estimation Equations (GEE) method for longitudinal data. For each respondent $\mathbf{y_j}$, they assumed a "working" correlation matrix $\mathbf{R_j}(\boldsymbol{\alpha})$, which is fully specified by a vector of unknown parameters $\boldsymbol{\alpha}$. The "working" covariance matrix is then given by

$$\mathbf{V_j} = \mathbf{A}_j^{1/2}\mathbf{R_j}(\alpha)\mathbf{A}_j^{1/2}/\phi$$

where $\mathbf{A_j}$ is an $n_j \times n_j$ diagonal matrix with the diagonal elements being functions of μ_{ij} with a common scale parameter ϕ.

The generalized estimation equations proposed by Liang and Zeger (1986) are

$$\sum_{j=1}^{n_j} \mathbf{D_j}'\mathbf{V_j}^{-1}\mathbf{s_j} = \mathbf{0} \tag{4.1}$$

where $\mathbf{s_j} = \mathbf{y_j} - \boldsymbol{\mu}_j$ and $\mathbf{D_j} = \partial\boldsymbol{\mu}_j/\partial\boldsymbol{\theta}$. They showed that the GEE method gives consistent estimates of the model parameters and of their variances under weak assumptions about the joint distribution, when consistent estimators of $\boldsymbol{\alpha}$ and ϕ are given. It does not require specifying a form for the joint distribution of the repeated measurements, only the first moment $\boldsymbol{\mu}_j$ and the correct structure of correlation $\mathbf{R_j}(\boldsymbol{\alpha})$. This approach thus provides a general estimation method that can be used even for non-normal response variables.

4.4 Polychoric Correlation in the Rasch Model

Now we discuss the implemetation of the polychoric correlation on the Rasch model. Consider the unidimensional polytomous Rasch model (2.3) in *Chapter 2*:

$$\begin{aligned}\pi_{ijk} = P(X_{ij} = k) &= \frac{\exp \sum_{k'=0}^{k}(\beta_j - \delta_{ik'})}{\sum_{k''=0}^{m} \exp \sum_{k'=0}^{k''}(\beta_j - \delta_{ik'})} \\ &= \frac{\exp(k\beta_j - \sigma_{ik})}{\sum_{k''=0}^{m} \exp(k''\beta_j - \sigma_{ik''})}\end{aligned}$$

We show that this model can also be derived from the concept of a latent variable approach, if the cut-off threshold parameters are chosen appropriately. Corresponding to the ordinal data X_{ij}, we assume the existence of a continuous random variable W_{ij} such that $X_{ij} = k$ if and only if $W_{ij} \in (c_{ij,k-1}, c_{ij,k}]$ where

$k = 0, \ldots, m$ with $c_{ij,-1} = -\infty$ and $c_{ij,m} = \infty$. For $k = 0, \ldots, m-1$, let

$$c_{ij,k} = \Psi^{-1}\left(\frac{\sum_{k'=0}^{k} \exp(k'\beta_j - \sigma_{ik'})}{\sum_{k'=0}^{m} \exp(k'\beta_j - \sigma_{ik'})}\right) \tag{4.2}$$

where $\Psi^{-1}(\alpha)$ is the upper $100(1-\alpha)\%$ point of an assumed distribution. Immediately, we get

$$\begin{aligned} P(X_{ij} = k) &= P(c_{ij,k-1} < W_{ij} \leq c_{ij,k}) \\ &= \Psi(c_{ij,k}) - \Psi(c_{ij,k-1}) \\ &= \frac{\sum_{k'=0}^{k} \exp(k'\beta_j - \sigma_{ik'})}{\sum_{k'=0}^{m} \exp(k'\beta_j - \sigma_{ik'})} - \frac{\sum_{k'=0}^{k-1} \exp(k'\beta_j - \sigma_{ik'})}{\sum_{k'=0}^{m} \exp(k'\beta_j - \sigma_{ik'})} \\ &= \frac{\exp(k\beta_j - \sigma_{ik})}{\sum_{k'=0}^{m} \exp(k'\beta_j - \sigma_{ik'})} \end{aligned} \tag{4.3}$$

This is the Rasch model (2.3), from a latent variables point of view. If we assume a normal distribution for W_{ij}, we take $c_{ij,k} = \Phi^{-1}(.)$, where $\Phi^{-1}(\alpha)$ is the upper $100(1-\alpha)\%$ point of the standard normal distribution.

We can extend this idea to a multi-item case, where there can be more than one response items. Suppose $\mathbf{W_j} = (W_{1j}, \ldots, W_{Ij})'$ is the latent response vector corresponding to the j-th individual. We assume that $\mathbf{W_j}$ follows a multivariate distribution with a joint c.d.f. $\mathbf{\Psi}_j$ such that each W_{ij} has the marginal c.d.f. Ψ_{ij}. Again, take $c_{ij,k}$ to be

$$c_{ij,k} = \Psi_{ij}^{-1}\left(\frac{\sum_{k'=0}^{k}\exp(k'\beta_j - \sigma_{ik'})}{\sum_{k'=0}^{m}\exp(k'\beta_j - \sigma_{ik'})}\right). \tag{4.4}$$

We observe that the Rasch model (2.3) holds for each component X_{ij}. Moreover they now have dependency according to the specification of $\mathbf{\Psi}_j$.

Intuitively, we treat the correlation between any two items as if it is same for all individuals. Thus one immediate choice of the joint distribution is Φ, the I-dimensional multivariate normal distribution with mean $\mathbf{0}$ and a variance matrix with compound symmetry structure and an equal correlation ρ among items:

$$\Sigma = (1-\rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}'$$

where $\mathbf{I}$ is the I-dimensional identity matrix. Then the joint probability that the j-th individual has the response k_i on the i-th item ($i = 1, \ldots, I$), is

$$\begin{aligned}\pi_{(1,\ldots,I),j}(k_1,\ldots,k_I) &= P(X_{1j} = k_1, \ldots, X_{Ij} = k_I) \\ &= \sum_{i_1,\ldots,i_I=(0,1)} (-1)^{i_1+\ldots+i_I} \times \\ &\quad \Phi(c_{1j,k_1-i_1}, \ldots, c_{Ij,k_I-i_I}) \end{aligned} \tag{4.5}$$

In the simple special case of bivariate normal situation where $I = 2$, or for any two items within a multi-item framework, we have

$$\begin{aligned}
\pi_{(i,i'),j}(k,k') &= P(X_{ij}=k, X_{i'j}=k') \\
&= \Phi(c_{ij,k}, c_{i'j,k'}) - \Phi(c_{ij,k}, c_{i'j,k'-1}) \\
&\quad -\Phi(c_{ij,k-1}, c_{i'j,k'}) + \Phi(c_{ij,k-1}, c_{i'j,k'-1}). \qquad (4.6)
\end{aligned}$$

for $i \neq i'$.

4.5 Estimation and Inference in the Rasch Model with Correlated Items

In this section, we will discuss the estimation and inference procedures in the extended Rasch model for the correlated data using the latent variable approach. The parameters to be estimated are $\boldsymbol{\beta}$ (which includes $I * m - 1$ distinct values of β_j's), $\boldsymbol{\delta}$, and ρ. Denote $\boldsymbol{\theta}$ as the collection of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, all together with $p = ((I * m - 1) + I * m) = 2I * m - 1$ parameters, and $\boldsymbol{\theta}_\rho$ is the collection of $\boldsymbol{\theta}$ and ρ.

4.5.1 Estimation of the Polychoric Correlation Coefficient

Given the values of $\boldsymbol{\theta}$, we can find the maximum likelihood estimator of $\rho(\boldsymbol{\theta})$ in the following manner. This method is similar to the full maximum likelihood (Poon and Lee, 1987), with consideration of the Rasch model in constructing the threshold parameters, as described in the previous section.

Given any two correlated items $\mathbf{x_i}$ and $\mathbf{x_{i'}}$, $(i, i') \in \{1, \ldots, I\}$, the full likelihood for ρ is:

$$\mathcal{L}_\rho(\boldsymbol{\theta}) = \prod_j \prod_{k,k'} \left(\pi_{(i,i'),j}(k, k')\right)^{I(j,k,k')}$$

where $I(j, k, k')$ is an indicator function of whether the j-th subject responds to categories k and k' to the respective two items, and $\pi_{(i,i'),j}(k, k')$ is given as in (4.6). Note that the likelihood function is not shown as a direct function of ρ, but through the expression of $\pi_{(i,i'),j}(k, k')$. The log-likelihood is

$$l_\rho(\boldsymbol{\theta}) = \log \mathcal{L}_\rho(\boldsymbol{\theta}) = \sum_j \sum_{k,k'} I(j, k, k') \log(\pi_{(i,i'),j}(k, k'))$$

Then the estimating equation for ρ is

$$\frac{\partial l_\rho(\boldsymbol{\theta})}{\partial \rho} = \sum_j \sum_{k,k'} \frac{I(j,k,k')}{\pi_{(i,i'),j}(k,k')} \times \pi'_{(i,i'),j}(k,k') = 0 \tag{4.7}$$

where

$$\begin{aligned}
\pi'_{(i,i'),j}(k,k') &= \frac{d\pi_{(i,i'),j}(k,k')}{d\rho} \\
&= \Big[\phi(c_{ij,k}, c_{i'j,k'}) - \phi(c_{ij,k}, c_{i'j,k'-1}) \\
&\quad -\phi(c_{ij,k-1}, c_{i'j,k'}) + \phi(c_{ij,k-1}, c_{i'j,k'-1})\Big]
\end{aligned}$$

and $\phi(.)$ is the density function of a bivariate standard normal distribution with correlation coefficient ρ.

We can plug in an initial estimate of $\boldsymbol{\theta}$ into (4.7) to get an estimate of ρ through an iterative method. The second-order derivative of the log-likelihood function with respect to ρ is:

$$\begin{aligned}
\frac{\partial^2 l_\rho(\boldsymbol{\theta})}{\partial \rho^2} &= \sum_j \sum_{k,k'} I(j,k,k') \times \\
&\quad \left[\frac{\pi''_{(i,i'),j}(k,k')}{\pi_{(i,i'),j}(k,k')} - \left(\frac{\pi'_{(i,i'),j}(k,k')}{\pi_{(i,i'),j}(k,k')}\right)^2\right]
\end{aligned} \tag{4.8}$$

where

$$\begin{aligned}
\pi''_{(i,i'),j}(k,k') &= \frac{d^2\pi_{(i,i'),j}(k,k')}{d\rho^2} \\
&= \Big[\phi'(c_{ij,k}, c_{i'j,k'}) - \phi'(c_{ij,k}, c_{i'j,k'-1}) \\
&\quad -\phi'(c_{ij,k-1}, c_{i'j,k'}) + \phi'(c_{ij,k-1}, c_{i'j,k'-1})\Big]
\end{aligned}$$

and

$$\phi'(x,y) = \Big(\frac{\rho}{1-\rho^2} + \frac{(x-\rho y)(y-\rho x)}{(1-\rho^2)^2}\Big)\phi(x,y).$$

Therefore, the iterative equation gives

$$\hat{\rho}^{(t)} = \hat{\rho}^{(t-1)} + \left[\frac{\partial^2 l_\rho(\boldsymbol{\theta})}{\partial\rho^2}\right]^{-1}\left[\frac{\partial l_\rho(\boldsymbol{\theta})}{\partial\rho}\right]\Bigg|_{\hat{\rho}^{(t-1)}}, \qquad t = 1,\ldots,T. \tag{4.9}$$

at some T.

This is how the polychoric correlation coefficient $\rho_{i,i'}$ between any two items i, i' is estimated. Based on the estimators of $\rho_{i,i'}$ for all $(I(I-1)/2)$ item pairs, the common ρ for all I items can be estimated by taking the average of these $\hat{\rho}_{i,i'}$ as follows:

$$\hat{\rho} = \frac{2\sum_{i<i'}\hat{\rho}_{i,i'}}{I(I-1)} \tag{4.10}$$

It follows from the general results of the maximum likelihood estimators that if the initial estimate of $\boldsymbol{\theta}$ is consistent given the true value of ρ, then $\hat{\rho}$ is consistent and asymptotically normal with mean ρ and variance estimated by

$$\widehat{Var(\hat{\rho})} = \frac{4\left(\frac{\partial^2 l_\rho(\boldsymbol{\theta})}{\partial \rho^2}\right)^{-1}\Big|_{\hat{\rho}^{(T)}}}{(I(I-1))^2}.$$

4.5.2 Estimation of the Rasch Parameters

For $(i, i') \in \{1, \ldots, I\}$, let us define

$$\begin{aligned}
\mu_{ij} &= E(X_{ij}) = \sum_{k=0}^{m} k\pi_{ijk} = \frac{\sum_{k=0}^{m} k \exp(k\beta_j - \sigma_{ik})}{\sum_{k=0}^{m} \exp(k\beta_j - \sigma_{ik})} \\
\sigma_{(i,i),j} &= Var(X_{ij}) = \sum_{k=0}^{m} k^2 \pi_{ijk} - \mu_{ij}^2 \\
\sigma_{(i,i'),j} &= Cov(X_{ij}, X_{i'j}) = \sum_{k,k'=0}^{m} kk'\pi_{(i,i'),j}(k,k') - \mu_{ij}\mu_{i'j}.
\end{aligned}$$

The estimating equation for $\boldsymbol{\theta}$ is

$$\mathbf{U}(\boldsymbol{\theta}) = \sum_{j=1}^{N} \mathbf{D_j}'\boldsymbol{\Omega}_j^{-1}(\mathbf{x_j} - \boldsymbol{\mu}_j) = \mathbf{0}, \tag{4.11}$$

where

$$\mathbf{x_j} = (X_{1j}, \ldots, X_{Ij})', \qquad \boldsymbol{\mu}_j = (\mu_{1j}, \ldots, \mu_{Ij})',$$

$\mathbf{\Omega}_j$ is the variance-covariance matrix of $\mathbf{x_j}$:

$$\mathbf{\Omega}_j = \begin{pmatrix} \sigma_{(1,1),j} & \cdots & \sigma_{(1,I),j} \\ \vdots & \ddots & \vdots \\ \sigma_{(I,1),j} & \cdots & \sigma_{(I,I),j} \end{pmatrix},$$

and the $I \times p$ matrix $\mathbf{D_j}$ is given by

$$\mathbf{D_j} = \frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\theta}}$$

Theorem 4.1 *The elements of* $\mathbf{D_j}$ *are given by*

$$\begin{aligned}
\frac{\partial \mu_{ij}}{\partial \beta_j} &= \frac{\sum_{k'=0}^{m} k'^2 \exp(k'\beta_j - \sigma_{ik'})}{\sum_{k'=0}^{m} \exp(k'\beta_j - \sigma_{ik'})} \\
&\quad - \left(\frac{\sum_{k'=0}^{m} k' \exp(k'\beta_j - \sigma_{ik'})}{\sum_{k'=0}^{m} \exp(k'\beta_j - \sigma_{ik'})} \right)^2 \\
\frac{\partial \mu_{ij}}{\partial \beta_{j'}} &= 0 \qquad (j' \neq j) \\
\frac{\partial \mu_{ij}}{\partial \delta_{ik}} &= -\frac{\sum_{k'=k}^{m} k' \exp(k'\beta_j - \sigma_{ik'})}{\sum_{k'=0}^{m} \exp(k'\beta_j - \sigma_{ik'})} \\
&\quad + \frac{(\sum_{k'=0}^{m} k' \exp(k'\beta_j - \sigma_{ik'}))(\sum_{k'=k}^{m} \exp(k'\beta_j - \sigma_{ik'}))}{(\sum_{k'=0}^{m} \exp(k'\beta_j - \sigma_{ik'}))^2} \\
\frac{\partial \mu_{ij}}{\partial \delta_{i'k}} &= 0 \qquad (i' \neq i)
\end{aligned} \tag{4.12}$$

Proof:

These follow directly from the rule of derivatives for quotients and the relation $\sigma_{ik} = \sum_{k'=0}^{k} \delta_{ik'}$. Note that only the terms $(k' \geq k)$ contain δ_{ik} in the expression of $\boldsymbol{\mu}_j$. Therefore only the terms $(k' \geq k)$ remain when taking derivatives.

□

The estimating equation (4.11) differs from the GEE in that it does not have a "working" covariance matrix. Instead, $\boldsymbol{\Omega}_j$ is the true covariance matrix.

Let us re-write the estimating equation (4.11) in the following way:

$$\mathbf{U}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta})) = \sum \mathbf{U}_\mathrm{j}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta})) = \mathbf{0}, \tag{4.13}$$

where $\hat{\rho}(\boldsymbol{\theta})$ is a consistent estimator given $\boldsymbol{\theta}$. The estimates $\hat{\boldsymbol{\theta}}$ are solutions to (4.13), and have the following asymptotic properties.

Theorem 4.2 *Under mild regularity conditions, and given that $\hat{\rho}$ is consistent given the true parameter $\boldsymbol{\theta}$, $N^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normally distributed with covariance matrix given by*

$$Var(N^{\frac{1}{2}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})) = lim_{N \to \infty} \left(\frac{\sum_j \mathbf{D}_\mathrm{j}' \boldsymbol{\Omega}_j^{-1} \mathbf{D}_\mathrm{j}}{N} \right)^{-1} \tag{4.14}$$

as $N \to \infty$.

Proof:

Upon expanding $\mathbf{U}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta}))$ around the true value of the parameter $\boldsymbol{\theta}$, we obtain the following:

$$\mathbf{U}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta})) + \frac{d\mathbf{U}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta}))}{d\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{0}$$

By rearranging the terms, $N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ can be expressed as

$$N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = -\left(\frac{1}{N}\frac{d\mathbf{U}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta}))}{d\boldsymbol{\theta}}\right)^{-1}\left(\frac{1}{N^{1/2}}\mathbf{U}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta}))\right)$$

We will show that

$$\frac{1}{N}\frac{d\mathbf{U}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta}))}{d\boldsymbol{\theta}} \xrightarrow{P} -\frac{\sum_j \mathbf{D_j}'\boldsymbol{\Omega}_j^{-1}\mathbf{D_j}}{N}$$

and

$$\frac{1}{N^{1/2}}\mathbf{U}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta})) \sim AN\left(0, \frac{\sum_j \mathbf{D_j}'\boldsymbol{\Omega}_j^{-1}\mathbf{D_j}}{N}\right).$$

Therefore by the Slutsky's Theorem (Cramér, 1946), $N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normally distributed with mean $\mathbf{0}$ and asymptotical covariance matrix given by

$$\left(\frac{\sum_j \mathbf{D_j}'\boldsymbol{\Omega}_j^{-1}\mathbf{D_j}}{N}\right)^{-1}\left(\frac{\sum_j \mathbf{D_j}'\boldsymbol{\Omega}_j^{-1}\mathbf{D_j}}{N}\right)\left(\frac{\sum_j \mathbf{D_j}'\boldsymbol{\Omega}_j^{-1}\mathbf{D_j}}{N}\right)^{\prime -1}$$

$$= \left(\frac{\sum_j \mathbf{D_j}'\boldsymbol{\Omega}_j{}^{-1}\mathbf{D_j}}{N} \right)^{-1}$$

First, we note that

$$\frac{1}{N}\frac{d\mathbf{U}(\boldsymbol{\theta},\hat{\rho}(\boldsymbol{\theta}))}{d\boldsymbol{\theta}} = \frac{1}{N}\left(\frac{\partial \mathbf{U}(\boldsymbol{\theta},\hat{\rho}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{U}(\boldsymbol{\theta},\hat{\rho}(\boldsymbol{\theta}))}{\partial \hat{\rho}(\boldsymbol{\theta})} \cdot \frac{\partial \hat{\rho}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)$$

However, as $N \to \infty$,

$$\begin{aligned} \frac{1}{N}\frac{\partial \mathbf{U}(\boldsymbol{\theta},\hat{\rho}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} &= \frac{1}{N}\frac{\partial \sum_j \mathbf{U_j}(\boldsymbol{\theta},\hat{\rho}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}} \\ &\xrightarrow{P} -\frac{\sum_j \mathbf{D_j}'\boldsymbol{\Omega}_j{}^{-1}\mathbf{D_j}}{N} \end{aligned} \tag{4.15}$$

On the other hand, $\partial\hat{\rho}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = O_p(1)$, and

$$\frac{\partial \mathbf{U}(\boldsymbol{\theta},\hat{\rho}(\boldsymbol{\theta}))}{\partial \hat{\rho}(\boldsymbol{\theta})} = \frac{\partial \sum_j \mathbf{U_j}(\boldsymbol{\theta},\hat{\rho}(\boldsymbol{\theta}))}{\partial \hat{\rho}(\boldsymbol{\theta})} = o_p(N)$$

since $\partial\mathbf{U_j}(\boldsymbol{\theta},\hat{\rho}(\boldsymbol{\theta}))/\partial\hat{\rho}(\boldsymbol{\theta})$ are linear functions of random variables with mean $\mathbf{0}$.

Therefore,

$$\begin{aligned} \frac{1}{N}\frac{d\mathbf{U}(\boldsymbol{\theta},\hat{\rho}(\boldsymbol{\theta}))}{d\boldsymbol{\theta}} &= -\frac{\sum_j \mathbf{D_j}'\boldsymbol{\Omega}_j{}^{-1}\mathbf{D_j}}{N} + \frac{1}{N} \cdot o_p(N) \cdot O_p(1) \\ &\xrightarrow{P} -\frac{\sum_j \mathbf{D_j}'\boldsymbol{\Omega}_j{}^{-1}\mathbf{D_j}}{N}. \end{aligned}$$

Second, for a fixed $\boldsymbol{\theta}$, expansion of $U(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta}))$ around the true value of ρ gives

$$\begin{aligned}
\frac{1}{N^{1/2}}\mathbf{U}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta})) &= \frac{1}{N^{1/2}}\left(\mathbf{U}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta})) + \left.\frac{d\mathbf{U}(\boldsymbol{\theta}, \rho)}{d\rho}\right|_{\rho^*} \cdot (\hat{\rho} - \rho) + o_p(\hat{\rho} - \rho)\right) \\
&= N^{1/2}\frac{1}{N}\sum_j \mathbf{U}_\mathrm{j}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta})) \\
&\quad + \left(\frac{1}{N}\left.\frac{\sum_j d\mathbf{U}_\mathrm{j}(\boldsymbol{\theta}, \rho)}{d\rho}\right|_{\rho^*}\right) \cdot \left(N^{1/2}(\hat{\rho} - \rho)\right) + o_p(1)
\end{aligned} \tag{4.16}$$

for some ρ^* falling between $\hat{\rho}$ and ρ.

Now $\left(N^{1/2}(\hat{\rho} - \rho)\right) = O_p(1)$ under the given conditions stated, and we have

$$\frac{1}{N}\left.\frac{\sum_j d\mathbf{U}_\mathrm{j}(\boldsymbol{\theta}, \rho)}{d\rho}\right|_{\rho^*} = o_p(1)$$

because $(d\mathbf{U}_\mathrm{j}(\boldsymbol{\theta}, \rho)/d\rho)\,|_{\rho^*}$ is a linear function of random variables with mean $\mathbf{0}$, and thus the second term on the right side of (4.16) is also $o_p(1)$.

Therefore, by the Central Limit Theorem, $\frac{1}{N}\sum_j \mathbf{U}_\mathrm{j}(\boldsymbol{\theta}, \hat{\rho}(\boldsymbol{\theta}))$ is asymptotically normal with mean $\mathbf{0}$ and an asymptotical covariance matrix

$$\begin{aligned} \frac{1}{N^2}\sum_j Var(\mathbf{U_j}(\boldsymbol{\theta},\hat{\rho}(\boldsymbol{\theta}))) &= \frac{1}{N^2}\sum_j(\mathbf{D_j}'\boldsymbol{\Omega}_j^{-1}\cdot\boldsymbol{\Omega}_j\cdot(\mathbf{D_j}'\boldsymbol{\Omega}_j^{-1})') \\ &= \frac{1}{N^2}\sum_j(\mathbf{D_j}'\boldsymbol{\Omega}_j^{-1}\mathbf{D_j}) \end{aligned}$$

Now $\frac{1}{N^{1/2}}\mathbf{U}(\boldsymbol{\theta},\hat{\rho}(\boldsymbol{\theta}))$ is also asymptotically normal:

$$\frac{1}{N^{1/2}}\mathbf{U}(\boldsymbol{\theta},\hat{\rho}(\boldsymbol{\theta})) \sim AN\left(0, \frac{\sum_j \mathbf{D_j}'\boldsymbol{\Omega}_j^{-1}\mathbf{D_j}}{N}\right)$$

This completes the proof.

□

The solution to (4.11) may also be obtained iteratively:

$$\hat{\boldsymbol{\theta}}^{(t)} = \hat{\boldsymbol{\theta}}^{(t-1)} + \left[\sum_{j=1}^{N}\mathbf{D_j}'\boldsymbol{\Omega}_j^{-1}\mathbf{D_j}\right]^{-1}\left[\sum_{j=1}^{N}\mathbf{D_j}'\boldsymbol{\Omega}_j^{-1}(\mathbf{x_j}-\boldsymbol{\mu}_j)\right]\Bigg|_{\hat{\boldsymbol{\theta}}^{(t-1)}}, \quad t=1,\ldots,T. \tag{4.17}$$

at some T.

4.5.3 Two-Step Iteration Method

Given some initial values, we can use (4.9) and (4.17) alternatively to get the estimates of $\boldsymbol{\theta}$ and ρ. That is, at each step of the iteration, we solve for (4.9) to obtain an estimator of ρ, and then given the updated value of ρ, solve for (4.17) to obtain an estimator of $\boldsymbol{\theta}$. We repeat this procedure until we reach convergence. Since, given any initial consistent estimators, the estimators obtained from (4.9) or (4.17), are consistent after the first step (Lehmann and Casella, 1998), the final estimators from this two-step iteration method (Olsson, 1979) after the last step T are also consistent and asymptotically normal. This method also allows us to compute a consistent estimator of its asymptotic variance matrix. The fact that in each of the two steps a portion of the parameters $\boldsymbol{\theta}_\rho$ is replaced by its consistent estimator implies some loss of efficiency. However, this loss is minor and negligible (Olsson, 1979).

Therefore, estimator $\hat{\rho}_{i,i'}^{(T)}$ obtained in (4.9) after the last iteration at step T is asymptotically normal with mean ρ and variance $Var(\hat{\rho}_{i,i'})$, which can be estimated by

$$\widehat{Var(\hat{\rho}_{i,i'})} = \left(\frac{\partial^2 l_\rho(\boldsymbol{\theta})}{\partial \rho^2}\right)^{-1}\Bigg|_{\hat{\rho}_{i,i'}^{(T)}}$$

where the right-hand side is evaluated at $\hat{\rho}_{i,i'}^{(T)}$ after the last iteration step T. Also $\hat{\boldsymbol{\theta}}^{(T)}$ finally obtained in (4.17) is asymptotically normal with mean $\boldsymbol{\theta}$ and variance matrix $Var(\hat{\boldsymbol{\theta}})$, which can be estimated by

$$\widehat{Var(\hat{\boldsymbol{\theta}})} = \left(\sum_{j=1}^{N} \mathbf{D_j}'\boldsymbol{\Omega}_j^{-1}\mathbf{D_j}\right)^{-1}\Bigg|_{\hat{\boldsymbol{\theta}}^{(T)}}$$

when the right hand side is evaluated at $\hat{\boldsymbol{\theta}}^{(T)}$ after the last iteration step T.

4.6 Efficiency Considerations

We now show the performance of our strategy in a computer simulation. Similarly as in the previous chapter, we replicated the simulation 1000 times to obtain the empirical distribution of the CML estimators.

4.6.1 Dichotomous Rasch Model

In this case, the response data are binary ($m = 1$). First, we generated the latent variable for $N = 100$ subjects for $I = 4$ items from a multivariate normal distribution. The correlation parameter ρ is chosen to be 0, 0.3, and 0.7 respectively. Then the ordinal response data are obtained according to the latent traits as generated, with a given distinct set of parameters for the ability of persons

$$\boldsymbol{\beta}' = (0, 0.25, 0.5, 0.75, 1)$$

and for the difficulty of items

$$\boldsymbol{\delta} = \begin{pmatrix} 0 & -0.3 \\ 0 & -0.1 \\ 0 & 0.1 \\ 0 & 0.3 \end{pmatrix}$$

with $\mathbf{1}'\boldsymbol{\delta}\mathbf{1} = 0$ for identifiability (see Wright and Masters, 1982).

Figure 4.1 gives the Q-Q plots of $\hat{\rho}$ when the true polychoric correlation coefficient is $0, 0.3, 0.7$ respectively. We can see that the estimators closely follow a normal distribution. *Table 4.1* reports the performance of our estimators of ρ in simulation. Both bias and variance increase with increasing values of ρ.

Figures 4.2–4.4 show the behavior of the estimates of $\boldsymbol{\beta}$ for various correlation settings (ρ=0, 0.3, and 0.7 respectively). From the Q-Q plots, we can see that all closely follow a normal distribution, with a slight departure for some extreme values. The smaller the polychoric correlation, the smaller the variance of the estimators. We can also see that the estimators of middle values of β (β=0.5) follow the normal distribution more closely than those of the two end points ($\beta = 0.25$, and 0.75).

Table 4.2 shows the relative efficiencies of the estimators for $\boldsymbol{\beta}$ with inter-item correlation over those assuming independence between items. When the true correlation is moderate, for example $\rho = .3$, there is little improvement in efficiency between the estimators obtained by recognizing the correlation among the items and those obtained under the independence assumption. The relative efficiencies are very close to 1. However, when the correlation is high, for example, $\rho = 0.7$, estimators based on recognizing the correlation gain a significant efficiency as compared to the traditional Rasch model, which assumes independence. The relative efficiencies could be as low as 0.75, and at most 0.83.

The result of $\boldsymbol{\delta}$ also gives a similar conclusion. Therefore, for the sake of brevity we have omitted the presentation of results on $\boldsymbol{\delta}$.

4.6.2 Polytomous Rasch Model

We also considered the polytomous response data. As in the previous simulation, the latent variables were generated for $N = 100$ subjects for $I = 4$ items from a multivariate normal distribution. The correlation parameter ρ are again chosen to be 0, 0.3, and 0.7 respectively. Then the ordinal response data are obtained according to the latent trait generated, with a given distinct set of parameters for the ability of persons

$$\boldsymbol{\beta}' = (0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2)$$

and the difficulty of items

$$\boldsymbol{\delta} = \begin{pmatrix} 0 & -0.2962 & 0.2185 \\ 0 & -0.4270 & 0.4717 \\ 0 & -0.3885 & 0.6263 \\ 0 & -0.3852 & 0.1803 \end{pmatrix}$$

with $\mathbf{1}'\boldsymbol{\delta}\mathbf{1} = 0$ for identifiability (see Wright and Masters, 1982).

Figure 4.5 gives the Q-Q plots of $\hat{\rho}$ in a polytomous Rasch Model when the true polychoric correlation coefficients are 0, 0.3, 0.7 respectively. We can see that the estimates follow very closely the assumed normal distribution. *Table 4.3* shows that the bias in estimation in the polytomous model is reduced, but the variance is nearly tripled, and it increases with an increasing level of ρ.

Figures 4.6–4.8 give the normal Q-Q plots of selected estimates of $\boldsymbol{\beta}$ (two extreme values of β=0.25, 1.75, and one middle value of β=1) for various correlation settings (ρ=0, 0.3, and 0.7 respectively). From the Q-Q plots, again we can see that the estimators follow normal distributions, with slight departure at extreme values. Smaller polychoric correlation leads to estimators with smaller variances. The variances of the estimators of β tend to be larger than those from the corresponding dichotomous case. Similar to what we observed in the dichotomous case, estimators of middle values of parameters (around 1) are better than those of the end points (β close 0 or 2).

Table 4.4 shows the relative efficiencies of the estimators for $\boldsymbol{\beta}$, adjusting for the inter-item correlation as compared to the traditional Rasch model that assumes independence between items. We observe that the results have a similar pattern to the dichotomous case. When the true correlation is moderate, for example $\rho = .3$, there is little improvement in efficiency between the estimators obtained by recognizing the correlation among the items and those obtained under the independence assumption. The relative efficiencies are close to 1 for all situations. However, when the correlation is high, for example, $\rho = 0.7$, the improvement in efficiency for recognizing the dependency becomes noticeable as compared to that under the assumption of independence. The highest relative efficiency was only about 0.83, as compared to the traditional Rasch model approach.

4.7 Summary

We have proposed a latent variable approach to the Rasch model in this chapter. Using the generalized estimating equations method, we developed an estimation method for the Rasch model parameters under item-to-item correlations. A simulation study has shown the relative efficiency of the estimators when inter-item correlation is considered. Generally, we observed normally distributed estimators for the parameters, and the efficiency loss of the estimates increased, as the level of polychoric correlation becomes high.

This method depends heavily on the performance of computing techniques. It may take a long time for the iteration procedure to converge, and sometimes it even diverges. Better numerical methods are needed to increase efficiency of the estimation. Further, a more general correlation pattern might be necessary to include unequal correlations between different item pairs.

Figure 4.1: Q-Q plots of $\hat{\rho}$ vs standard normal quantiles in the dichotomous model ($I = 4, m = 1, N = 100$)

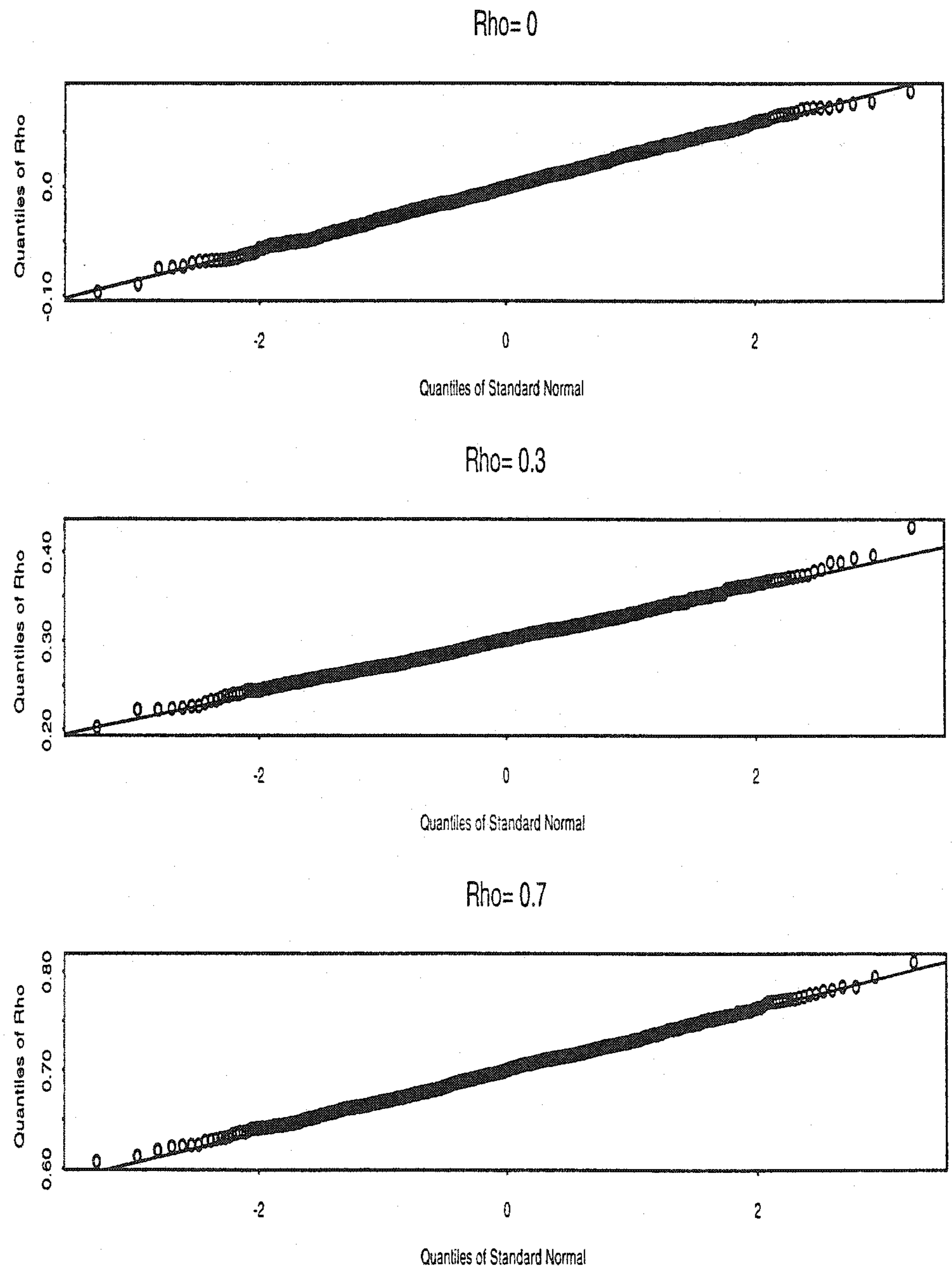


Figure 4.2: Q-Q plots of $\hat{\beta}$ vs standard normal quantiles in the dichotomous model ($\rho = 0, I = 4, m = 1, N = 100$)

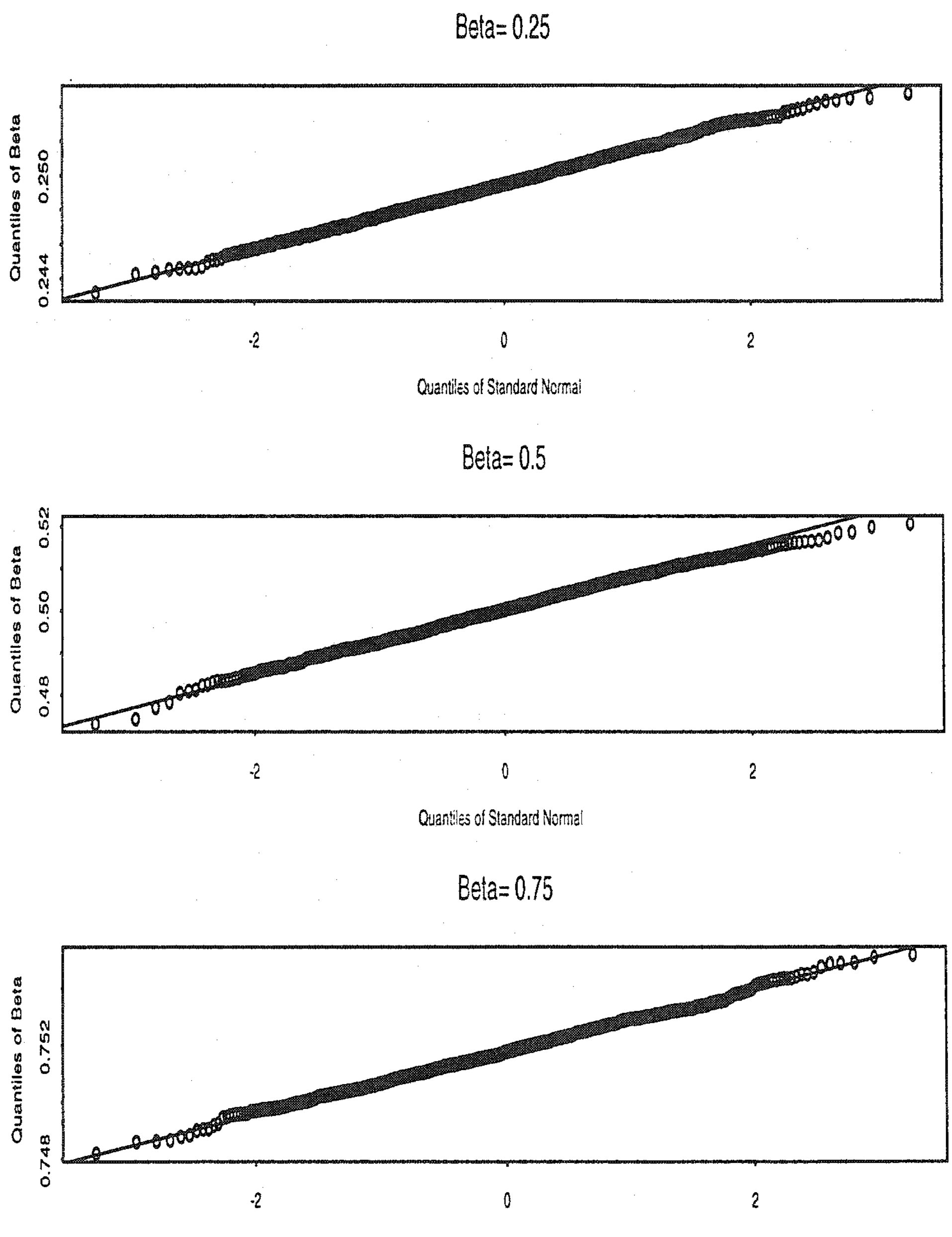


Figure 4.3: Q-Q plots of $\hat{\beta}$ vs standard normal quantiles in the dichotomous model ($\rho = 0.3, I = 4, m = 1, N = 100$)

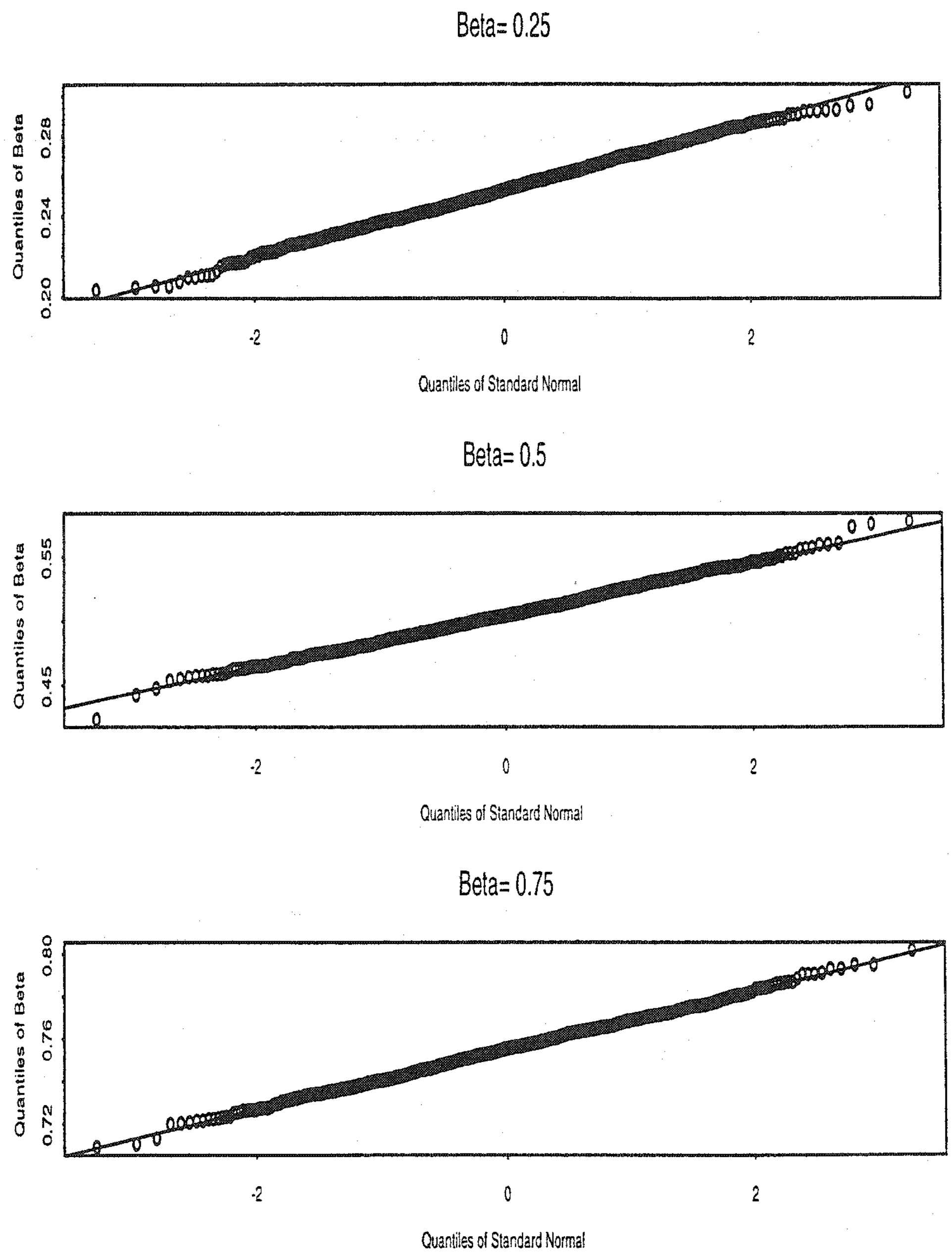


Figure 4.4: Q-Q plots of $\hat{\beta}$ vs standard normal quantiles in the dichotomous model ($\rho = 0.7, I = 4, m = 1, N = 100$)

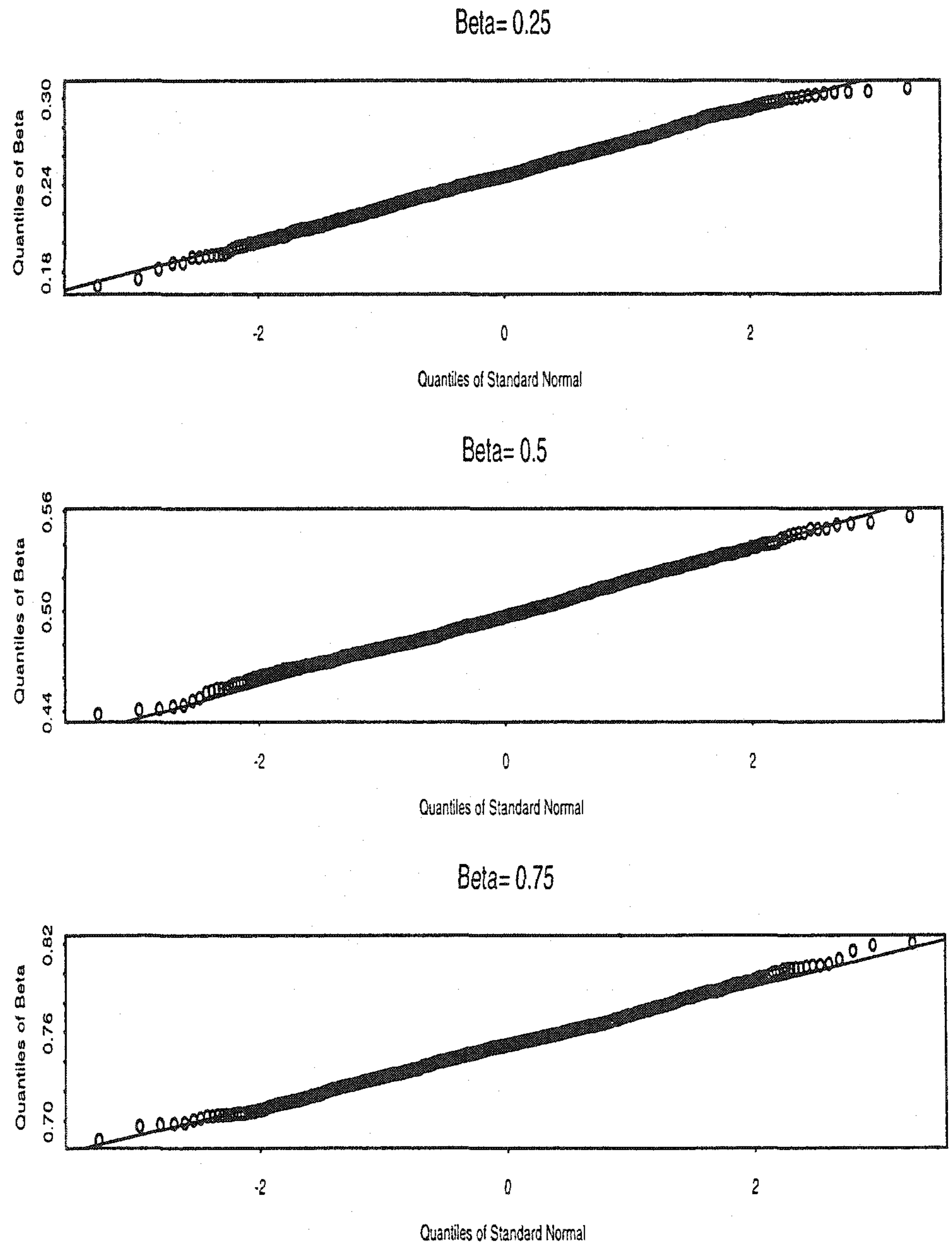


Figure 4.5: Q-Q plots of $\hat{\rho}$ vs standard normal quantiles in the polytomous model ($I = 4, m = 2, N = 100$)

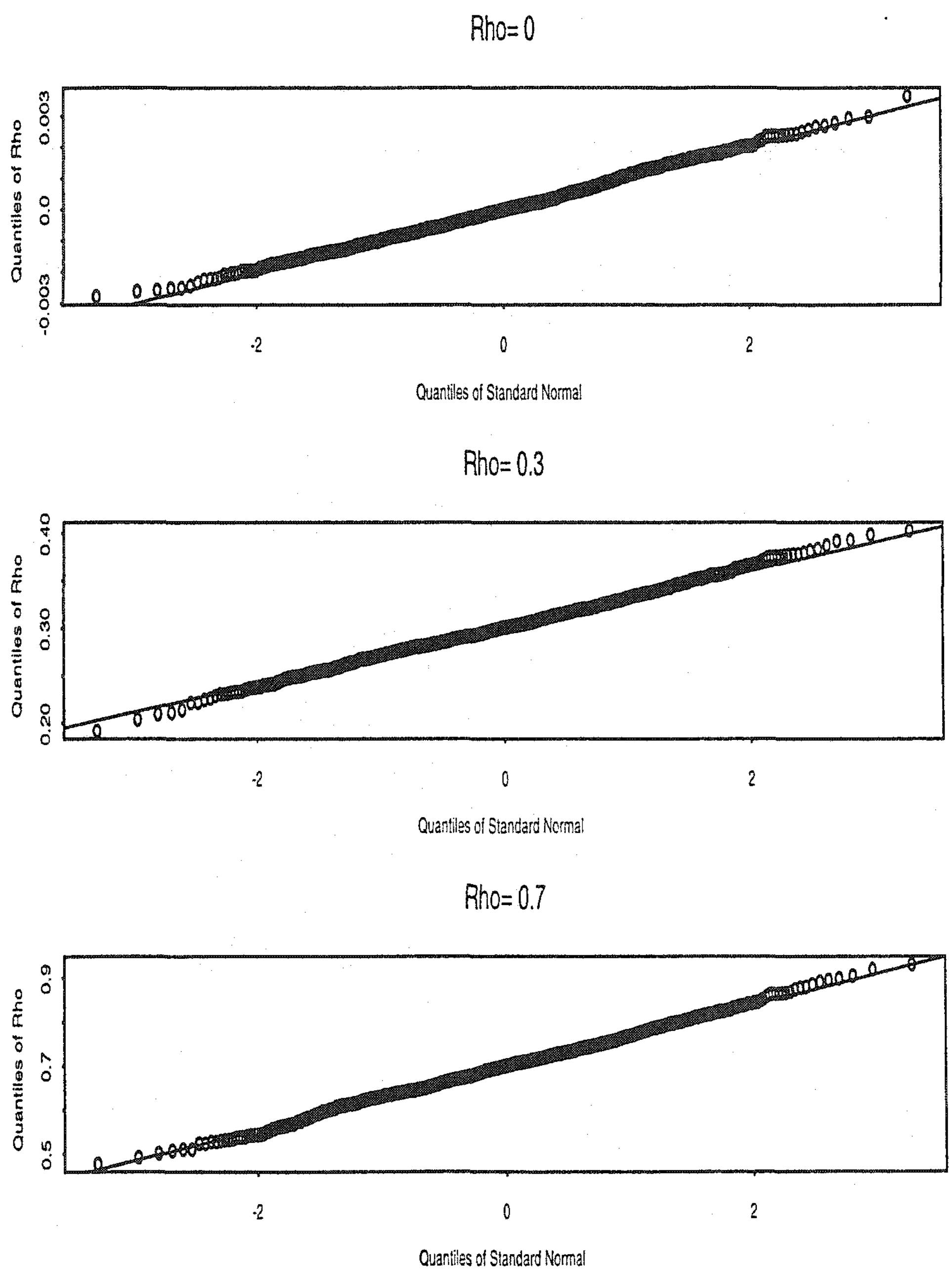


Figure 4.6: Q-Q plots of selected $\hat{\beta}$ vs standard normal quantiles in the polytomous model ($\rho = 0, I = 4, m = 2, N = 100$)

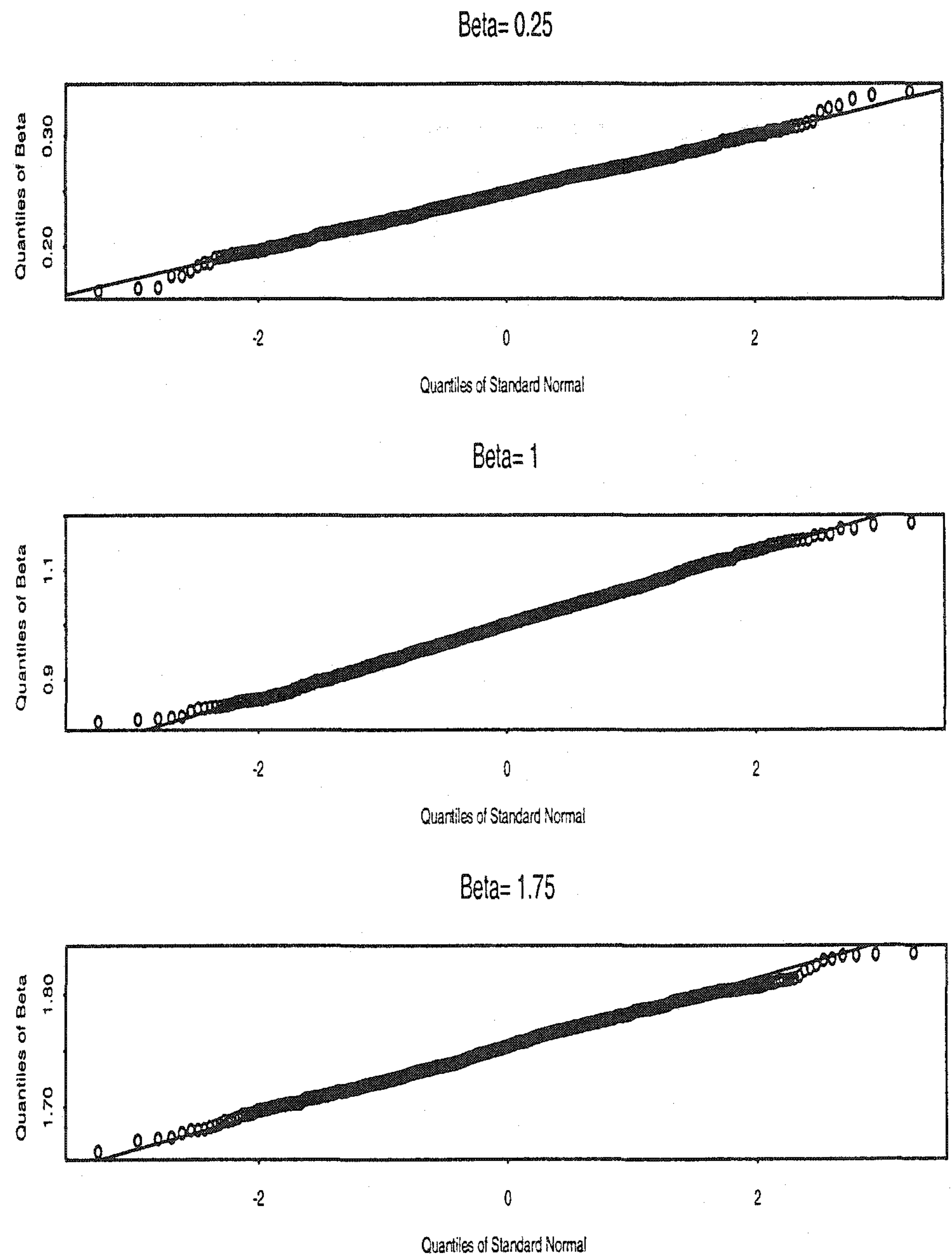


Figure 4.7: Q-Q plots of selected $\hat{\beta}$ vs standard normal quantiles in the polytomous model ($\rho = 0.3, I = 4, m = 2, N = 100$)

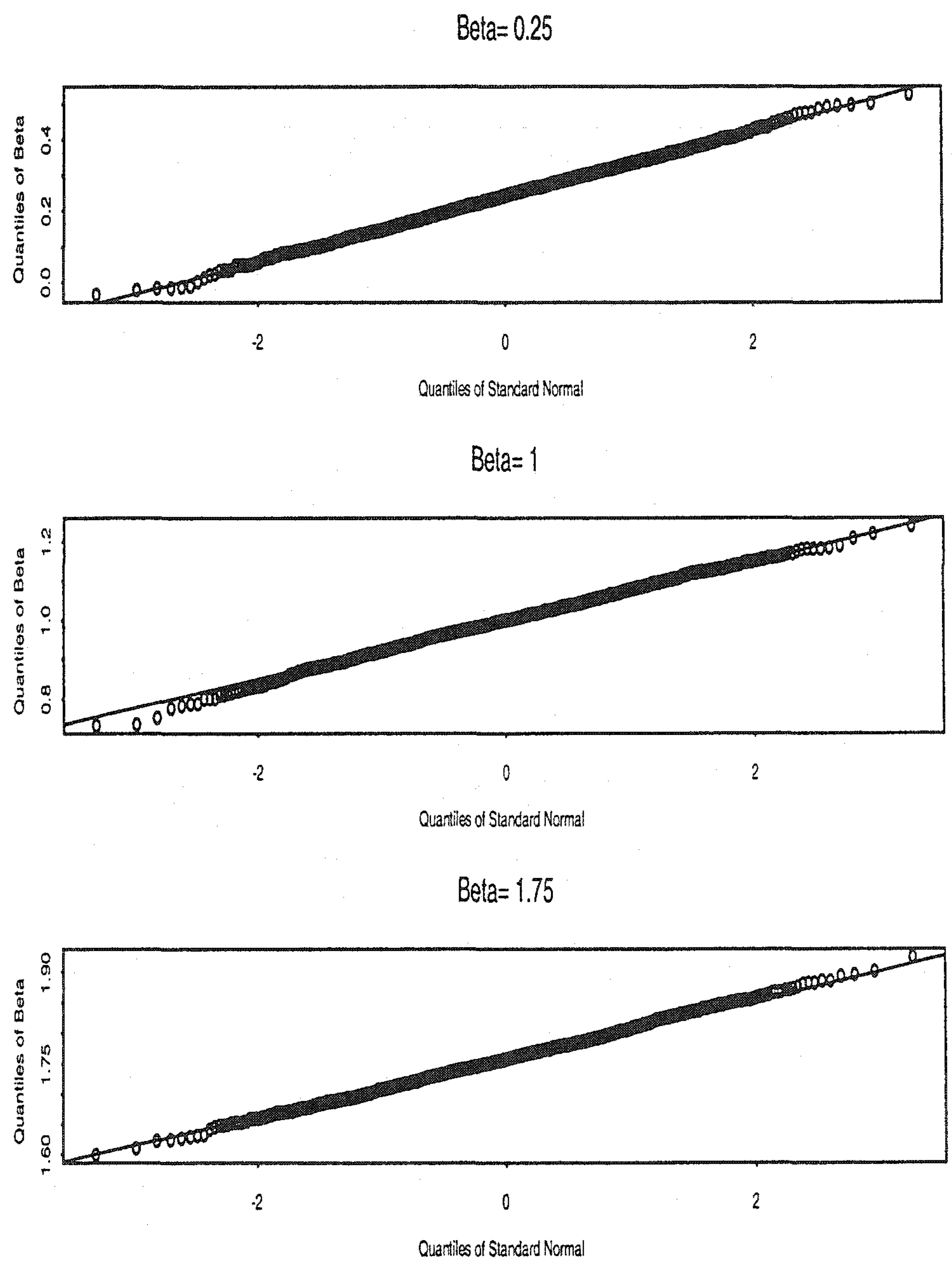


Figure 4.8: Q-Q plots of selected $\hat{\beta}$ vs standard normal quantiles in the polytomous model ($\rho = 0.7, I = 4, m = 2, N = 100$)

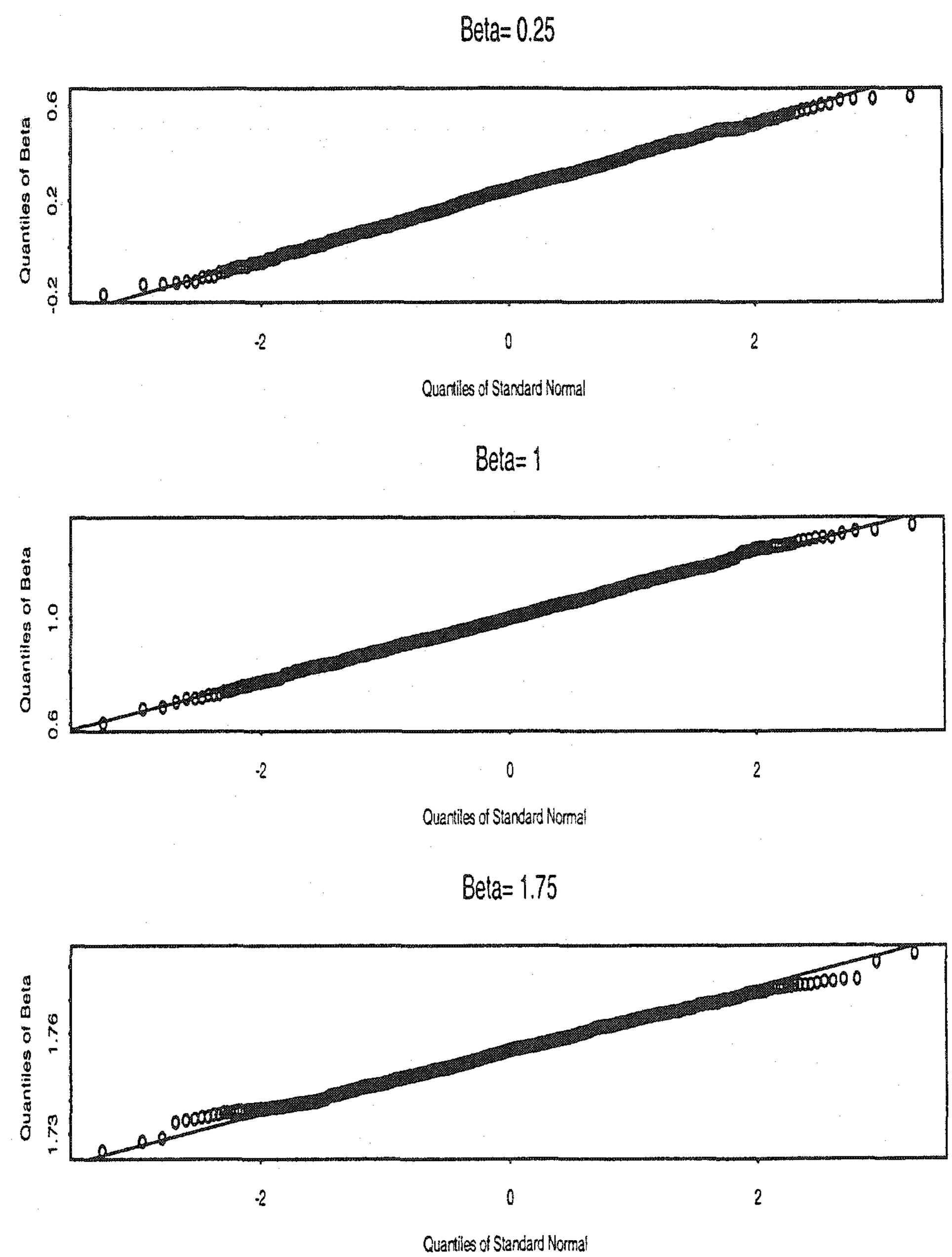


Table 4.1: Summary of estimates of ρ in the dichotomous model ($I = 4, m = 1, N = 100$)

ρ	$Mean(\hat{\rho})$	$\widehat{Var(\hat{\rho})}$	$\widehat{MSE(\hat{\rho})}$
0.0	0.0289	0.0096	0.0104
0.3	0.2516	0.0178	0.0201
0.7	0.7511	0.0301	0.0327

Note: The entries show the average, variance, and MSE of the estimated $\hat{\rho}$ in 1000 simulations.

Table 4.2: Relative efficiencies of the estimators for β in the dichotomous model ($I = 4, m = 1, N = 100$)

β	ρ=0	ρ=0.3	ρ=0.7
0.25	1.000	0.9763	0.8250
0.50	1.000	0.9847	0.7502
0.75	1.000	0.9755	0.7958
overall	1.000	0.9794	0.7994

Note: The entries are relative efficiencies of the estimators for β, based on empirical distributions, and adjusted by the inter-item correlation compared to models assuming the independence of items. The row labelled 'overall' corresponds to those of the trace of the variance matrix. Ratios of less than 1 indicate improvement produced by taking inter-item correlation into consideration.

Table 4.3: Summary of estimates of ρ in the polytomous model ($I = 4, m = 2, N = 100$)

ρ	$Mean(\hat{\rho})$	$\widehat{Var}(\hat{\rho})$	$\widehat{MSE}(\hat{\rho})$
0.0	0.0203	0.0397	0.0402
0.3	0.3242	0.0740	0.0746
0.7	0.7309	0.1185	0.1195

Note: The entries show the average, variance, and MSE of the estimated $\hat{\rho}$ in 1000 simulations.

Table 4.4: Relative efficiencies of the estimators for β in the polytomous model ($I = 4, m = 2, N = 100$)

β	ρ=0	ρ=0.3	ρ=0.7
0.25	1.000	0.9459	0.7455
0.50	1.000	0.9635	0.7196
0.75	1.000	0.9545	0.8259
1.00	1.000	0.9293	0.7403
1.25	1.000	0.9777	0.7993
1.50	1.000	0.9329	0.7587
1.75	1.000	0.9798	0.8149
overall	1.000	0.9687	0.7769

Note: The entries are relative efficiencies of the estimators for β, based on empirical distributions, and adjusted by the inter-item correlation compared to models assuming the independence of items. The row labelled 'overall' corresponds to those of the trace of the variance matrix. Ratios of less than 1 indicate improvement produced by taking inter-item correlation into consideration.

Chapter 5

Analysis of Missing Item Response Data

5.1 Introduction

In the item response data, there often exist missing responses in some of the items. This could be caused by many reasons. For example, patients may be unable to respond to some questions in clinical trial settings because of sickness, unconsciousness, or other reasons. In a household survey, one could refuse to answer certain sensitive questions, such as questions related to income or family violence for privacy or confidential reasons.

Some missing data are ignorable, while others are not. As reviewed in

Chapter 2, ignoring the missing observations and missing data mechanisms could lead to serious estimation bias. Little has been done to date about missing data problems in item response theory, especially in connection with the Rasch models, except in some special situations. De Gruijter (1988) dealt with missing data by tailoring test items in a Rasch model. Patz and Junker (1999) described a general Markov Chain Monte Carlo strategy for Bayesian inference in complex item response theory settings, which addresses the non-response issue. Huisman and Molenaar (2001) compared several imputation techniques of missing scale data in item response models when the pattern and the reason for the non-response data are ignorable.

However, these research are rather limited in that either they only deal with specific situations, or they use techniques that are applicable only when certain conditions are satisfied. Further, no work has been done to resolve general missing data problems, arisen under various missing data mechanisms. In particular, we are interested in the implications of various missing data strategies, when applied to the Rasch models based on item response data.

In this chapter, we discuss the bootstrap technique applied to the Rasch models when missing data are present. We focus on the method of bootstrap under imputation. Large sample behavior of the bootstrap estimators is established and compared to those of other methods including simple imputation methods, as described in *Chapter 2*.

5.2 Bootstrap for Imputed Data

Imputation for missing data is widely used for statistical inference about the population characteristic. Consider a simple example: suppose y_i is the response variable in a sample $\mathcal{S}$. Let $\mathbf{y}_R = \{y_i : i \in \mathcal{S}_{\mathcal{R}}\}$, and $\mathbf{y}_M = \{y_i : i \in \mathcal{S}_M\}$, where $\mathcal{S}_R$ and $\mathcal{S}_M$ are subsets of $\mathcal{S}$ corresponding to respondents with full observations and non-respondents with missing observations. In case of no missing values, $\hat{\theta} = \hat{\theta}(\mathbf{y})$, a function of $\mathbf{y}$, is used to estimate an unknown population parameter θ. When some data are missing, we often use $\mathbf{y}_R$ to obtain imputed values η_i for $i \in \mathcal{S}_M$ and then treat these imputed data as if they were true observations, and use $\hat{\theta}_I = \hat{\theta}(\mathbf{y}_I)$ to estimate θ where $\mathbf{y}_I = \{y_i : i \in \mathcal{S}_R\} \cup \{\eta_i : i \in \mathcal{S}_M\}$.

If the imputation method is well defined and true to the actual situation, then the estimator $\hat{\theta}_I$ is asymptotically valid, although not as efficient as $\hat{\theta}$. However, in general such cannot be assumed and treating the imputed values as if they were true observations could lead to serious underestimation of the variance of $\hat{\theta}_I$ as well as biased estimation of θ, especially when the proportion of missing data is rather large (Rubin, 1978).

Bootstrap (Efron, 1979, 1994) is an extremely useful tool of obtaining the sampling properties of random variables. In the case of no missing values, Efron's bootstrap method can be described as follows. Let P be the statistical model that generates the data $\mathbf{y}$ and let $\hat{P}$ be an estimate of P based on the generated data $\mathbf{y}$. Let $\mathbf{y}^*$ be the bootstrap sample generated from $\hat{P}$. The spirit of the bootstrap method, summarized in *Figure 6.1* (Efron and Tibshirani, 1986), is to mimic the sampling behavior of (P, $\mathbf{y}$, $\hat{\theta}$) by using the conditional (given $\mathbf{y}$) sampling behavior of ($\hat{P}$, $\mathbf{y}^*$, $\hat{\theta}^*$), where $\hat{\theta}^* = \hat{\theta}(\mathbf{y}^*)$ is the bootstrap analog of $\hat{\theta}$.

In bootstrap analysis, we are mainly interested in the variance and the distribution of the bootstrap estimator $\hat{\theta}^*$. According to Efron and Tibshirani (1986), the bootstrap estimator of $Var(\hat{\theta})$ is

$$v_B(\mathbf{y}) = Var^*(\hat{\theta}^*) \tag{5.1}$$

where Var^* is the conditional variance with respect to $\mathbf{y}^*$, given $\mathbf{y}$. Also the bootstrap estimator of the distribution of $\hat{\theta} - \theta$, denoted by H_B, is

$$H_B(\mathbf{y}) = H^*_{\hat{\theta}^* - \hat{\theta}} \tag{5.2}$$

where H^* is the conditional distribution, given $\mathbf{y}$.

If $v_B(\mathbf{y})$ or $H_B(\mathbf{y})$ has no explicit form, then one may use the Monte Carlo approximation

$$v_B(\mathbf{y}) \approx \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_b^* - \bar{\theta}^*)^2 \tag{5.3}$$

and

$$H_B(\mathbf{y}) \approx \frac{1}{B}\sum_{b=1}^{B}\delta_{\hat{\theta}_b^* - \hat{\theta}} \tag{5.4}$$

with $\hat{\theta}_b^* = \hat{\theta}(\mathbf{y}_b^*)$, where $\mathbf{y}_b^*$, $b = 1, \ldots, B$, are independent bootstrap data sets generated from $\hat{P}$, $\bar{\theta}^* = \sum_{b=1}^{B}\hat{\theta}_b^*/B$, and δ_x is the distribution degenerated at point x. See Efron and Tibshirani (1986) for details.

5.3 Bootstrap under Imputation

When imputation is used for missing values, naive bootstrap estimators are obtained by applying the standard bootstrap formulas (5.1)–(5.2) and treating $\mathbf{y}_I$ as $\mathbf{y}$. However, $v_B(\mathbf{y}_I)$ does not capture the inflation in variance due to imputation and/or missing data and may lead to serious underestimation, and therefore as a result, both $v_B(\mathbf{y}_I)$ and $H_B(\mathbf{y}_I)$ can be inconsistent (Rubin, 1978). Ideally, the bootstrap data sets should also be imputed in the same way as the original data set was imputed (Shao and Sitter, 1996). The procedure is as follows:

1. Draw a random sample $\mathbf{y}^* = \{y_i^* : i = 1, \ldots, n\}$ with replacement from

the original data set $\mathbf{y}$, according to the sampling scheme to obtain $\mathbf{y}$.

2. Apply the same imputation procedure used in constructing $\mathbf{y}_I$ to the units in the bootstrap sample $\mathbf{y}^*$, to form $\mathbf{y}_I^*$, the bootstrap analog of $\mathbf{y}_I$.

3. Obtain the bootstrap analog $\hat{\theta}_I^*$ of $\hat{\theta}_I$, and apply standard bootstrap formulas or their Monte Carlo approximations to obtain the variance and distribution for the bootstrap estimators for $\hat{\theta}_I^*$, based on the imputed bootstrap data set $\mathbf{y}_I^*$.

According to Shao and Sitter (1996), this method is the only method thus far that works irrespective of the sampling design (single stage or multistage, simple random sampling or stratified sampling), the imputation method (random or non-random, proper or improper), or the type of $\hat{\theta}$ (smooth or non-smooth). However, whether this superiority would hold under any missing data mechanisms is not yet known.

5.4 Bootstrap under Imputation: Implementation in Rasch Models

We now consider the bootstrap procedure under imputation to ordinal item response data with missing values in the context of the Rasch Model analysis.

We consider the data structure $\mathbf{X}$ introduced in *Chapter 2*. Assume that the data set carries identification flags W_{ij} indicating whether X_{ij} is a respondent (W_{ij}=0) or not (W_{ij}=1).

For the imputation method, we choose the hot deck imputation method, which imputes missing values for $\mathbf{X}_M$ with a random sample from $\mathbf{X}_R$. Using such a random imputation method may introduce more variation in estimating the population characteristics. However this increment in variation is relatively small when Monte Carlo method is used, compared to the variation caused from randomness of choosing a bootstrap sample (Shao and Sitter, 1996). The disadvantage of deterministic imputation techniques such as ratio or regression imputation methods lies in that they may not preserve the distribution of the data (Shao and Tu, 1995). Therefore they are not ideal for situations when the parameter of interest is a function of the population distribution, for example, the population median.

The bootstrap procedure is as follows, taking into account the imputation effect, applying the method described in the previous section.

1. Draw a random sample $\mathbf{X}^* = \{\mathbf{x}_1^*, \ldots, \mathbf{x}_N^*\}$ with replacement from the original data set $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, according to the sampling scheme to obtain $\mathbf{X}$.

2. Use the identification flags $\mathbf{W}$ to obtain respondents and non-respondents

in $\mathbf{X}^*$. Let $\mathbf{X}^*_R = \{X^*_{ij} : W_{ij} = 1\}$ and $\mathbf{X}^*_M = \{X^*_{ij} : W_{ij} = 0\}$ denote the set of respondents and non-respondents in the bootstrap sample.

3. Applying the hot deck imputation procedure used in constructing $\mathbf{X}_I$ to impute $\mathbf{X}^*_M$ using $\mathbf{X}^*_R$, i.e., for each item i, $i = 1, \ldots, I$, draw a random sample from $\mathbf{x_i}^*_R$, the set of respondents, to fit in $\mathbf{x_i}^*_M$, the set of respondents for item i. Denote the bootstrap analog of $\mathbf{X}_I$ by $\mathbf{X}^*_I$.

4. Obtain the bootstrap analog $\hat{\delta}^*_I$ of $\hat{\delta}_I$, and the bootstrap analog $\hat{\beta}^*_I$ of $\hat{\beta}_I$, using the improved CML method introduced in *Chapter 3*, based on the imputed bootstrap data set $\mathbf{X}^*_I$.

5. Apply the bootstrap formulas (5.1)–(5.2) to obtain the variance and distribution of the bootstrap estimators $\hat{\delta}^*_I$ and $\hat{\beta}^*_I$, using $\hat{\delta}^*_I$ in place of $\hat{\delta}^*$ and $\hat{\beta}^*_I$ in place of $\hat{\beta}^*$.

6. If necessary, repeat steps 1–5, and use Monte Carlo approximations as given in (5.3)–(5.4).

It follows from Shao and Sitter (1996) that the bootstrap estimators described above are consistent as far as the imputation method provides consistent estimators for the population parameter.

5.5 Simulation Results

In this section, we evaluate the performance of our implementation of missing data analysis strategy for item response data under Rasch models in computer simulation. We generate missing data patterns using the three types: data missing completely at random, data missing at random, and data missing with nonignorable reasons, as descried in *Chapter 2*. In each situation, two missing proportions are used: 20% and 50%. We will compare the performance of estimators obtained by means of four different approaches: using the complete data as generated initially (no missing values), using the complete data subset (removing subjects with missing values), using the standard hot deck imputation method, and using the bootstrap method under the hot deck imputation.

5.5.1 Case 1: Data Missing Completely at Random

When some data are missing completely at random, we can use a random number generator, independent of the process that generated the data, to randomly choose a portion (20% and 50%) of the observations that will be assumed missing.

For the dichotomous case, we use the same given distinct set of parameters for the ability of persons

$$\beta' = (0, 0.25, 0.5, 0.75, 1)$$

and the difficulty of items

$$\delta = \begin{pmatrix} 0 & -0.3 \\ 0 & -0.1 \\ 0 & 0.1 \\ 0 & 0.3 \end{pmatrix}$$

as in the previous chapters. For details on the simulated data, please see *Chapter 3*.

Table 5.1 gives the results for β of the three methods when the missing proportion is 20%. We can see that the bootstrap technique and the imputation method produce almost identical results to that using the complete original data with no missing values. There were also no discernible differences between the bootstrap result and that using complete data. On the other hand, removing subjects with missing data and analyzing only the complete subset data reduces the accuracy of the estimates and increases their variances. Results for δ again give similar conclusion, and will be omitted in this presentation.

Table 5.2 gives the results for β when the missing proportion is 50% to evaluate the impact of an increased missing data proportion. We can see that despite for the fact that the proportion of missing data increases, the bootstrap technique still produces results that are nearly identical to those using the complete original data. The standard hot deck imputation method also gives a similar result, but the efficiency is not as good as that using the bootstrap method. This indicates that the bootstrap method combined with the hot deck imputation is the best strategy, especially in situations when the proportion of missing data is large and the data are missing completely at random. The analysis based on the complete subset data was the the least efficient method of analysis, similarly as before and as others have demonstrated (Carrière, 1994, 1999).

For the polytomous case, we use the set of true parameters as follows: the distinct set of parameters for the ability of persons is

$$\beta' = (0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2)$$

and the difficulty of items is

$$\delta = \begin{pmatrix} 0 & -0.2962 & 0.2185 \\ 0 & -0.4270 & 0.4717 \\ 0 & -0.3885 & 0.6263 \\ 0 & -0.3852 & 0.1803 \end{pmatrix}$$

Table 5.3 and *Table 5.4* give the results for β when the missing proportions are 20% and 50%, respectively. We obtain similar conclusions with the polytomous case as in the dichotomous case in these two situations. That is, the bootstrap method under imputation outperforms that of the standard imputation and the analysis based on the complete subset data is not efficient and suffers from large bias in estimation. One notable finding is that bias of the estimators are quite small in the polytomous case, especially when the missing data proportion is moderate at 20%.

5.5.2 Case 2: Data Missing at Random

In this section, we evaluate the estimation strategy when some data are missing at random. Here the variables where data are missing are not the cause of the incomplete data. Instead, the cause of the missing data is due to some other external influences. We generate the "external" influence in the simulation in the following manner. We let the first several items in the data to be the "external" influence, and remaining items are simulated to be missing or non-missing according to the outcome of the generated external influence. To see this as a reasonable generation of data missing at random, consider a health survey. Patients with severe throat problems may not be able to speak. Then the speech ability of this patient cannot be measured, and therefore the response to that item is missing, although the response to the throat problem is available.

In our study, we simulate this missing data pattern according to the raw score of the first half items. If they are some extremely large or small. then the outcomes of remaining items are set to be missing; the cut-off points are determined by the lower $100(\alpha/2)\%$ and upper $100(1-\alpha/2)\%$ quantiles with the missing proportion set to α. For the two situations with low (20%) and high (50%) missing proportion, the quantiles are (0.1, 0.9) and (0.25, 0.75) respectively. In this simulation, we only consider the polytomous Rasch Model, because in a dichotomous model, using the above missing mechanism generates either too many or too few missing observations because of rounding error. The same set of true parameter values as in the previous section is used.

Table 5.5 and *Table 5.6* give the results for β when the missing proportions are 20% and 50%, respectively. From the tables, we see that bias is small no matter which method is used. Bootstrap method generally gives smaller MSE than the standard imputation method does, except for a few extreme cases when both perform similarly. However, as the missing proportion increases, the bootstrap method becomes much better (in terms of MSE, since bias are small in all situations) than the other two methods.

5.5.3 Case 3: Nonignorable Missing Data

Missing data can arise due to the nature of questions being asked or the treatment being administrated. In other words, the data are missing with reasons that can be explained but are unmeasurable. In our simulation, we generate such situations by setting a fraction of observations to be missing, if the raw score of a subject is larger than $100(1-\alpha/2)\%$ quantile or less than $100(\alpha/2)\%$ quantile with a missing proportion α. This is tenable in reality for persons in extreme situations not to participate in the study any further. For example, persons with a very low or high gross income may refrain from reporting their actual income. Similar scenarios can be constructed in a family violence survey. Again we focus on the polytomous Rasch Model only. We use the same set of true parameters as in the previous sections. The results are summarized in *Table 5.7.*

From *Table 5.7*, we see the clear advantage of the bootstrap method in this situation. Even when the missing proportion is only 20%, this method gives much more efficient estimates than those given by the standard imputation method alone. The method that uses only the complete subset data is the worst of all. It does not give any consistent estimates, producing large bias overall.

When the missing proportion is high at 50%, none of the methods gives reasonably consistent estimates of the parameters. This may be explained by the nonignorable missing data mechanism with too many data missing to draw on information independently in our simulation. In practice, however, there may be more information available to draw on about the missing data that are nonignorable. Further, it is highly unusual to have over 50% of the data missing with nonignorable reasons. Therefore, our overall conclusion is that the bootstrap under imputation method is the best strategy.

5.6 Summary

In this chapter, we investigated several missing data analytic strategies in a Rasch model with item response data. Among several imputation and re-sampling methods used for missing data analysis, the bootstrap under imputation method is found to be the best method overall in its accuracy and precision of estimating the true parameters. This was true whether the missing data occurred completely at random, at random, or with nonignorable reasons.

Figure 5.1: Bootstrap method (Efron and Tibshirani, 1986)

$$\begin{array}{ccccccc} P & \longrightarrow & \mathbf{y} & \Longrightarrow & \hat{P} & \longrightarrow & \mathbf{y}^* \\ \searrow & & \swarrow & & \searrow & & \swarrow \\ & \hat{\theta}(\mathbf{y}) & & & & \hat{\theta}(\mathbf{y}^*) & \end{array}$$

Table 5.1: Comparison of estimates of β when 20% of the data is missing completely at random in the dichotomous model ($I = 4, m = 1, N = 100$)

β	Method	$\hat{\beta}$	Bias (%)	$\widehat{MSE(\hat{\beta})}$
0.25	Complete	0.2460	-1.60	0.00152
	Complete subset	0.2425	-3.00	0.00227
	Imputation	0.2443	-2.28	0.00152
	Bootstrap	0.2457	-1.72	0.00152
0.50	Complete	0.5006	0.12	0.00113
	Complete subset	0.5034	0.68	0.00344
	Imputation	0.5012	0.24	0.00152
	Bootstrap	0.5007	0.14	0.00112
0.75	Complete	0.7543	0.57	0.00151
	Complete subset	0.7593	1.24	0.00441
	Imputation	0.7549	0.65	0.00152
	Bootstrap	0.7546	0.61	0.00151

Note: Shown are estimates of β, biases in relative frequency, and MSEs, based on the complete with no missing data (Complete), the complete data subset removing the subject with missing observations (Complete Subset), the standard hot deck imputation method (Imputation), and the bootstrap technique under the hot deck imputation (Bootstrap).

Table 5.2: Comparison of estimates of β when 50% of the data is missing completely at random in the dichotomous model ($I = 4, m = 1, N = 100$)

β	Method	$\hat{\beta}$	Bias (%)	$\widehat{MSE}(\hat{\beta})$
0.25	Complete	0.2460	-1.60	0.00152
	Complete Subset	0.2313	-7.48	0.00403
	Imputation	0.2469	-1.24	0.00152
	Bootstrap	0.2475	-1.00	0.00151
0.50	Complete	0.5006	0.12	0.00113
	Complete Subset	0.5038	0.76	0.00594
	Imputation	0.5014	0.28	0.00152
	Bootstrap	0.5009	0.18	0.00112
0.75	Complete	0.7543	0.57	0.00151
	Complete Subset	0.7637	1.83	0.00593
	Imputation	0.7568	0.91	0.00152
	Bootstrap	0.7562	0.83	0.00150

Note: Shown are estimates of β, biases in relative frequency, and MSEs, based on the complete with no missing data (Complete), the complete data subset removing the subject with missing observations (Complete Subset), the standard hot deck imputation method (Imputation), and the bootstrap technique under the hot deck imputation (Bootstrap).

Table 5.3: Comparison of estimates of β when 20% of the data is missing completely at random in the polytomous model ($I = 4, m = 2, N = 100$)

β	Method	$\hat{\beta}$	Bias (%)	$\widehat{MSE(\hat{\beta})}$
0.25	Complete	0.2484	-0.66	0.13446
	Complete Subset	0.2518	0.72	0.24501
	Imputation	0.2482	-0.70	0.15142
	Bootstrap	0.2484	-0.65	0.13133
0.50	Complete	0.4992	-0.16	0.05204
	Complete Subset	0.5014	0.27	0.15557
	Imputation	0.4991	-0.17	0.06316
	Bootstrap	0.4992	-0.16	0.05343
0.75	Complete	0.7500	-0.01	0.01412
	Complete Subset	0.7508	0.11	0.06698
	Imputation	0.7499	-0.01	0.01838
	Bootstrap	0.7499	-0.01	0.01457
1.00	Complete	1.0005	0.05	0.00933
	Complete Subset	1.0002	0.02	0.01832
	Imputation	1.0005	0.05	0.00901
	Bootstrap	1.0005	0.05	0.00851
1.25	Complete	1.2509	0.07	0.02563
	Complete Subset	1.2495	-0.04	0.02885
	Imputation	1.2509	0.07	0.02652
	Bootstrap	1.2508	0.07	0.02532
1.50	Complete	1.5010	0.07	0.05777
	Complete Subset	1.4987	-0.09	0.13224
	Imputation	1.5011	0.07	0.06785
	Bootstrap	1.5010	0.07	0.05283
1.75	Complete	1.7509	0.05	0.10503
	Complete Subset	1.7480	-0.12	0.37368
	Imputation	1.7510	0.06	0.13094
	Bootstrap	1.7510	0.06	0.11244

Note: Shown are estimates of β, biases in relative frequency, and MSEs, based on the complete with no missing data (Complete), the complete data subset removing the subject with missing observations (Complete Subset), the standard hot deck imputation method (Imputation), and the bootstrap technique under the hot deck imputation (Bootstrap).

Table 5.4: Comparison of estimates of β when 50% of the data is missing completely at random in the polytomous model ($I = 4, m = 2, N = 100$)

β	Method	$\hat{\beta}$	Bias (%)	$\widehat{MSE}(\hat{\beta})$
0.25	Complete	0.2423	-3.06	0.67098
	Complete Subset	0.2400	-3.99	1.05244
	Imputation	0.2414	-3.44	0.76753
	Bootstrap	0.2416	-3.34	0.70010
0.50	Complete	0.4950	-1.01	0.29003
	Complete Subset	0.4937	-1.25	0.41759
	Imputation	0.4944	-1.11	0.32358
	Bootstrap	0.4947	-1.07	0.28969
0.75	Complete	0.7476	-0.32	0.06762
	Complete Subset	0.7474	-0.35	0.07682
	Imputation	0.7474	-0.34	0.07129
	Bootstrap	0.7476	-0.32	0.06153
1.00	Complete	1.0002	0.02	0.00328
	Complete Subset	1.0007	0.07	0.00527
	Imputation	1.0003	0.03	0.00415
	Bootstrap	1.0004	0.04	0.00481
1.25	Complete	1.2527	0.22	0.08476
	Complete Subset	1.2538	0.30	0.15019
	Imputation	1.2531	0.25	0.10073
	Bootstrap	1.2531	0.24	0.09483
1.50	Complete	1.5051	0.34	0.29734
	Complete Subset	1.5066	0.44	0.45324
	Imputation	1.5057	0.38	0.33726
	Bootstrap	1.5055	0.37	0.30558
1.75	Complete	1.7574	0.42	0.62523
	Complete Subset	1.7590	0.51	0.86067
	Imputation	1.7581	0.46	0.69021
	Bootstrap	1.7578	0.44	0.61303

Note: Shown are estimates of β, biases in relative frequency, and MSEs, based on the complete with no missing data (Complete), the complete data subset removing the subject with missing observations (Complete Subset), the standard hot deck imputation method (Imputation), and the bootstrap technique under the hot deck imputation (Bootstrap).

Table 5.5: Comparison of estimates of β when 20% of the data is missing at random in the polytomous model ($I = 4, m = 2, N = 100$)

β	Method	$\hat{\beta}$	Bias (%)	$\widehat{MSE}(\hat{\beta})$
0.25	Complete	0.2490	-0.42	0.03376
	Complete Subset	0.2495	-0.19	0.04234
	Imputation	0.2482	-0.71	0.03312
	Bootstrap	0.2485	-0.58	0.04062
0.50	Complete	0.4998	-0.03	0.00258
	Complete Subset	0.5001	0.03	0.01348
	Imputation	0.4998	-0.04	0.00506
	Bootstrap	0.4997	-0.07	0.00624
0.75	Complete	0.7503	0.05	0.00225
	Complete Subset	0.7505	0.06	0.00641
	Imputation	0.7507	0.10	0.00521
	Bootstrap	0.7503	0.04	0.00418
1.00	Complete	1.0006	0.06	0.00759
	Complete Subset	1.0005	0.05	0.01675
	Imputation	1.0011	0.11	0.01192
	Bootstrap	1.0007	0.07	0.00825
1.25	Complete	1.2506	0.05	0.00967
	Complete Subset	1.2504	0.03	0.01768
	Imputation	1.2511	0.08	0.01139
	Bootstrap	1.2508	0.06	0.01029
1.50	Complete	1.5005	0.03	0.00797
	Complete Subset	1.5001	0.01	0.01504
	Imputation	1.5007	0.05	0.01595
	Bootstrap	1.5007	0.05	0.01112
1.75	Complete	1.7503	0.02	0.00594
	Complete Subset	1.7498	-0.01	0.03272
	Imputation	1.7503	0.02	0.01226
	Bootstrap	1.7505	0.03	0.01491

Note: Shown are estimates of β, biases in relative frequency, and MSEs, based on the complete with no missing data (Complete), the complete data subset removing the subject with missing observations (Complete Subset), the standard hot deck imputation method (Imputation), and the bootstrap technique under the hot deck imputation (Bootstrap).

Table 5.6: Comparison of estimates of β when 50% of the data is missing at random in the polytomous model ($I = 4, m = 2, N = 100$)

β	Method	$\hat{\beta}$	Bias (%)	$\widehat{MSE(\hat{\beta})}$
0.25	Complete	0.2486	-0.55	0.05094
	Complete Subset	0.2491	-0.36	0.09447
	Imputation	0.2485	-0.58	0.05032
	Bootstrap	0.2483	-0.66	0.04399
0.50	Complete	0.4995	-0.09	0.01157
	Complete Subset	0.5000	-0.04	0.03404
	Imputation	0.4996	-0.08	0.00965
	Bootstrap	0.4994	-0.11	0.00729
0.75	Complete	0.7501	0.02	0.00238
	Complete Subset	0.7505	0.06	0.01142
	Imputation	0.7502	0.03	0.00264
	Bootstrap	0.7502	0.02	0.00230
1.00	Complete	1.0005	0.05	0.00440
	Complete Subset	1.0006	0.07	0.00650
	Imputation	1.0006	0.06	0.00406
	Bootstrap	1.0006	0.06	0.00410
1.25	Complete	1.2507	0.05	0.00972
	Complete Subset	1.2506	0.05	0.01383
	Imputation	1.2507	0.06	0.01082
	Bootstrap	1.2508	0.06	0.00921
1.50	Complete	1.5007	0.05	0.01675
	Complete Subset	1.5004	0.02	0.03666
	Imputation	1.5007	0.05	0.01554
	Bootstrap	1.5009	0.06	0.01198
1.75	Complete	1.7507	0.04	0.02807
	Complete Subset	1.7506	0.04	0.08208
	Imputation	1.7506	0.04	0.02276
	Bootstrap	1.7508	0.05	0.01568

Note: Shown are estimates of β, biases in relative frequency, and MSEs, based on the complete with no missing data (Complete), the complete data subset removing the subject with missing observations (Complete Subset), the standard hot deck imputation method (Imputation), and the bootstrap technique under the hot deck imputation (Bootstrap).

Table 5.7: Comparison of estimates of β with 20% nonignorable missing data in the polytomous model ($I = 4, m = 2, N = 100$)

β	Method	$\hat{\beta}$	Bias (%)	$\widehat{MSE(\hat{\beta})}$
0.25	Complete	0.2495	-0.22	0.01953
	Complete Subset	0.2549	1.95	0.68244
	Imputation	0.2495	-0.19	0.06606
	Bootstrap	0.2492	-0.30	0.02324
0.50	Complete	0.5000	0.01	0.00239
	Complete Subset	0.5036	0.72	0.36095
	Imputation	0.5001	0.03	0.02581
	Bootstrap	0.5000	-0.00	0.00210
0.75	Complete	0.7503	0.05	0.00208
	Complete Subset	0.7521	0.28	0.12523
	Imputation	0.7504	0.06	0.01051
	Bootstrap	0.7504	0.05	0.00267
1.00	Complete	1.0004	0.04	0.00462
	Complete Subset	1.0005	0.05	0.00677
	Imputation	1.0005	0.05	0.00580
	Bootstrap	1.0005	0.05	0.00619
1.25	Complete	1.2503	0.03	0.00558
	Complete Subset	1.2487	-0.11	0.04918
	Imputation	1.2504	0.03	0.00931
	Bootstrap	1.2505	0.04	0.00698
1.50	Complete	1.5002	0.01	0.00518
	Complete Subset	1.4966	-0.23	0.31622
	Imputation	1.5001	0.01	0.02575
	Bootstrap	1.5003	0.02	0.00565
1.75	Complete	1.7504	-0.02	0.00513
	Complete Subset	1.7444	-0.32	0.88419
	Imputation	1.7498	-0.01	0.06328
	Bootstrap	1.7500	0.00	0.00454

Note: Shown are estimates of β, biases in relative frequency, and MSEs, based on the complete with no missing data (Complete), the complete data subset removing the subject with missing observations (Complete Subset), the standard hot deck imputation method (Imputation), and the bootstrap technique under the hot deck imputation (Bootstrap).

Chapter 6

Numerical Example: Study on Family Health of Lung Cancer Patients

6.1 Introduction

In this chapter, we will illustrate the methods described in previous chapters using the data from a study of family health of lung cancer patients. The study was designed by Kristjanson et. al. (Kristjanson et. al., 1997).

The impact of lung cancer on the patient and the family members is especially devastating because of its rapid course and grave prognosis – less than

15% of lung cancer patients will survive over five years following diagnosis. As families care for the patient, they witness his/her rapid physical deterioration and symptom distress. Family members experience both mental and physiological health changes during a cancer illness in the family and in the bereavement period, which may affect their abilities to function productively. All these demand rapid adjustment in work, marriage, family roles, and social activities on both part of patients and family members.

This study is undertaken to examine family care characteristics (family care expectations, perceptions, care satisfaction) and family health status across the illness trajectory. The objective is to understand the level of "family care satisfaction" in association with other family health and care measures.

All of the family health and care variables are measured using a rating scale. In this chapter, we will demonstrate the utility of our methods developed in this thesis by comparing (1) the approach of simply analyzing the average scores in a model, (2) using the scores obtained from the traditional Rasch method, (3) using those obtained from our improved CML approach, and (4) using those obtained from the GEE along with the consideration of the inter-item correlation in a chosen statistical analysis model. For this purpose, we use the initial response only. The full longitudinal data shall be analyzed upon development of an appropriate longitudinal Rasch model, planned for a future research.

The data are ordinal responses. We consider fitting the Rasch model for each of these variables to obtain Rasch scores, i.e. person ability parameters, for each subject.

6.2 The Data

There were 117 patients included in the study. For each patient, the patient and their family member's information were collected at the first time when the patient entered the study. Other family health and care variables were repeatedly measured using the study questionnaires at various times in up to twelve different occasions.

Before we attempt Rasch modelling and statistical analysis with the lung cancer data, we cleaned up the data for consistency in the scale and regroup some variables to facilitate interpretation in the analysis results, as follows.

The study collected the following demographical information: marital status, age, sex, educational level, occupation, income, ethnicity, religion, and the relationship of family member to the patient. To facilitate interpretation, we simplified these variables as follows, with baseline set to be the first category (with the assigned category number of 0):

1. Marital Status: married (=0) ("married" or "common-law") or not mar-

ried (=1) ("divorced", "never married" or "widowed").

2. Age: "18–50" (=0), "51–65" (=1) or "65 or older" (=2).

3. Sex: male (=0) or female (=1).

4. Educational Level: less than high school (=0), high school graduates (=1) or higher than high school education (=2).

5. Occupation: not working (=0) or working (=1).

6. Income: less than \$20,000 (=0), \$20,000 – \$40,000 (=1) or higher than \$40,000 (=2).

7. Ethnicity: non-European (=0) or European (=1).

8. Religion: Christian (=0) or non-Christian (=1).

9. Relationship of family member to the patient: spouse (=0) or non-spouse (=1).

Table 6.1 shows the frequencies for each level of these demographical variables from the patients and their family members.

We also have the following family health variables: SDS (symptom distress scale with 13 items; 1-normal, 2-occasional distress, 3-frequent distress, 4-usual distress, 5-constant distress), QOLR (current quality of life rating, ranging from 1 = poor to 10=excellent), QOLS (satisfaction with current quality of life, ranging from 1="not at all" to 10="very satisfied"), FAMCAR (family care satisfaction with 20 items; 1=very satisfied, 2=satisfied, 3=undecided, 4=dissatisfied, 5=very dissatisfied), SOS (symptom of stress scale with 94 items; 0=never, 1=infrequently, 2=sometimes, 3=often, 4=very frequently), FAD (family assessment device with 12 items; 1=strongly agree, 2=agree, 3=disagree, 4=strongly disagree).

The family care variables include: FEXP (family expectations scale with 16 items, ranging from 0="not at all important to me" to 10="very important to me") and FPER (family perceptions scale with 21 items; 1=strongly agree, 2=agree, 3=uncertain, 4=disagree, 5=strongly disagree).

The first three variables, SDS, QOLR, and QOLS, were measured from patients themselves and the remaining variables were observed from their family members. The outcome variable of interest is the family care satisfaction, FAMCAR, explained by others conditionally.

The variables QOLR (current quality of life rating) and QOLS (satisfaction with current quality of life) were rearranged to binary response: 0="negative rating or not satisfied", 1="positive rating or satisfied". All other family health and care variables were similarly rearranged from 0 to m ($m+1$ is the number of categories in that variable) so that high values represent positive attitudes. For example, for FAMCAR, before the rearrangement, 1 stood for "very satisfied", and 5 meant "very dissatisfied". After the rearrangement, 0 now stands for "very dissatisfied", and 4 for "very satisfied", consistently with others to indicate that a high score means a high degree of satisfaction expressed. The frequencies of the 20 FAMCAR items are shown in *Table 6.2*.

6.3 Preliminary Analysis

In this section, we consider several possible ways to analyze the given data. As the family health and care variables are rating scale data, there are a number of ways one can measure a score for each study subject on a linear scale, suitable to be used in a statistical model. We first recognize that there are statistical issues to deal with such correlated items and missing values in the data.

First, we deal with the missing data issue. As discussed in *Chapter 5*, we will use the hot deck imputation method with or without the bootstrap technique. Second, to account for the correlations among items, we will use the two-step iteration method (4.9) and (4.17) for the Rasch model and obtain the person ability parameters for each subject adjusted for these polychoric correlations among items. We do this for each of the family health and care variables measured in a rating scale.

Table 6.3 gives the estimates of the polychoric correlations among items on the basis of three competing missing data methods. The correlations among family assessment device items (FAD), family expectation items (FEXP), and symptom distress scale (SDS) items do not vary greatly by different methods, mainly because of a very small proportion of missing values (3.65%, 3.09%, and 8.06% respectively). The correlations among family care satisfaction (FAMCAR) items, family perceptions (FPER) items, and symptom of stress (SOS) items were changed slightly from using the complete data subset to using the imputed data, due to increased missing proportions compared to the others (12.87%, 16.13%, and 16.08% respectively). The larger the missing proportion in a variable, the larger the discrepancy between the correlation from complete data subset and that from imputed data. The latter two methods produced nearly identical estimates, while the complete subset data estimated them slightly higher than the other two.

However, overall there were no appreciable amount of variation in the level of correlations estimated. This is also consistent with our findings in *chapter 5*, where we found that the three approaches to missing data did not result in any substantial differences, when the missing proportions are not large (overall proportion about 10% in this example). Therefore, we shall not present results from all three methods, but only that from the bootstrap under hot deck imputation method, which was found to the best under all missing data structures considered.

We note that some of the correlations are quite high. For example, the correlation among the family expectation items is greater than 0.85. That among the family assessment device items is also close to 0.5, regardless which method is used. Ignoring the existence of such high correlations can lead to misleading conclusions.

6.4 Obtaining Linear Scores

We will demonstrate the utility of our research by contrasting the following four approaches in the analysis: (1) an average scoring which takes a simple summary score for each subject on a number of given items; (2) the traditional Rasch method using the CON procedure (Wright and Masters, 1982), which assumes no dependency among items or parameters; (3) the simultaneous method with the improved CML procedure (Sheng and Carrière, 2002), with dependency among parameters incorporated but not among items; and (4) the simultaneous Rasch method with polychoric correlation incorporated.

As an example, *Table 6.4* gives the estimates of distinct set of person ability parameters $\boldsymbol{\beta}$, adjusted for item difficulty, for the family assessment device variables, by the average scoring and the three Rasch methods. We observe that average scores take positive values only (range from 0 to m), while Rasch scores adjusted by item difficulty differences can take both positive and negative values. The person ability parameters with consideration of the polychoric correlation have the smallest range (-0.8 to 2.2), followed by those from the independent assumption of items (-1.4 to 2.4), and those by the traditional Rasch method (-2.2 to 2.9). The estimates of $\boldsymbol{\beta}$ for other family health and care variables are obtained in a similar way, as well as the item difficulty parameters $\boldsymbol{\delta}$.

Given the values of person ability and item difficulty parameters, the probability π_{ijk} of any subject j responds on category k to item i can be easily calculated. As an example, *Tables 6.5 - 6.7* give selected estimates of π_{ijk} from the three Rasch approaches for the family assessment device variable. We can see that the entries to the same category for each item are decreasing gradually as the subject raw scores increases; however, due to too many categories, there is no substantial difference between the two subjects with adjacent raw scores. The probabilities of subjects with the same raw score responding to the same category for each item are quite different from one method to the others. The difference is particularly large for those with extreme raw scores (close to 0 or $I * m$), since the person abilities vary greatly according to the methods used. Estimates of π_{ijk} from other variables are not presented, but exhibit the similar pattern.

6.5 Using the Person Scores in Regression Analysis

After obtaining the Rasch scores for each subject, i.e., the estimates for the person ability parameters adjusted by the item difficulty, investigators often use them in regression studies to describe the relationship between family care satisfaction and other family health and care measures. The purpose of this section is to demonstrate varying degree of conclusions possible by different linearizing methods using the regression analytic method, a typical method of analysis by practioners. For each family health and care variable, we created one score for each subject, upon accounting for item difficulty.

We also created a new variable upon requests by the primary investigators of the family care study project. It is named "discrepancy" (DIS), which describes the discrepancy between the family expectations scale and the family perceptions scale for each subject. This variable will be used in the subsequent analysis in lieu of the two variables, "family expectations" and "family perceptions". The investigators also strongly believe that there exists a relationship between family care satisfaction scale and the quadratic term of the discrepancy, and this is verified graphically. That is, the level of satisfaction rises rapidly when DIS is negative with higher health perceptions than expected, but the rate of growth tapers down as DIS becomes positive with higher health expectation than perceived.

Table 6.8 gives the result of these models using person scores from the four approaches in the regression analysis of the family care satisfaction score on the other covariates. The full model includes all covariates in the regression model, and the reduced model only includes significant ones. After fitting the regression model, we assessed the model assumptions such as normality, linearity and equal variance in the data and confirmed these to be satisfactory. *Figure 6.1–6.4* show the relevant plots of the results from the four reduced models. Other than a few points from the data set of 117 patients, we see that these model assumptions are basically satisfied by the fitted model. Although the residuals are a little skewed to the right, the departure from normality appeared to be moderate. The regression coefficients and their standard errors are robust against slight non-normality (Ramsay and Schafer, 1996).

The results from the average scoring method and the traditional Rasch method (based on CON procedure) are worse than the two competing methods that consider simultaneous estimation of parameters and incorporate the polychoric correlation among items: the first two have larger MSEs and much smaller R^2 values that implies large variation left unexplained by the respective models. In the reduced model, although the MSEs are similar for all approaches, the R^2 values from the latter two approaches are much better than those from the first two methods. Here we conclude that the simple average scoring method has relatively poor fitting, and the traditional Rasch method is not satisfactory. Even the improved CML is not perfect, as it does not account for the polychoric correlations among items, especially those that are too large to be ignorable. This will have big impact in the evaluation of each individual covariate's contribution to the significant results.

Table 6.9 displays the ordinary least squares estimates of all significant coefficients ($P < 0.05$) in the reduced model. Each model contains a different set of significant covariates. The following variables are common in all four models: the family member's age, the family assessment device level, the symptom distress scale, discrepancy score, and the squared discrepancy score. Among them, the symptom distress score and the family member's age (51–65) are moderately significant, while the family assessment device level and discrepancy level (both the linear and quadratic terms) are highly significant with one exception that the linear effect of the discrepancy score is not significant in the average scoring method. When the family members are between 51 and 65 years old and have poor family environment measured by FAD, they are less satisfied with health care. When the patients develop less symptom distress, the family care satisfaction level tended to be higher.

The use of Rasch scores revealed a few more covariates being significant in addition to the above.

In the model using the traditional Rasch method, the patient's age and education level and the family member's ethnic background are also found to be significantly related to family care satisfaction level. The family care satisfaction level is positively associated with middle-aged patients, European family background, and higher than high school education of the patients.

According to the model using the improved CML method, the patient and family member's education level are also relevant. Those who have higher than high school education tend to have higher satisfaction level than those who have less than high school education.

In the model using the simultaneous Rasch method with the polychoric correlation, the patient's education and the family member's ethnic background have significant relationships with the family satisfaction level. The family care satisfaction level is positively associated with higher than high school education of the patients and their European family background.

In summary, the directions and the levels of association between family health and care variables remained consistent among different scoring methods. However, the significance levels and the key variables identified were not the same, as discussed above. Overall, use of the Rasch methods produced results with higher precision than the simple average scoring method. As pointed out numerously throughout in this thesis, clearly the first three approaches have some limitations, either unfulfilling the properties of the fundamental measurement (the average scoring method), assuming independence of related parameters (the traditional Rasch method), or assuming independence of the inter-related questionnaire items. On the other hand, the method based on GEE with the polychoric correlation incorporated rectifies all these limitations in the first three methods. Therefore, we conclude based on the latter method that the family's ethnic background do matter in their family care satisfaction with the patient care – people of European descent tend to be happier with the family health care. Patients who had some college education seem to manage better, leading to generally positive attitude by family members about the patient care.

Table 6.1: Frequency table of all demographical variables

Variable	0	1	2	Missing
Family Member:				
Marital Status	100	14		3
Age	36	40	38	3
Sex	32	81		4
Education	55	27	31	4
Income	22	44	34	17
Occupation	56	55		6
Ethnicity	49	65		3
Religion	104	10		3
Relationship	72	39		6
Patient:				
Marital Status	89	26		2
Age	7	39	70	1
Sex	66	50		1
Education	79	19	17	2
Income	32	55	19	11
Occupation	79	37		1
Ethnicity	45	71		1
Religion	104	11		2

Note: Shown are frequencies in each level of the demographical variables from patients and their family members.

Table 6.2: Frequency table of FAMCAR variables

Item	0	1	2	3	4	Missing
FAMCAR1	54	162	55	17	3	98
FAMCAR2	32	197	89	47	12	12
FAMCAR3	37	203	86	35	8	20
FAMCAR4	38	201	73	27	8	42
FAMCAR5	73	186	25	14	3	88
FAMCAR6	64	129	25	11	5	155
FAMCAR7	43	175	69	46	12	44
FAMCAR8	54	219	45	32	7	32
FAMCAR9	64	212	55	23	5	30
FAMCAR10	68	240	38	6	3	34
FAMCAR11	52	200	62	28	8	39
FAMCAR12	56	172	65	5	1	90
FAMCAR13	41	247	44	18	4	35
FAMCAR14	55	186	50	56	18	24
FAMCAR15	39	214	67	29	6	34
FAMCAR16	33	152	51	38	7	108
FAMCAR17	35	204	63	37	12	38
FAMCAR18	46	215	59	32	7	30
FAMCAR19	50	224	46	30	8	31
FAMCAR20	60	223	52	23	9	22

Note: Shown are frequencies in the rating scale of the 20 family care satisfaction items.

Table 6.3: Polychoric correlations based on three missing data techniques: family health and care variables in lung cancer study

	Polychoric Correlations					
Method	FAMCAR	FAD	FEXP	FPER	SDS	SOS
Missing Proportion (%)	12.87	3.65	3.09	16.13	8.06	16.08
Complete Subset	0.418	0.496	0.866	0.341	0.270	0.369
Imputation	0.359	0.478	0.859	0.259	0.252	0.288
Bootstrap	0.352	0.477	0.857	0.255	0.253	0.284

Note: Shown are the missing proportions (first row) for each of the family health and care variables, as well as the estimates of the polychoric correlations obtained from using the complete data subset removing the missing observations (Complete Subset), the standard hot deck imputation method (Imputation), and the bootstrap technique under imputation (Bootstrap).

Table 6.4: Estimates of β for the family assessment device variables

r	Average Score	Traditional	Improved CML	With Correlation
1	0.0833	-2.2272	-1.4476	-0.8036
2	0.1667	-1.9944	-1.2840	-0.7659
3	0.2500	-1.9657	-0.9670	-0.5927
4	0.3333	-1.6836	-0.4151	-0.4748
5	0.4167	-1.5692	-0.4115	-0.2392
6	0.5000	-1.1232	-0.2989	-0.1820
7	0.5833	-0.8264	-0.1696	-0.1053
8	0.6667	-0.6455	0.0873	-0.0330
9	0.7500	-0.3960	0.1186	-0.0269
10	0.8333	-0.3320	0.1346	0.1002
11	0.9167	-0.1350	0.2240	0.1907
12	1.0000	-0.1098	0.2332	0.2712
13	1.0833	-0.0137	0.3010	0.3005
14	1.1667	0.0295	0.3992	0.4068
15	1.2500	0.1265	0.4079	0.4738
16	1.3333	0.3729	0.4289	0.6081
17	1.4167	0.4407	0.4526	0.6342
18	1.5000	0.4890	0.4882	0.6351
19	1.5833	0.6978	0.6051	0.6585
20	1.6667	0.7190	0.6981	0.8042
21	1.7500	0.7831	0.7081	0.9467
22	1.8333	0.9824	0.7665	0.9710
23	1.9167	0.9853	0.8604	1.1032
24	2.0000	0.9979	0.9771	1.2064
25	2.0833	1.1314	1.0398	1.2152
26	2.1667	1.4012	1.0847	1.2591
27	2.2500	1.6472	1.1211	1.4537
28	2.3333	1.7013	1.1280	1.5116
29	2.4167	1.7186	1.3423	1.5958
30	2.5000	2.1179	1.4537	1.5989
31	2.5833	2.2911	1.4927	1.6972
32	2.6667	2.4414	1.6638	1.8264
33	2.7500	2.5280	1.9342	1.8833
34	2.8333	2.6141	1.9985	1.8972
35	2.9167	2.8927	2.2486	2.1036

Note: Entries are the estimates of β, ordered by subject's raw score r, from the simple average scoring method ("Average Score"), the traditional Rasch method ("Traditional"), the improved CML procedure ("Improved CML"), and simultaneous Rasch analysis with a polychoric correlation ("With Correlation").

Table 6.5: Estimates of π_{ijk} from the traditional Rasch method for the family assessment device variables, for selected subjects with raw score of r

	Item Number											
r k	1	2	3	4	5	6	7	8	9	10	11	12
1 0	0.64	0.90	0.96	0.99	0.96	0.94	0.72	0.89	0.97	0.81	0.99	0.96
1	0.33	0.08	0.03	0.01	0.04	0.03	0.22	0.10	0.02	0.18	0.01	0.04
2	0.02	0.02	0.01	0.00	0.00	0.02	0.05	0.01	0.01	0.01	0.00	0.00
3	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00
2 0	0.63	0.90	0.95	0.99	0.96	0.93	0.71	0.89	0.97	0.81	0.99	0.96
1	0.34	0.08	0.03	0.01	0.04	0.03	0.23	0.11	0.02	0.19	0.01	0.04
2	0.02	0.02	0.02	0.00	0.00	0.03	0.05	0.00	0.01	0.00	0.00	0.00
3	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00
17 0	0.01	0.08	0.15	0.70	0.35	0.03	0.01	0.15	0.12	0.16	0.84	0.15
1	0.08	0.11	0.08	0.14	0.20	0.02	0.06	0.27	0.05	0.56	0.13	0.08
2	0.07	0.43	0.50	0.06	0.03	0.21	0.22	0.18	0.14	0.24	0.02	0.12
3	0.84	0.38	0.27	0.10	0.41	0.74	0.71	0.39	0.69	0.04	0.01	0.65
18 0	0.01	0.06	0.13	0.67	0.31	0.03	0.01	0.13	0.09	0.14	0.83	0.12
1	0.06	0.10	0.07	0.14	0.19	0.01	0.05	0.25	0.04	0.55	0.14	0.08
2	0.07	0.43	0.50	0.06	0.03	0.20	0.21	0.19	0.14	0.26	0.02	0.12
3	0.86	0.41	0.30	0.13	0.47	0.76	0.73	0.43	0.73	0.05	0.01	0.68
19 0	0.01	0.06	0.13	0.66	0.31	0.03	0.01	0.13	0.09	0.14	0.82	0.12
1	0.06	0.10	0.07	0.15	0.19	0.01	0.05	0.25	0.04	0.55	0.14	0.07
2	0.07	0.43	0.50	0.06	0.03	0.20	0.21	0.19	0.14	0.26	0.03	0.12
3	0.86	0.41	0.30	0.13	0.47	0.76	0.73	0.43	0.73	0.05	0.01	0.69
34 0	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00
1	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.01	0.00	0.12	0.16	0.00
2	0.01	0.13	0.19	0.07	0.01	0.04	0.04	0.06	0.03	0.39	0.21	0.03
3	0.99	0.87	0.81	0.90	0.98	0.96	0.96	0.93	0.97	0.49	0.50	0.97
35 0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00
1	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.07	0.09	0.00
2	0.01	0.09	0.14	0.05	0.01	0.02	0.03	0.04	0.02	0.32	0.19	0.02
3	0.99	0.91	0.86	0.94	0.99	0.98	0.97	0.96	0.98	0.61	0.67	0.98

Note: Entries are the estimates of π_{ijk} for subject j with raw score of r, responding on category k to item i, where subjects are ordered by their raw score r.

Table 6.6: Estimates of π_{ijk} from the improved CML Rasch procedure for the family assessment device variables, for selected subjects with raw score of r

		Item Number											
r	k	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0.33	0.71	0.84	0.97	0.91	0.73	0.39	0.74	0.89	0.61	0.97	0.88
	1	0.45	0.16	0.07	0.03	0.09	0.07	0.32	0.22	0.06	0.36	0.03	0.08
	2	0.08	0.11	0.08	0.00	0.00	0.13	0.19	0.03	0.03	0.03	0.00	0.02
	3	0.14	0.02	0.01	0.00	0.00	0.07	0.10	0.01	0.02	0.00	0.00	0.02
2	0	0.32	0.70	0.83	0.96	0.90	0.71	0.37	0.73	0.88	0.60	0.97	0.87
	1	0.45	0.16	0.08	0.04	0.09	0.07	0.32	0.23	0.06	0.37	0.03	0.09
	2	0.08	0.12	0.08	0.00	0.00	0.14	0.20	0.03	0.03	0.03	0.00	0.02
	3	0.15	0.02	0.01	0.00	0.01	0.08	0.11	0.01	0.03	0.00	0.00	0.02
17	0	0.01	0.09	0.17	0.73	0.39	0.04	0.01	0.17	0.14	0.17	0.85	0.17
	1	0.08	0.11	0.08	0.13	0.21	0.02	0.07	0.29	0.05	0.57	0.12	0.09
	2	0.08	0.44	0.49	0.05	0.03	0.22	0.23	0.18	0.15	0.22	0.02	0.12
	3	0.83	0.36	0.26	0.09	0.37	0.72	0.69	0.36	0.66	0.04	0.01	0.62
18	0	0.01	0.07	0.14	0.68	0.32	0.03	0.01	0.14	0.10	0.15	0.83	0.13
	1	0.07	0.10	0.07	0.14	0.20	0.02	0.06	0.26	0.04	0.56	0.13	0.08
	2	0.07	0.43	0.50	0.06	0.03	0.20	0.21	0.19	0.14	0.25	0.03	0.12
	3	0.85	0.40	0.29	0.12	0.45	0.75	0.72	0.42	0.72	0.04	0.01	0.67
19	0	0.01	0.06	0.12	0.66	0.30	0.02	0.01	0.12	0.09	0.14	0.82	0.11
	1	0.06	0.09	0.07	0.15	0.19	0.01	0.05	0.25	0.04	0.55	0.14	0.07
	2	0.07	0.43	0.50	0.06	0.03	0.20	0.21	0.19	0.13	0.26	0.03	0.12
	3	0.86	0.42	0.31	0.13	0.48	0.77	0.73	0.44	0.74	0.05	0.01	0.70
34	0	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.01	0.26	0.00
	1	0.00	0.01	0.01	0.04	0.02	0.00	0.00	0.02	0.00	0.19	0.21	0.01
	2	0.02	0.17	0.25	0.08	0.01	0.05	0.06	0.08	0.04	0.42	0.20	0.03
	3	0.98	0.82	0.74	0.84	0.97	0.95	0.94	0.90	0.96	0.38	0.33	0.96
35	0	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.01	0.25	0.00
	1	0.00	0.01	0.01	0.04	0.02	0.00	0.00	0.02	0.00	0.18	0.21	0.01
	2	0.02	0.17	0.24	0.08	0.01	0.05	0.05	0.08	0.04	0.42	0.20	0.03
	3	0.98	0.82	0.75	0.85	0.97	0.95	0.95	0.90	0.96	0.39	0.34	0.96

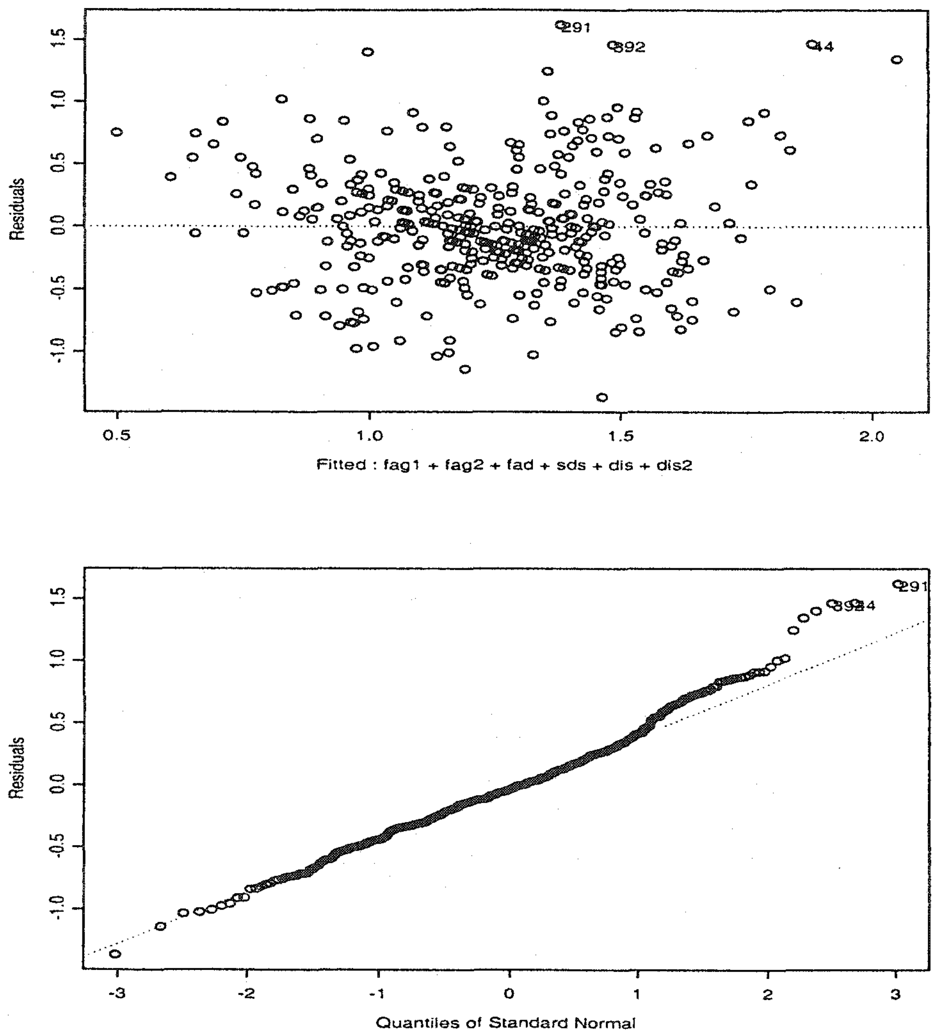
Note: Entries are the estimates of π_{ijk} for subject j with raw score of r, responding on category k to item i, where subjects are ordered by their raw score r.

Table 6.7: Estimates of π_{ijk} from the simultaneous Rasch analysis with polychoric correlation for the family assessment device variables, for selected subjects with raw score of r

	Item Number											
r k	1	2	3	4	5	6	7	8	9	10	11	12
1 0	0.17	0.52	0.69	0.94	0.84	0.47	0.20	0.61	0.75	0.48	0.96	0.76
1	0.39	0.20	0.10	0.05	0.13	0.07	0.28	0.30	0.08	0.47	0.04	0.12
2	0.11	0.23	0.18	0.01	0.01	0.23	0.27	0.06	0.07	0.05	0.00	0.05
3	0.33	0.05	0.03	0.00	0.02	0.23	0.24	0.03	0.10	0.00	0.00	0.07
2 0	0.15	0.48	0.66	0.93	0.82	0.42	0.18	0.58	0.72	0.45	0.95	0.73
1	0.37	0.20	0.11	0.06	0.14	0.07	0.26	0.32	0.08	0.48	0.05	0.13
2	0.11	0.25	0.20	0.01	0.01	0.25	0.28	0.06	0.08	0.06	0.00	0.05
3	0.37	0.07	0.03	0.00	0.03	0.26	0.27	0.04	0.12	0.01	0.00	0.09
17 0	0.01	0.07	0.13	0.68	0.32	0.03	0.01	0.13	0.10	0.15	0.83	0.13
1	0.07	0.10	0.08	0.14	0.20	0.02	0.06	0.26	0.04	0.56	0.13	0.08
2	0.07	0.43	0.50	0.06	0.03	0.20	0.21	0.19	0.14	0.25	0.03	0.12
3	0.85	0.40	0.29	0.12	0.45	0.75	0.72	0.42	0.72	0.04	0.01	0.67
18 0	0.01	0.07	0.13	0.68	0.32	0.03	0.01	0.13	0.10	0.15	0.83	0.13
1	0.07	0.10	0.08	0.14	0.20	0.02	0.06	0.26	0.04	0.56	0.13	0.08
2	0.07	0.43	0.50	0.06	0.03	0.20	0.21	0.19	0.14	0.25	0.03	0.12
3	0.85	0.40	0.29	0.12	0.45	0.75	0.72	0.42	0.72	0.04	0.01	0.67
19 0	0.01	0.06	0.13	0.67	0.32	0.03	0.01	0.13	0.10	0.15	0.83	0.12
1	0.07	0.10	0.07	0.15	0.19	0.01	0.06	0.26	0.04	0.56	0.14	0.08
2	0.07	0.43	0.50	0.06	0.03	0.20	0.21	0.19	0.14	0.25	0.02	0.12
3	0.85	0.41	0.30	0.12	0.46	0.76	0.72	0.42	0.72	0.04	0.01	0.68
34 0	0.00	0.00	0.01	0.15	0.02	0.00	0.00	0.01	0.00	0.03	0.50	0.01
1	0.01	0.02	0.02	0.09	0.05	0.00	0.01	0.06	0.01	0.32	0.24	0.01
2	0.03	0.26	0.36	0.11	0.02	0.08	0.09	0.12	0.06	0.43	0.13	0.05
3	0.96	0.72	0.61	0.65	0.91	0.92	0.90	0.81	0.93	0.22	0.13	0.93
35 0	0.00	0.00	0.01	0.10	0.02	0.00	0.00	0.01	0.00	0.02	0.44	0.00
1	0.01	0.02	0.01	0.08	0.03	0.00	0.01	0.04	0.01	0.29	0.24	0.01
2	0.02	0.24	0.33	0.11	0.02	0.07	0.08	0.11	0.05	0.43	0.15	0.05
3	0.97	0.74	0.65	0.71	0.93	0.93	0.91	0.84	0.94	0.26	0.17	0.94

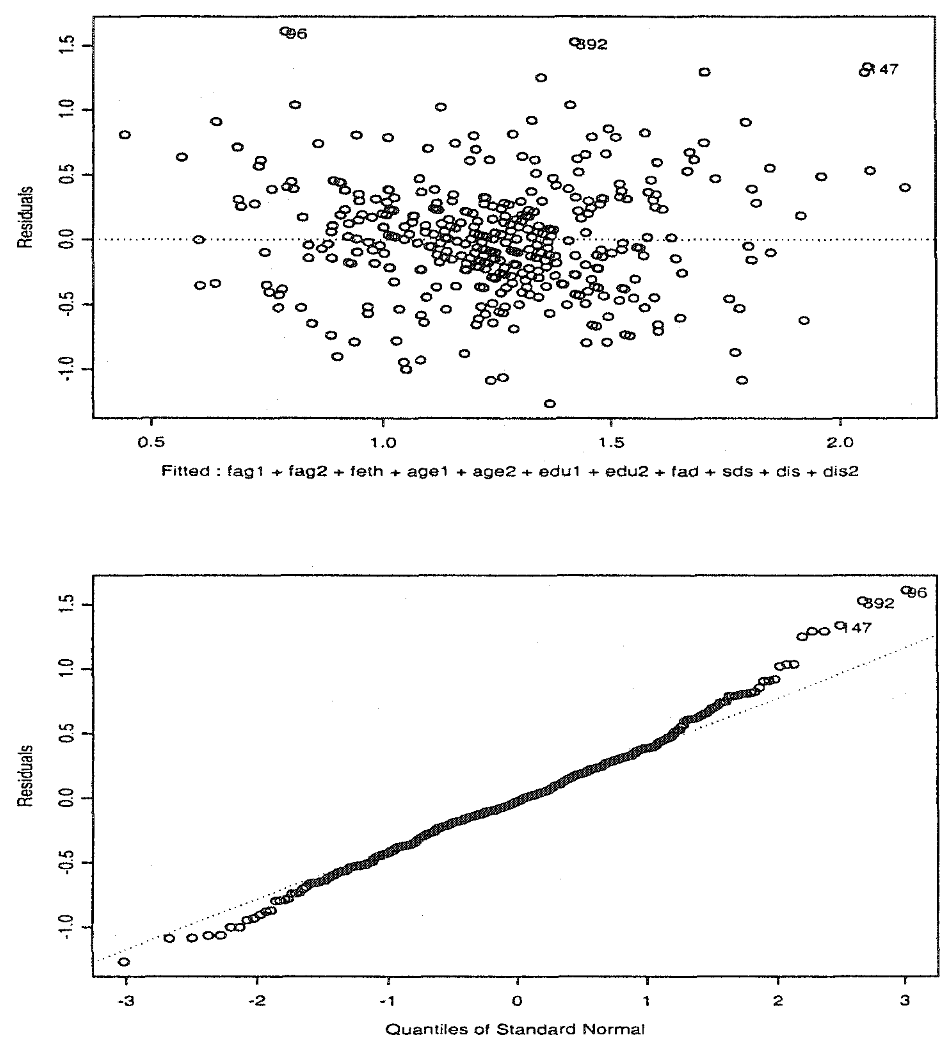
Note: Entries are the estimates of π_{ijk} for subject j with raw score of r, responding on category k to item i, where subjects are ordered by their raw score r.

Figure 6.1: Plots from the reduced regression model using the simple average scoring method



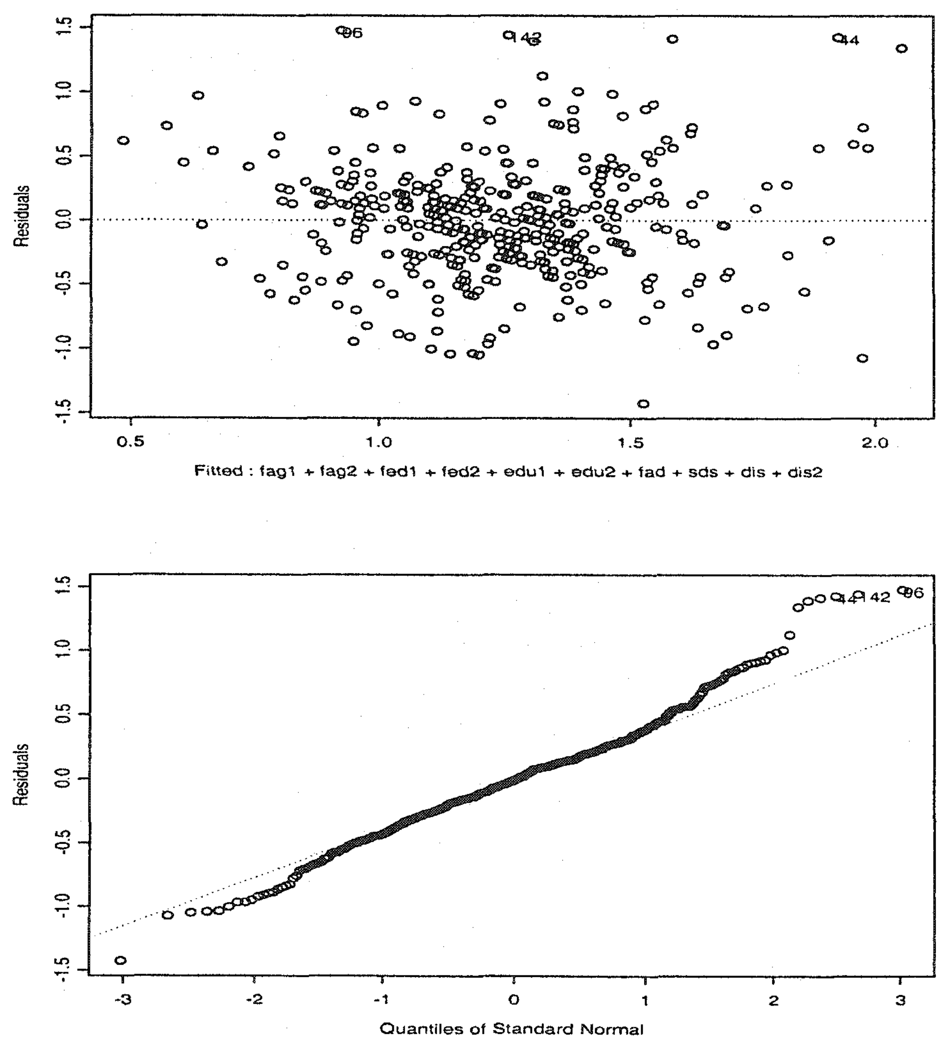
Note: Plotted are: residual versus fitted values, and the normal Q-Q plot of the residuals.

Figure 6.2: Plots from the reduced regression model using the traditional Rasch method



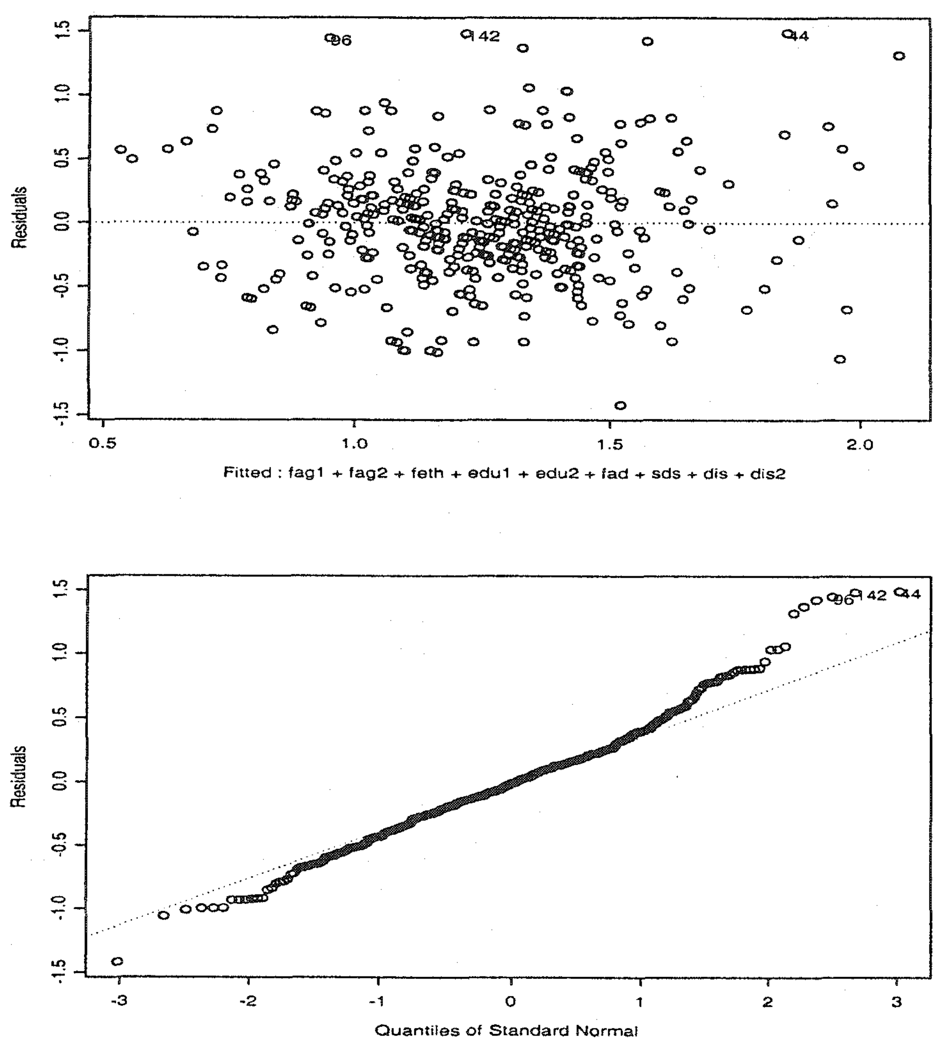
Note: Plotted are: residual versus fitted values, and the normal Q-Q plot of the residuals.

Figure 6.3: Plots from the reduced regression model using the improved CML procedure



Note: Plotted are: residual versus fitted values, and the normal Q-Q plot of the residuals.

Figure 6.4: Plots from the reduced regression model using the simultaneous Rasch analysis with the polychoric correlation



Note: Plotted are: residual versus fitted values, and the normal Q-Q plot of the residuals.

Table 6.8: Result of regression models of family care satisfaction scores in the lung cancer study using person scores obtained from different methods

	Full Model			Reduced Model		
	Root MSE	R^2	F-stat	Root MSE	R^2	F-stat
Average Score	0.460	0.287	3.388	0.464	0.257	9.89
Traditional	0.455	0.298	4.776	0.455	0.270	12.02
Improved CML	0.437	0.513	5.209	0.452	0.369	13.62
With Correlation	0.415	0.606	5.511	0.442	0.438	14.19

Note: Shown are the regression models characteristics (Root MSE, R^2, F-statistic) from the simple average scoring method ("Average Score"), the traditional Rasch method ("Traditional"), the improved CML procedure ("Improved CML"), and the simultaneous Rasch analysis with the polychoric correlation ("With Correlation"). The significance levels of all F-statistics are less than 0.001.

Table 6.9: Significant coefficients of reduced regression models of family care satisfaction scores in lung cancer study

	Method 1	Method 2	Method 3	Method 4
(Int)	1.402(0.312)	1.303(0.175)	1.209(0.158)	1.397(0.154)
Family member:				
Age				
51–65	**-0.346(0.106)	***-0.210(0.058)	*-0.134(0.058)	*-0.149(0.058)
65+	-0.026(0.098)	-0.041(0.061)	0.000(0.065)	-0.048(0.059)
Education				
HighSch	N/S	N/S	0.003(0.064)	N/S
College	N/S	N/S	*0.143(0.064)	N/S
Ethnicity				
European	N/S	**0.112(0.048)	N/S	*0.101(0.047)
FAD	**-0.346(0.101)	***-0.254(0.051)	***-0.230(0.051)	***-0.268(0.052)
DIS	-0.114(0.074)	***-0.140(0.020)	***-0.129(0.020)	***-0.121(0.019)
DIS2	***0.033(0.009)	***0.035(0.004)	***0.034(0.004)	***0.033(0.004)
Patient:				
Age				
51–65	N/S	**0.273(0.105)	N/S	N/S
65+	N/S	0.070(0.103)	N/S	N/S
Education				
HighSch	N/S	-0.106(0.078)	-0.097(0.079)	-0.045(0.078)
College	N/S	*0.149(0.065)	**0.175(0.067)	*0.166(0.066)
SDS	**0.211(0.071)	*0.109(0.042)	*0.104(0.042)	*0.099(0.042)

Note: Shown are the estimates (standard errors) of coefficients with $P < 0.1$ from regression models using the simple average scoring method (Method 1), the traditional Rasch method (Method 2), the improved CML procedure (Method 3). The simultaneous Rasch analysis with polychoric correlation (Method 4), with the level of significance is as indicated (*: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; N/S: non-significant).

Chapter 7

Conclusions

In this thesis, we discussed several statistical issues that could arise from the analyses of item response data upon recognizing the lack of appropriate methodologies to a number of realistic situations.

We implemented a simultaneous CML estimation method for dichotomous and polytomous Rasch models and derived their asymptotic properties in *Chapter 3*. The advantage to this approach of carrying out the simultaneous estimation method was rather substantial in both the polytomous and dichotomous Rasch models. The improvement in efficiency of the estimators, as compared to those of the currently used conditional approach, was especially apparent for intermediately valued person ability parameters. The simultaneous estimation method also had a huge impact on the model fit. We constructed the conditional likelihood ratio test for the goodness-of-fit of the model. The test statistic was shown to be distributed according to the assumed asymptotic χ^2 distribution. On the other hand, the corresponding results based on the current conditional approach deviates significantly from the expected distribution. In summary, our conclusion is that the current approach has shortcomings in not considering correlations implicit in the Rasch model parameters. Our implementation, based on a conditional likelihood function, improved the fit of the data as well as the precision of the estimators in comparison to those of the other CML methods such as CON.

In *Chapter 4*, we developed a method to account for the inter-item correlations, known as the polychoric correlation, among items that are measured on an ordinal scale. We proposed a latent variable approach to the Rasch model in this chapter. Using the idea of generalized estimating equations, we expanded the estimation method for the Rasch model parameters under item-to-item correlations. Generally, we observe consistent estimators for the Rasch model parameters and the polychoric correlation coefficient with normal distributions. Simulation study shows the relative efficiency of the estimators when the inter-item correlation is considered. The efficiency loss of the estimators is shown to be worse as the polychoric correlation rises.

In *Chapter 5*, we discussed several methods to deal with missing observations under three missing data mechanisms. When data are missing completely at random, most imputation methods can adequately reduce the impact of the missing data, especially when a small proportion of data is missing. When data are missing at random, the bootstrap under hot deck imputation method is better than the other imputation methods in terms of small MSE of estimators (with small bias from all methods). This becomes particularly apparent when the missing proportion is large. On the other hand, when the missing data are nonignorable, only the bootstrap under hot deck imputation method could produce efficient estimators. As expected, using the complete data subset did not give satisfactory results. It produced either very inefficient or biased estimators. Overall, the bootstrap under hot deck imputation method is proven to be the most superior in producing efficiently consistent estimators and their variances, when there are missing values present in the item response data.

In *Chapter 6*, we applied the methods discussed in the thesis to a real data set on the family satisfaction study involving lung cancer patients to demonstrate the utility of our research contributions. Advantages of the methods we developed are clearly demonstrated. Comparison with other commonly used methods revealed that although overall qualitative results may be similar, there are important differences in the set of covariates being recognized as significant and in some key variables. We based the interpretation of the results on our methods, because ours accommodate simultaneous estimation of parameters, correlation among items and the most efficient missing data method.

This thesis also identified a number of areas that needs further research. First, when we accounted for the inter-item correlation, we assumed that the polychoric correlations are the same among all items. This may not be necessarily satisfied in the real world situations. Quite often, the correlations may be found to be similar among all subjects, but not so among items. Further research is needed that will include this possibility. Second, the methods developed here may be extended to accommodate longitudinal item response data that collect repeated measurements over time from the subjects. Further investigation is necessary to confirm this conjecture. Third, we used large sample simulation results to show the superior performance of the proposed methods over the others in dealing with missing data. However, the asymptotical behavior of the estimators has not been well established. We plan to develop it under either specific settings of Rasch models, or general situations of item response data.

Bibliography

[1] Anderson, E.B. (1970). Asymptotic properties of conditional maximum likelihood estimation. *Journal of the Royal Statistical Society B*, 32, 283–301.

[2] Anderson, E.B. (1973a). *Conditional Inference and Models for Measuring.* Copenhagen: Mentalhygiejnisk Forlag.

[3] Anderson, E.B. (1973b). A goodness of fit for the Rasch model. *Psychometrika*, 38, 123–140.

[4] Andrich, D. (1978a). A binomial latent trait model for the study of Likert-style attitude questionaires. *British Journal of Mahtematical and Statistical Psychology*, 31, 84–98.

[5] Andrich, D. (1978b). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–573.

[6] Andrich, D. (1978c). Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement*, 38, 665–680.

[7] Andrich, D. (1978d). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 2, 581–594.

[8] Andrich, D. (1979). A model for contingency tables having an ordered response classification. *Biometrics*, 1979, 403–415.

[9] Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. Lord, and M. Novick (Eds.), *Statistical Theories of Mental Test Scores.* Reading, MA: Addison-Wisley.

[10] Carrière, K.C. (1994). Incomplete Repeated Measures Data Analysis in the Presence of Treatment Effects. *Journal of American Statistical Association,* 89, 680–686.

[11] Carrière, K.C. (1999). Methods for Repeated Measures Data Analysis with Missing Values. *Journal of Statistical Planning and Inference*, 77, 221–236.

[12] Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton, NJ: Princeton University Press.

[13] De Gruijter, D.N.M. (1988). The Rasch model and missing data, with an emphasis on tailoring test items. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

[14] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B*, 56, 463–474.

[15] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7, 1–26.

[16] Efron, B. (1994). Missing data, imputation, and the bootstrap. *Journal of American Statistical Association*, 89, 463–479.

[17] Efron, B., and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statitical accuracy, *Statistical Science*, 1, 54–77.

[18] Fischer, G.H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika*, 46, 59–77.

[19] Fischer, G.H., and Parzer, P. (1991). An extension of the rating-scale model with an application to the measument of change. *Psychometrika*, 56, 637–651.

[20] Fischer, G.H., and Ponocny, I. (1994). An extension of the partial credit model with an application to the measument of change. *Psychometrika*, 59, 177–192.

[21] Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society*, A, 144, 285–307.

[22] Graham, J.W., Hofer, S.M., and MacKinnon, D.P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, 31, 197–218.

[23] Heitjan, D.F. (1997). Annotation: What can be done about missing data? Approaches to imputation. *American Journal of Public Health*, 87, 548–550.

[24] Huang, R. (2001). PhD dissertation, University of Alberta.

[25] Huisman, M., and Molenaar, I.W. (2001). Imputation of missing scale data with item response models. In A. Boomsma, M.A.J. van Duijn, and T.A.B. Snijders (Eds.), *Essays on item response theory*, 221-244. New York, NY: Springer-Verlag.

[26] Jacobsen, M. (1989). Existence and unicity of MLEs in discrete exponential family distributuions. *Scandinavian Journal of Statistyics*, 16, 335–349.

[27] Kelderman, H., and Rijkes, C.P.M. (1994). Loglinear multidimentional IRT models for polytomously scored items. *Psychometrika*, 59, 147–177.

[28] Kristjanson, L., Leis, A., Koop, P., Carrière, K.C., and Mueller, B. (1997). Family Members' Care Expectations, Care Perceptions and Satisfaction with Advanced Cancer Care: A Pilot Study. *Journal of Palliative Care*, 13, 5–13.

[29] Lehmann, E.L. (1986). *Testing Statistical Hypotheses.* New York, NY: Springer–Verlag.

[30] Lehmann, E.L. (1999). *Elemetns of Large-Sample Theory.* New York, NY: Springer–Verlag.

[31] Lehmann, E.L., and Casella, G. (1998). *Theory of Point Estimation.* New York, NY: Springer–Verlag.

[32] Liang, K.Y., and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.

[33] Linacre, J.M. (1999). Understanding Rasch Measurement. *Journal of Outcome Measurement.* 3, 382–405.

[34] Little, R.J.A., and Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* New York, NY: John Wiley & Sons.

[35] Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.

[36] Miller, C.R., Sabagh, G., and Dingman, H.F. (1962). Latent class analysis and differential mortality, *Journal of the American Statistical Association*, 57, 430–438.

[37] Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44, 443–460.

[38] Patz, R.J., and Junker, B.W. (1999). Applications and extensions of MCMC in IRT: multiple item types, missing data, and rated responses. *Journal of Educational and Behaviour Statistics*, 24, 342–366.

[39] Poon, W.Y., and Lee, S.Y. (1987). Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficient. *Psychometrika*, 52, 409–430.

[40] Poon, W.Y., Lee, S.Y., and Bentler, P.M. (1990). Pseudo maximum likelihood estimation of multivariate polychoric and polyserial correlations. In *Proceedings of the Annual Meeting of the American Statistical Association*, 291–296.

[41] Ramsay, F.L., and Schafer, D.W. (1996). *The Statistical Sleuth: a Course in Methods of Data Analysis*. Belmont, CA: Duxbury Press.

[42] Rao, C.R. (1965). *Linear Statistical Inference and Its Applications*. New York: John Wiley & Sons.

[43] Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* [Danish Institute of Educational Research 1960, University of Chicago Press 1980, MESA Press 1993] Chicago, IL: MESA Press.

[44] Rasch, G. (1972). Objektivitet i samfundsvidenskaberne et metodeproblem. Paper Presented at the University of Copenhagen.

[45] Ronning, G., and Kukuk, M. (1996) Efficient estimation of ordered probit models. *Journal of American Statistical Association,* 91, 1120–1129.

[46] Rubin, D.B. (1978). Multiple Imputations in Sample Surveys – a Pheonomenological Bayesian Approach to Nonresponse. In *Proceedings of the Survey Research Methods Section, American Statistical Association,* pp 20–34.

[47] Schenker, N., and Welsh, A. H. (1988). Asymptotic results for multiple imputation. *Annals of Statistics*, 16, 1550–1566.

[48] Shao, J., and Sitter, R.R. (1996). Bootstrap for Imputed Survey Data. *Journal of American Statistical Association*, 91, 1278–1288.

[49] Shao, J., and Tu, D. (1995). *The jackknife and Bootstrap.* New York, NY: Springer–Verlag.

[50] Sheng, X., and Carrière, K. C. (2002). An Improved CML Estimation Procedure for the Rasch Model with Item Response Data. *Statistics in Medicine*, 21, 407–416.

[51] Stegelmann, W. (1983). Expanding the Rasch model to a general model having more than one dimension. *Psychometrika*, 48, 259–267.

[52] Thurstone, L.L. (1931). Measurement of social attitudes. *Journal of Abnormal and Social Psychology*, 26, 249–269.

[53] Tukey, J. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics*, 29, 614.

[54] Wright, B.D. (1997). *MESA Memo 64: Measurement for Psychology.* MESA Psychometric Laboratory, University of Chicago.

[55] Wright, B.D., and Douglas, G.A. (1977). Conditional versus unconditional procedures for sample-free item analysis. *Educational and Psychological Measurement*, 37, 47–60.

[56] Wright, B.D., and Masters, G.N. (1982). *Rating Scale Analysis: Rasch Measurement.* Chicago, IL: MESA Press.

[57] Wothke, W. (1998). Longitudinal and multi-group modeling with missing data. In T.D. Little, K.U. Schnabel, and J. Baumert (Eds.), *Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples.* Mahwah, NJ: Lawrence Erlbaum Associates.

[58] Zeger, S.L., and Liang, K.Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42, 121–130.