A Composite Likelihood Approach for
Factor Analyzing Ordinal Data

Dissertation

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Nuo Xi, M.A., M.S.
Graduate Program in Psychology

The Ohio State University
2011

Dissertation Committee:
Dr. Michael W. Browne, Advisor
Dr. Nancy E. Betz
Dr. Michael C. Edwards
Abstract

Ordinal variables are widely used in psychological and educational assessment. One popular model for analyzing ordinal variables is based on a modification of the standard factor analysis model. This modified factor analysis model postulates a continuous latent response variable that corresponds to each observed ordinal variable and then specifies a usual factor analysis model on the item-level latent response variables. Various methodologies are available for estimating parameters in the modified factor analysis model and they usually involve two to three computational steps.

This dissertation considers the Underlying Bivariate Normal (UBN) approach for estimating the modified factor analysis model from ordinal data. The UBN method relies on univariate and bivariate margins only, and it estimates all the model parameters in a single step. As a limited-information procedure, the UBN approach has asymptotic properties that are different from full-information Maximum Likelihood Estimation (MLE) procedure and requires theoretical results that are specifically developed for limited-information approaches. The current research applies the Godambe information matrix, instead of the usual Fisher information matrix, to obtain standard error estimates and evaluates the model fit by a residual based quadratic form test statistic.
The UBN approach is illustrated by using both simulation studies and two real data examples. Performance of the UBN approach is evaluated by checking the closeness between the UBN estimates and the corresponding true population values. The UBN estimates are also compared with results given by the Expectation-Maximization (EM) algorithm and the Metropolis-Hastings Robbins-Monro (MH-RM) method within the Item Response Theory (IRT) framework. Simulation results show that: (1) the UBN approach is able to recover the true population values of model parameters; (2) the UBN estimates become more precise when sample size increases; (3) the standard error estimates given by the Godambe information matrix are consistent with the observed empirical standard errors; (4) the residual based test statistic possesses the assumed asymptotic properties and is useable as a goodness-of-fit test statistic. The two real data illustrations demonstrate that the UBN approach is feasible for large scale real world problems.
Dedicated to my parents
Acknowledgments

I am greatly indebted to my advisor, Dr. Michael W. Browne, for his guidance on the current research and his constant support through my graduate study, without which the dissertation would not have been possible.

My gratitude goes to Dr. Nancy E. Betz and Dr. Michael C. Edwards, for their insightful comments and helpful suggestions on the dissertation. I wish to thank Dr Stephen H.C. du Toit for his kind help in offering the IRTPRO package and the Quality of Life data set. I want to thank Dr Li Cai for his help in explaining multiple technical issues I encountered when using IRTPRO. I also want to thank Dr Hao Wu and Dr Guangjian Zhang for their advice on improving the quality of this research.

I wish to thank Chih-Lin Li for encouraging and entertaining me during the whole process of our graduate study. I also thank my parents for their support, patience, and love.

This work was partially supported by the Society of Multivariate Experimental Psychology (SMEP) Dissertation Award and Scientific Software International (SSI) funding for program development.
Vita

2005 ................................. B.S. Statistics, Peking University, China
2008 ................................. M.A. Psychology, The Ohio State University
2009 ................................. M.S. Statistics, The Ohio State University
2005 - present .................. Graduate Teaching Associate and Statistical Consultant
Department of Psychology, The Ohio State University

Fields of Study

Major Field: Psychology

Area of Concentration: Quantitative Psychology
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xi</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Ordinal Variable</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Modified Factor Analysis Model</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Common Estimation Procedures</td>
<td>4</td>
</tr>
<tr>
<td>2. The UBN Approach</td>
<td>8</td>
</tr>
<tr>
<td>2.1 The UBN Fit Function</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Covariance Matrix $P$</td>
<td>11</td>
</tr>
<tr>
<td>2.3 Estimating Model Parameters</td>
<td>13</td>
</tr>
<tr>
<td>2.3.1 Gradient Vector of $F_{ubn}(\gamma)$</td>
<td>13</td>
</tr>
<tr>
<td>2.3.2 Expected Hessian Matrix of $F_{ubn}(\gamma)$</td>
<td>18</td>
</tr>
<tr>
<td>2.3.3 Equality and Inequality Constraints</td>
<td>19</td>
</tr>
<tr>
<td>2.3.4 Constraint Function Jacobian Matrix</td>
<td>20</td>
</tr>
<tr>
<td>2.3.5 Constrained Optimization Procedure</td>
<td>22</td>
</tr>
<tr>
<td>2.4 Comparison between UBN and Other Approaches</td>
<td>23</td>
</tr>
<tr>
<td>3. Composite Likelihood</td>
<td>25</td>
</tr>
</tbody>
</table>
Appendices:

A. Tables and Figures .................................................. 79
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Simulation results for the one-factor model at $N = 200$</td>
<td>79</td>
</tr>
<tr>
<td>A.2 Simulation results for the one-factor model at $N = 1000$</td>
<td>80</td>
</tr>
<tr>
<td>A.3 Population threshold values of the two-factor model</td>
<td>80</td>
</tr>
<tr>
<td>A.4 Factor loadings of the independent-cluster model</td>
<td>81</td>
</tr>
<tr>
<td>A.5 Factor loadings of the bi-factor model</td>
<td>82</td>
</tr>
<tr>
<td>A.6 Summary of computation time study</td>
<td>82</td>
</tr>
<tr>
<td>A.7 The UBN estimates of the PSYCH101 data set</td>
<td>83</td>
</tr>
<tr>
<td>A.8 The UBN factor loading estimates of the Quality of Life data set</td>
<td>84</td>
</tr>
<tr>
<td>A.9 The UBN threshold estimates of the Quality of Life data set</td>
<td>85</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1</td>
<td>86</td>
</tr>
<tr>
<td>A.2</td>
<td>87</td>
</tr>
<tr>
<td>A.3</td>
<td>88</td>
</tr>
<tr>
<td>A.4</td>
<td>89</td>
</tr>
<tr>
<td>A.5</td>
<td>90</td>
</tr>
<tr>
<td>A.6</td>
<td>91</td>
</tr>
<tr>
<td>A.7</td>
<td>92</td>
</tr>
<tr>
<td>A.8</td>
<td>93</td>
</tr>
<tr>
<td>A.9</td>
<td>93</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Ordinal Variable

Ordinal variables are commonly encountered in educational, social, and behavioral research. An example is the 5-point Likert scale (Likert, 1932) that is widely used in opinion surveys. A person’s response can be classified into one of the following five categories, “Strongly Disagree”, “Disagree”, “Neutral”, “Agree”, and “Strongly Agree”, usually scored by numbers 1 to 5. It is assumed that the response categories of an ordinal variable have a natural order to them, as is shown in the Likert scale example, and a higher value represents more of a characteristic than a lower value. This characteristic may have several ordinal variables as its indicators so that more information with respect to the characteristic can be collected.

Ordinal variables provide order of the response categories, but they do not establish numeric difference between categories. Thereby the numeric values assigned to each response category have no metric meaning but only represent the ordered categories a subject selects. In fact, one can use other labels, such as alphabetic characters, to record the response categories as long as the ordered information is retained. For this reason, arithmetic operations on the numeric values, like addition and substraction, are inappropriate for ordinal variables without additional assumptions. Similarly, the usual linear regression is not appropriate to apply on ordinal
variables directly. Special models and estimation procedures have been developed to analyze ordinal variables.

1.2 Modified Factor Analysis Model

One popular model for analyzing multivariate ordinal variables is based on a modification of the standard factor analysis model. This modified factor analysis model postulates an unobserved continuous variable, the latent response variable, that corresponds to each ordinal variable. A specific response category of an ordinal variable is chosen if and only if the corresponding latent response variable falls between the two thresholds that define this category. The usual factor analysis model is then specified for the latent response variables.

The item-level latent response variables are necessary because a proper implementation of factor analysis model assumes that the dependent variables are continuous and the unique factors are normally distributed. Such an assumption is usually not suitable for ordinal variables, but it will be properly met for continuous latent response variables.

To specify the modified factor analysis model mathematically, let \( x = (x_1, \ldots, x_q)' \) denote a vector of \( q \) observed ordinal variables and \( k_i \) denote the number of ordered response categories that the \( i \)th ordinal variable has. Write \( x_i = a_i \) to mean that the ordered category \( a_i \) is chosen for the \( i \)th ordinal variable \( x_i \), where \( a_i \in \{1, \ldots, k_i\} \). The actual numbers used to label the categories are arbitrary and irrelevant as long as the ordinal outcomes are retained. Assume a continuous latent response variable \( x_i^* \) that corresponds to the observed ordinal variable \( x_i \) with the relationship,

\[
x_i = a_i \iff \tau_{a_i-1}^{(i)} < x_i^* \leq \tau_{a_i}^{(i)}, \quad a_i = 1, \ldots, k_i,
\]

(1.1)
where
\[ \tau_0^{(i)} = -\infty, \tau_1^{(i)} < \tau_2^{(i)} < \ldots < \tau_{k_i-1}^{(i)}, \tau_{k_i}^{(i)} = +\infty, \]  

(1.2)

are thresholds on the latent response variable continuum and they define the \( k_i \) categories of \( x_i \). For a variable \( x_i \) with \( k_i \) categories, there are \( k_i - 1 \) unknown threshold parameters, i.e., \( \tau_1^{(i)}, \tau_2^{(i)}, \ldots, \tau_{k_i-1}^{(i)} \).

Since only ordinal information is available, the scale of latent response variable is indeterminate (Muthén, 1984). Without loss of generality, set the mean and variance of each \( x_i^* \) to zero and one, respectively. Let \( x^* = (x_1^*, \ldots, x_q^*)' \) denote a \( q \)-dimensional vector of latent response variables. Assume \( x^* \) follow a \( q \)-dimensional multivariate normal distribution with a null mean vector and a \( q \times q \) covariance matrix \( P \), i.e., \( x^* \sim N_q(0, P) \), where the diagonal elements of \( P \) are all equal to one. Then specify a standard factor analysis model on \( x^* \),

\[ x^* = \Lambda z + u, \]  

(1.3)

where \( z = (z_1, \ldots, z_m)' \) is an \( m \)-dimensional \((m < q)\) vector of latent factors, \( \Lambda \) is a \( q \times m \) matrix containing factor loadings, and \( u = (u_1, \ldots, u_q)' \) is a \( q \)-dimensional vector of unique factors that may include specific factors and measurement errors. Vectors \( z \) and \( u \) are assumed to be independently and normally distributed. The difference between the standard factor analysis model and this modified version is that in the standard factor analysis model, vector \( x^* \) consists of directly observed manifest variables but here it is unobserved.

It follows directly from data model (1.3) that each latent response variable is a linear combination of latent factors and unique factors,

\[ x_i^* = \lambda_{i1}z_1 + \lambda_{i2}z_2 + \ldots + \lambda_{im}z_m + u_i, \quad i = 1, \ldots, q, \]  

(1.4)
where factor loading $\lambda_{ij}$ represents the influence of the $j$th latent factor $z_j$ on the $i$th latent response variable $x_i^*$. Having $x_i^*$, the outcome of the corresponding ordinal variable $x_i$ is determined by categorizing $x_i^*$ into the response categories $1$ to $k_i$, according to the thresholds in (1.1). This shows that in the modified factor analysis model each latent response variable is directly influenced by the latent factors and the corresponding ordinal variable is indirectly affected by the latent factors, through the usage of latent response variable.

1.3 Common Estimation Procedures

Several approaches have been proposed by different authors to estimate parameters in the modified factor analysis model, i.e., thresholds, factor loadings, and factor correlations.

Using all the information in the data, Lee, Poon and Bentler (1990) developed a full simultaneous Maximum Likelihood (ML) estimation procedure to estimate the modified factor analysis model. This full-information estimation procedure requires calculating a $q$-dimensional integration over $x^*$ to obtain the probability function of a full response vector $x$. Lee et al. (1990) apply the Fisher scoring algorithm to compute the full information MLE. They provide first derivatives of the full likelihood with respect to model parameters and the Fisher information matrix, which are necessary in applying the Fisher scoring algorithm. Also, Lee et al. (1990) apply the likelihood ratio criterion to test overall model fit.

Lee et al. (1990) regarded this full-information approach as a classical optimal solution and a standard for comparison of other less optimal procedures. However, the involved $q$-dimensional integral, although it utilizes all the information in the data, is extremely difficult to evaluate when $q$ is large, which is not uncommon in practice.
Jöreskog and Moustaki (2001) commented that the full-information approach is not computationally feasible for models with $q$ larger than 4.

To overcome the difficulty with evaluating a $q$-dimensional integration, methods based on only univariate and bivariate margins have been introduced. In particular, Muthén (1984) discusses a three-stage, limited-information, Generalized Least Square (GLS) estimator for analyzing general structural equation models to ordinal variables, of which the modified factor analysis model is a special case. In the first stage, the threshold parameters are estimated from the univariate margins of the data. In the second stage, an estimate of the correlation between latent response variables, the polychoric correlation, is derived from the bivariate margins of ordinal variables given the threshold estimates. The estimates obtained in the first two stages are limited-information ML solutions. Finally, the factor loadings are estimated by Weighted Least Squares (WLS) using a weight matrix that is the inverse of the asymptotic covariance matrix of polychoric correlations. One problem with the asymptotic covariance matrix is that it is often unstable in small samples, particularly if there are zeros or small frequencies in the bivariate margins (Jöreskog & Moustaki, 2001).

De Leon (2005) proposed another multi-stage estimation procedure, the Maximum Pairwise Likelihood (MPL) approach. The Pairwise Likelihood is a product of bivariate marginal likelihoods, which depend on every possible pair of the observed ordinal variables. The maximizer of Pairwise Likelihood is called the Maximum Pairwise Likelihood Estimator (MPLE). The MPL procedure combines the first two stages of Muthén’s approach into one: the thresholds and polychoric correlations are estimated simultaneously by maximizing the Pairwise Likelihood. In the second stage,
the factor loadings are obtained through the GLS approach using the asymptotic covariance matrix of polychoric correlations as the weight matrix. More details of the MPL approach are provided by Liu (2007).

Both Muthén’s three-stage approach and the MPL approach involve computing polychoric correlation estimates in a separate stage and obtaining the factor loading estimates through GLS. In contrast, Jöreskog and Moustaki (2001) consider a single-stage limited-information method, the Underlying Bivariate Normal (UBN) approach, to factor analyze ordinal variables. They define a UBN fit function that is the sum of all univariate and bivariate fit functions. The factor loadings and thresholds are estimated by minimizing the newly defined UBN fit function in one single stage. Polychoric correlations are not estimated in the UBN approach, and no subsequent GLS stage is involved.

Here, it is proven that minimizing the UBN fit function is equivalent to maximizing the sum of all univariate and bivariate log likelihoods. Hence, UBN can be regarded as a special case of composite likelihood, a likelihood formed by adding together individual component likelihoods, each of which corresponds to a marginal or conditional event (Lindsay, 1988). The Pairwise Likelihood used in MPL also belongs to the composite likelihood family. In general, composite likelihood approaches avoid the severe computational burden of calculating a full information likelihood function by relying on marginal or conditional likelihoods that only require low-dimensional integrations. They are limited-information methodologies. Composite likelihood estimates, i.e., the parameter estimates that maximize a composite likelihood, are proven to be consistent and asymptotically normal. However, the corresponding statistical inference on the composite likelihood estimates require theoretical development that
take the information loss into consideration. This topic will be discussed in this dissertation.

The purpose of current dissertation is to consider the UBN approach for factor analyzing ordinal variables and investigate properties of the UBN estimates. It is of particular interest to develop theoretical result for estimating standard errors of the UBN estimates and testing the goodness of model fit with limited information. The remainder of the dissertation proceeds as follows. Chapter 2 gives a brief introduction of the UBN approach and discusses computational details in minimizing the UBN fit function. Chapter 3 describes the composite likelihood and provides a formal proof of UBN being a special case of the composite likelihood. Chapter 4 provides detailed information on estimating standard errors and evaluating the model fit with a marginal residual based goodness-of-fit test statistic. Chapter 5 includes three simulation studies that focus on investigating different aspects of the UBN estimates and the associated test statistic. Chapter 6 applies the UBN approach to two real data examples. Chapter 7 concludes the dissertation with a summary and some discussion.
Chapter 2: The UBN Approach

The Underlying Bivariate Normal (UBN) approach proposed by Jöreskog and Moustaki (2001) recovers model parameters by minimizing a weighted sum of all univariate and bivariate fit functions, which is referred to as the UBN fit function. This fit function compares the difference between observed sample proportions and model fitted probabilities in univariate and bivariate margins. It can be written as a log likelihood ratio between a saturated model and a hypothesized model, and the objective is to minimize the log likelihood ratio between these two models and obtain parameter estimates.

2.1 The UBN Fit Function

Suppose that we have \( q \) ordinal variables \( x_1, \ldots, x_q \). Correspondingly, there are \( q \) latent response variables \( x^*_1, \ldots, x^*_q \), distributed as \( N_q(0, \Phi) \). A standard factor analysis model (1.3) is specified for these latent response variables,

\[
x^* = \Lambda z + u,
\]

where \( \Lambda \) is a \( q \times m \) factor loading matrix, latent factors \( z_1, \ldots, z_m \) are distributed as \( N_m(0, \Phi) \), unique factors \( u_1, \ldots, u_q \) are distributed as \( N_q(0, \Psi) \), and the latent factors are independent from the unique factors. In order to identify the model, the diagonal elements of \( P \) and \( \Phi \) are set to one. The off-diagonal elements of \( P \) and \( \Phi \) are therefore
correlation coefficients. Let $\gamma$ represent a vector of all model parameters,

$$\gamma = (\tau', \lambda', \phi')' \quad (2.1)$$

where vector $\tau$ includes thresholds associated with $q$ ordinal variables, vector $\lambda$ consists of elements in the factor loading matrix $\Lambda$, and vector $\phi$ contains non-duplicated off-diagonal elements in $\Phi$. When the latent factors are uncorrelated, $\Phi$ is an identity matrix and $\gamma$ only contains $\tau$ and $\lambda$.

The UBN approach estimates model parameters by comparing the observed sample proportions to model fitted probabilities in univariate and bivariate margins. Let $p_a^{(i)}$ denote the sample proportion of a response in category $a$ of variable $i$ and $\pi_a^{(i)}(\gamma)$ be the corresponding univariate marginal probability,

$$\pi_a^{(i)}(\gamma) = Pr(x_i = a) = \int_{\tau_{a-1}^{(i)}}^{\tau_a^{(i)}} \phi_1(x_i^*)dx_i^*, \quad (2.2)$$

where $\phi_1(\cdot)$ is the standard normal density function. Let $p_{ab}^{(ij)}$ denote the sample proportion of a response in category $a$ of variable $i$ and category $b$ of variable $j$ and $\pi_{ab}^{(ij)}(\gamma)$ be the corresponding bivariate marginal probability,

$$\pi_{ab}^{(ij)}(\gamma) = Pr(x_i = a, x_j = b) = \int_{\tau_{a-1}^{(i)}}^{\tau_a^{(i)}} \int_{\tau_{b-1}^{(j)}}^{\tau_b^{(j)}} \phi_2(x_i^*, x_j^*|\rho_{ij})dx_i^*dx_j^*, \quad (2.3)$$

where $\phi_2(\cdot, \cdot | \rho_{ij})$ is the density function of standardized bivariate normal distribution and $\rho_{ij}$ is the correlation between $x_i^*$ and $x_j^*$.

Equations (2.2) and (2.3) are derived from the latent response variables’ distribution $N_q(0, P)$ and the relationship between ordinal variables and latent response variables, i.e., equation (1.1). They show that probabilities $\pi_a^{(i)}(\gamma)$ and $\pi_{ab}^{(ij)}(\gamma)$ are functions of the model parameters specified in (2.1). In contrast, the corresponding observed sample proportions $p_a^{(i)}$ and $p_{ab}^{(ij)}$ only depend on the ordinal data. If the modified factor analysis model fits perfectly, the observed sample proportions are
expected to be close to the corresponding model fitted probabilities, given sampling errors. If not, we expect to see some discrepancy between them.

The UBN fit function defined by Jöreskog and Moustaki (2001) is such a fit function that measures the difference between observed sample proportions and model fitted probabilities,

\[
F_{ubn}(\gamma) = \sum_{i=1}^{q} \sum_{a=1}^{k_i} P_a^{(i)} \ln \left[ \frac{P_a^{(i)}}{\pi_a^{(i)}(\gamma)} \right] + \sum_{i=2}^{q} \sum_{j=1}^{i-1} \sum_{a=1}^{k_i} \sum_{b=1}^{k_j} P_{ab}^{(ij)} \ln \left[ \frac{P_{ab}^{(ij)}}{\pi_{ab}^{(ij)}(\gamma)} \right].
\] (2.4)

It is a sum of all univariate and bivariate fit functions and only involves one- and two-dimensional integrations (2.2) and (2.3). Let \( N \) denote the sample size, \( n_a^{(i)} = N p_a^{(i)} \) denote the observed sample frequency of a response in category \( a \) of variable \( i \), and \( n_{ab}^{(ij)} = N p_{ab}^{(ij)} \) denote the observed sample frequency of a response in category \( a \) of variable \( i \) and category \( b \) of variable \( j \). The UBN fit function can be written as

\[
\frac{1}{N} \left[ \ln \left( \prod_{i=1}^{q} \prod_{a=1}^{k_i} \left( p_a^{(i)} \right)^{n_a^{(i)}} \right) + \ln \left( \prod_{i=2}^{q} \prod_{j=1}^{i-1} \prod_{a=1}^{k_i} \prod_{b=1}^{k_j} \left( p_{ab}^{(ij)} \right)^{n_{ab}^{(ij)}} \right) \right]
\]

and it has two terms. Each term is a log likelihood ratio between a saturated model, in which every cell in the univariate/bivariate margins is fitted perfectly, and a hypothesized model. Because the saturated model always fits perfectly, neither of the two terms can be negative. Hence \( F_{ubn}(\gamma) \) is non-negative. In addition, it will be equal to zero if all the model fitted probabilities are equal to the corresponding observed sample proportions.

Therefore the objective is to find a set of parameter values that make \( F_{ubn}(\gamma) \) as close to zero as possible. The obtained parameter estimate is called the UBN estimate and represented by \( \hat{\gamma}_{ubn} \).
2.2 Covariance Matrix $P$

The correlation $\rho_{ij}$ seen in (2.3) is not a model parameter. Rather, it will be a function of model parameters. The data model (1.3) implies the corresponding covariance structure,

$$P = \Lambda \Phi \Lambda' + \Psi,$$

(2.5)

where $P$, $\Phi$, and $\Psi$ are covariance matrices of the latent response variables $x^*$, latent factors $z$, and unique factors $u$, respectively.

Let $[\cdot]_{ij}$ represent the element on the $i$th row and $j$th column of a designated matrix. Because the diagonal elements of $P$ are set to one for the purpose of model identification,

$$\rho_{ii} = [P]_{ii} = 1, \quad i = 1, \ldots, q,$$

(2.6)

the off-diagonal elements of $P$ are correlation coefficients between the latent response variables. It is further assumed that the unique factors are independent from each other, making $\Psi$ a diagonal matrix. Therefore, the covariance structure (2.5) implies that,

$$\rho_{ij} = [\Lambda \Phi \Lambda']_{ij}, \quad i \neq j.$$

(2.7)

Thus the correlation coefficient $\rho_{ij}$ involved in calculating $\pi^{(ij)}(\gamma)$ is a function of the model parameters.

Since the diagonal elements of $P$ are equal to one and the covariance matrix $\Psi$ is a diagonal matrix, the covariance structure (2.5) also indicates that

$$\Psi = \text{Diag}[I - \Lambda \Phi \Lambda'],$$

(2.8)
meaning that the diagonal elements of $\Psi$ also depend on the model parameters. To summarize, the ordinal variables’ thresholds, factor loadings in $\Lambda$, and off-diagonal elements of $\Phi$ are unknown model parameters, as is shown in (2.1). The matrices $P$ and $\Psi$ are functions of the model parameters, as is seen in (2.7) and (2.8).

When the latent factors are uncorrelated, i.e., $\Phi$ is an identity matrix, functions (2.7) and (2.8) can be further simplified. Now the model parameter vector $\gamma$ only contains thresholds and factor loadings. The correlation between two latent response variables is

$$\rho_{ij} = \left[\Lambda \Lambda^\prime\right]_{ij}, \quad i \neq j,$$

and the covariance matrix $\Psi$ is,

$$\Psi = \text{Diag}[I - \Lambda \Lambda^\prime].$$

Hence, $P$ and $\Psi$ only depend on factor loadings.

Correlation between two unobservable latent response variables, $\rho_{ij}$, is not directly measurable. The polychoric correlation is a popular technique for estimating $\rho_{ij}$ between two theorized normally distributed continuous latent variables from two observed ordinal variables. The multi-stage procedures introduced in Chapter 1 apply the polychoric correlation technique to estimate the true correlation $\rho_{ij}$ in one separate stage and use the asymptotic covariance matrix of polychoric correlations as the weight matrix in the following GLS stage. Because the polychoric correlations are obtained using different marginals, the resulting correlation matrix may not be non-negative definite, especially when the sample size is small and there are many ordinal variables (Yuan, Wu, & Bentler, 2011). In contrast, the UBN approach does not apply this polychoric correlation technique and computes $\rho_{ij}$ directly by using (2.7),
given factor loadings and factor correlations. The estimated correlation matrix $P$ is non-negative definite as long as $\Phi$ is non-negative definite.

2.3 Estimating Model Parameters

The current study applies a constrained Fisher scoring type of algorithm (c.f., Browne & du Toit, 1992; Aitchison & Silvey, 1960) to minimize the UBN fit function subject to appropriate equality and inequality constraints. The usual Fisher scoring algorithm is related to Newton’s method and is used to optimize the likelihood function numerically. It finds successively better solution using a gradient and expected Hessian matrix. Because equality and inequality constraints must be imposed on model parameters, as will be discussed in more detail later, we apply the constrained optimization method described in Browne and du Toit (1992, pp. 272-277).

Besides gradient function and expected Hessian matrix of the UBN fit function, this optimization approach also requires equality and inequality constraints, as well as a constraint function Jacobian matrix. The required information together with a brief introduction of the optimization procedure will be given in this section.

2.3.1 Gradient Vector of $F_{ubn}(\gamma)$

Let $g(\gamma)$ denote the gradient vector of the UBN fit function $F_{ubn}(\gamma)$. Given the definition of $F_{ubn}(\gamma)$ specified in (2.4), the gradient vector $g(\gamma)$ is,

$$
g(\gamma) = \frac{\partial F_{ubn}(\gamma)}{\partial \gamma} = - \sum_{i=1}^{q} \sum_{a=1}^{k_i} \frac{k_i}{p_a^{(i)}(\gamma)} \frac{\partial \pi_a^{(i)}(\gamma)}{\partial \gamma} - \sum_{i=2}^{q} \sum_{j=1}^{i-1} \sum_{a=1}^{k_i} \sum_{b=1}^{k_j} \frac{k_j}{p_{ab}^{(ij)}(\gamma)} \frac{\partial \pi_{ab}^{(ij)}(\gamma)}{\partial \gamma}.
$$

(2.9)

It can be seen that the gradient vector is a weighted sum of the first derivatives of model fitted probabilities $\pi_a^{(i)}(\gamma)$ and $\pi_{ab}^{(ij)}(\gamma)$ with respect to $\gamma$, where elements of $\gamma$
are specified in (2.1). To compute the gradient vector, one needs to calculate elements in vectors \( \frac{\partial \pi_i^a(\gamma)}{\partial \gamma} \) and \( \frac{\partial \pi_{ab}^{ij}(\gamma)}{\partial \gamma} \).

Two possible situations are under discussion. The first situation is when the latent factors are uncorrelated, i.e., \( \Phi \) is an identity matrix. The model parameter vector \( \gamma \) is

\[
\gamma = (\tau', \lambda')',
\]

where vector \( \tau \) contains thresholds and vector \( \lambda \) contains factor loadings in \( \Lambda \).

Equation (2.2) shows that the univariate marginal probability \( \pi_i^a(\gamma) \) only involves two thresholds \( \tau_{a-1}^i \) and \( \tau_a^i \). Therefore the vector \( \frac{\partial \pi_i^a(\gamma)}{\partial \gamma} \) only has two nonzero elements, and they are

\[
\frac{\partial \pi_i^a(\gamma)}{\partial \tau_{a-1}^i} = -\phi_1(\tau_{a-1}^i), \tag{2.10}
\]

\[
\frac{\partial \pi_i^a(\gamma)}{\partial \tau_a^i} = \phi_1(\tau_a^i). \tag{2.11}
\]

The elements in \( \frac{\partial \pi_{ab}^{ij}(\gamma)}{\partial \gamma} \) with respect to parameters other than \( \tau_{a-1}^i \) and \( \tau_a^i \) are equal to zero.

To compute the bivariate marginal probability \( \pi_{ab}^{ij}(\gamma) \), it involves four associated thresholds, \( \tau_{a-1}^i, \tau_a^i, \tau_{b-1}^j, \tau_b^j \), and the correlation between \( x_i^a \) and \( x_j^b \), \( \rho_{ij} \), which is equal to

\[
\rho_{ij} = [\Lambda\Lambda']_{ij} = \sum_{l=1}^m \lambda_{il}\lambda_{jl}.
\]

Olsson (1979) shows that

\[
\pi_{ab}^{ij} = \Phi_2(\tau_a^i, \tau_b^j | \rho_{ij}) - \Phi_2(\tau_{a-1}^i, \tau_b^j | \rho_{ij}) - \Phi_2(\tau_a^i, \tau_{b-1}^j | \rho_{ij}) + \Phi_2(\tau_{a-1}^i, \tau_{b-1}^j | \rho_{ij}), \tag{2.12}
\]
\[
\frac{\partial \Phi_2(u, v|\rho)}{\partial u} = \phi_1(u) \cdot \Phi_1\left\{\frac{v - \rho u}{(1 - \rho^2)^{1/2}}\right\}, \tag{2.13}
\]
\[
\frac{\partial \Phi_2(u, v|\rho)}{\partial v} = \phi_1(v) \cdot \Phi_1\left\{\frac{u - \rho v}{(1 - \rho^2)^{1/2}}\right\}, \tag{2.14}
\]
\[
\frac{\partial \Phi_2(u, v|\rho)}{\partial \rho} = \phi_2(u, v|\rho), \tag{2.15}
\]

where \(\phi_2(\cdot, \cdot|\rho)\) is the density function of the standardized bivariate normal distribution with correlation \(\rho\), \(\Phi_2(\cdot, \cdot|\rho)\) is the corresponding distribution function, \(\phi_1(\cdot)\) is the standard normal density function, and \(\Phi_1(\cdot)\) is the corresponding distribution function. Equations (2.12) to (2.15) imply that the nonzero elements in \(\frac{\partial \pi^{(ij)}(\gamma)}{\partial \gamma}\) are

\[
\frac{\partial \pi^{(ij)}_{ab}}{\partial \tau^{(i)}_{a-1}} = -\phi_1(\tau^{(i)}_{a-1})\Phi_1\left\{\frac{\tau^{(j)}_b - \rho_{ij}\tau^{(i)}_a}{(1 - \rho_{ij}^2)^{1/2}}\right\} - \Phi_1\left\{\frac{\tau^{(j)}_b - \rho_{ij}\tau^{(i)}_a}{(1 - \rho_{ij}^2)^{1/2}}\right\}, \tag{2.16}
\]
\[
\frac{\partial \pi^{(ij)}_{ab}}{\partial \tau^{(i)}_a} = \phi_1(\tau^{(i)}_a)\Phi_1\left\{\frac{\tau^{(j)}_b - \rho_{ij}\tau^{(i)}_a}{(1 - \rho_{ij}^2)^{1/2}}\right\} - \Phi_1\left\{\frac{\tau^{(j)}_b - \rho_{ij}\tau^{(i)}_a}{(1 - \rho_{ij}^2)^{1/2}}\right\}, \tag{2.17}
\]
\[
\frac{\partial \pi^{(ij)}_{ab}}{\partial \tau^{(j)}_{b-1}} = -\phi_1(\tau^{(j)}_{b-1})\Phi_1\left\{\frac{\tau^{(i)}_a - \rho_{ij}\tau^{(j)}_b}{(1 - \rho_{ij}^2)^{1/2}}\right\} - \Phi_1\left\{\frac{\tau^{(i)}_a - \rho_{ij}\tau^{(j)}_b}{(1 - \rho_{ij}^2)^{1/2}}\right\}, \tag{2.18}
\]
\[
\frac{\partial \pi^{(ij)}_{ab}}{\partial \tau^{(j)}_b} = \phi_1(\tau^{(j)}_b)\Phi_1\left\{\frac{\tau^{(i)}_a - \rho_{ij}\tau^{(j)}_b}{(1 - \rho_{ij}^2)^{1/2}}\right\} - \Phi_1\left\{\frac{\tau^{(i)}_a - \rho_{ij}\tau^{(j)}_b}{(1 - \rho_{ij}^2)^{1/2}}\right\}, \tag{2.19}
\]
\[
\frac{\partial \pi^{(ij)}_{ab}}{\partial \lambda_{il}} = \frac{\partial \pi^{(ij)}_{ab}}{\partial \rho_{ij}} \frac{\partial \rho_{ij}}{\partial \lambda_{il}} = \frac{\partial \pi^{(ij)}_{ab}}{\partial \rho_{ij}} \lambda_{il}, \quad l = 1, \ldots, m, \tag{2.20}
\]
\[
\frac{\partial \pi^{(ij)}_{ab}}{\partial \lambda_{jl}} = \frac{\partial \pi^{(ij)}_{ab}}{\partial \rho_{ij}} \frac{\partial \rho_{ij}}{\partial \lambda_{jl}} = \frac{\partial \pi^{(ij)}_{ab}}{\partial \rho_{ij}} \lambda_{jl}, \quad l = 1, \ldots, m, \tag{2.21}
\]

where

\[
\frac{\partial \pi^{(ij)}_{ab}}{\partial \rho_{ij}} = \phi_2(\tau^{(i)}_a, \tau^{(j)}_b|\rho_{ij}) - \phi_2(\tau^{(i)}_{a-1}, \tau^{(j)}_b|\rho_{ij}) - \phi_2(\tau^{(i)}_a, \tau^{(j)}_{b-1}|\rho_{ij}) + \phi_2(\tau^{(i)}_{a-1}, \tau^{(j)}_{b-1}|\rho_{ij}). \tag{2.22}
\]

The second situation is when the latent factors are correlated to each other, i.e., \(\Phi\) is no longer an identity matrix. To ensure that \(\Phi\) is non-negative definite, we
use the following parameterization of Cholesky decomposition,

\[ \Phi = TD_\alpha T', \quad (2.23) \]

where \( T \) is a lower triangular matrix with unit diagonals,

\[
T = \begin{pmatrix}
1 & & & \\
t_{21} & 1 & & \\
& \ddots & \ddots & \\
t_{m1} & t_{m2} & \cdots & 1
\end{pmatrix},
\]

and \( D_\alpha \) is a diagonal matrix

\[
D_\alpha = \begin{pmatrix}
1 & & & \\
& \alpha_2 & & \\
& & \ddots & \\
& & & \alpha_m
\end{pmatrix}
\]

with inequality constraints

\[ \alpha_i \geq 0, \quad i = 2, \ldots, m, \quad (2.24) \]

to make sure that \( \Phi \) is non-negative definite, and equality constraints

\[ [TD_\alpha T']_{ii} = 1, \quad i = 2, \ldots, m, \quad (2.25) \]

for the purpose of model identification. The covariance matrix \( P \) can be written as,

\[ P = \Lambda \Phi \Lambda' + \Psi = \Lambda^* D_\alpha \Lambda^* + \Psi, \quad (2.26) \]

where \( \Lambda^* = \Lambda T \), and the off-diagonal elements of \( P \) are

\[ \rho_{ij} = \sum_{l=1}^{m} \alpha_l \lambda^*_l \lambda^*_j = \sum_{l=1}^{m} \alpha_l (\sum_{d=1}^{m} \lambda_{dl} t_{dl})(\sum_{d=1}^{m} \lambda_{jd} t_{dl}), \quad i \neq j \quad (2.27) \]

where \( t_{dl} \) represents element in matrix \( T, \ t_{dl} = [T]_{dl} \).
Now the parameter vector $\gamma$ is

$$
\gamma = (\tau', \lambda', t', \alpha')',
$$

where vector $\tau$ contains thresholds, vector $\lambda$ contains factor loadings, vector $t$ includes lower triangular elements in $T$, and vector $\alpha$ consists of diagonal elements $\alpha_2$ to $\alpha_m$ in $D_\alpha$. Notice that there are equality constraints (2.25) imposed on $T$ and $D_\alpha$. The number of independent parameters in the model equals to the number of elements in $\gamma$ minus the number of equality constraints in (2.25).

The first derivatives of $\pi_a(i)(\gamma)$ and $\pi_{ab}(ij)(\gamma)$ with respect to thresholds $\tau$ do not change. The first derivatives of $\pi_{ab}(ij)(\gamma)$ with respect to other parameters do change. According to the chain rule, the first derivatives of $\pi_{ab}(ij)(\gamma)$ with respect to $\lambda$, $t$ and $\alpha$ are,

$$
\frac{\partial \pi_{ab}(ij)}{\partial \lambda_{id}} = \frac{\partial \pi_{ab}(ij)}{\partial \rho_{ij}} \frac{\partial \rho_{ij}}{\partial \lambda_{id}} = \frac{\partial \pi_{ab}(ij)}{\partial \rho_{ij}} \sum_{l=1}^{m} \alpha_l t_{dl} \lambda_{jl}^*, \ d = 1, \ldots, m, \quad (2.28)
$$

$$
\frac{\partial \pi_{ab}(ij)}{\partial \lambda_{jd}} = \frac{\partial \pi_{ab}(ij)}{\partial \rho_{ij}} \frac{\partial \rho_{ij}}{\partial \lambda_{jd}} = \frac{\partial \pi_{ab}(ij)}{\partial \rho_{ij}} \sum_{l=1}^{m} \alpha_l t_{dl} \lambda_{dl}^*, \ d = 1, \ldots, m, \quad (2.29)
$$

$$
\frac{\partial \pi_{ab}(ij)}{\partial t_{dl}} = \frac{\partial \pi_{ab}(ij)}{\partial \rho_{ij}} \frac{\partial \rho_{ij}}{\partial t_{dl}} = \frac{\partial \pi_{ab}(ij)}{\partial \rho_{ij}} \alpha_l (\lambda_{jl}^* \lambda_{id} + \lambda_{il}^* \lambda_{jd}), \ \forall d > l, \quad (2.30)
$$

$$
\frac{\partial \pi_{ab}(ij)}{\partial \alpha_l} = \frac{\partial \pi_{ab}(ij)}{\partial \rho_{ij}} \frac{\partial \rho_{ij}}{\partial \alpha_l} = \frac{\partial \pi_{ab}(ij)}{\partial \rho_{ij}} \lambda_{il}^* \lambda_{jl}^*, \ \forall l = 2, \ldots, m, \quad (2.31)
$$

where $\frac{\partial \pi_{ab}(ij)}{\partial \rho_{ij}}$ has been given in (2.22).

The above equations specify nonzero elements in $\frac{\partial \pi_a(i)(\gamma)}{\partial \gamma}$ and $\frac{\partial \pi_{ab}(ij)(\gamma)}{\partial \gamma}$ under two situations, i.e., uncorrelated factors and correlated factors. By substituting $\frac{\partial \pi_a(i)(\gamma)}{\partial \gamma}$ and $\frac{\partial \pi_{ab}(ij)(\gamma)}{\partial \gamma}$ in (2.9), one can calculate the gradient vector $g(\gamma)$. 

17
2.3.2 Expected Hessian Matrix of $F_{\text{ubn}}(\gamma)$

Whether the latent factors are correlated or not, the Hessian matrix of the UBN fit function is

$$\frac{\partial^2 F_{\text{ubn}}(\gamma)}{\partial \gamma \partial \gamma'} = \sum_{i=1}^{k_i} \sum_{a=1}^{p_a^{(i)}} \left\{ \frac{\pi_a^{(i)}(\gamma) \pi_a^{(i)}(\gamma)}{[\pi_a^{(i)}(\gamma)]^2} \left[ \frac{\pi_a^{(i)}(\gamma)}{\partial \gamma} \frac{\partial^2 \pi_a^{(i)}(\gamma)}{\partial \gamma \partial \gamma'} - \frac{\pi_a^{(i)}(\gamma)}{\partial \gamma} \right] \right\} + \sum_{i=2}^{q} \sum_{j=1}^{k_j} \sum_{a=1}^{k_i} \sum_{b=1}^{k_j} \left\{ \frac{\pi_{ab}^{(ij)}(\gamma)}{\partial \gamma} \frac{\partial^2 \pi_{ab}^{(ij)}(\gamma)}{\partial \gamma \partial \gamma'} - \pi_{ab}^{(ij)}(\gamma) \right\}.$$  

(2.32)

Because

$$\sum_{a=1}^{k_i} \pi_a^{(i)}(\gamma) = 1, \quad \sum_{a=1}^{k_i} \sum_{b=1}^{k_j} \pi_{ab}^{(ij)}(\gamma) = 1,$$

we have

$$\sum_{a=1}^{k_i} \frac{\partial^2 \pi_a^{(i)}(\gamma)}{\partial \gamma \partial \gamma'} = 0, \quad \sum_{a=1}^{k_i} \sum_{b=1}^{k_j} \frac{\partial^2 \pi_{ab}^{(ij)}(\gamma)}{\partial \gamma \partial \gamma'} = 0.$$

Also note that under the model, $E(p_a^{(i)}) = \pi_a^{(i)}(\gamma)$ and $E(p_{ab}^{(ij)}) = \pi_{ab}^{(ij)}(\gamma)$. When taking expectation of the Hessian matrix, the two terms involving the second derivatives of $\pi_a^{(i)}(\gamma)$ and $\pi_{ab}^{(ij)}(\gamma)$ with respect to $\gamma$ in (2.32) can be omitted, because

$$E \left( \sum_{a=1}^{k_i} \frac{p_a^{(i)}}{[\pi_a^{(i)}(\gamma)]^2} \left\{ \pi_a^{(i)}(\gamma) \frac{\partial^2 \pi_a^{(i)}(\gamma)}{\partial \gamma \partial \gamma'} \right\} \right) = \sum_{a=1}^{k_i} E(p_a^{(i)}) \frac{\partial^2 \pi_a^{(i)}(\gamma)}{\partial \gamma \partial \gamma'} = \sum_{a=1}^{k_i} \frac{\partial^2 \pi_a^{(i)}(\gamma)}{\partial \gamma \partial \gamma'} = 0,$$

$$E \left( \sum_{i=2}^{q} \sum_{j=1}^{k_j} \sum_{a=1}^{k_i} \sum_{b=1}^{k_j} \left\{ \pi_{ab}^{(ij)}(\gamma) \frac{\partial^2 \pi_{ab}^{(ij)}(\gamma)}{\partial \gamma \partial \gamma'} \right\} \right) = \sum_{i=2}^{q} \sum_{j=1}^{k_j} E(p_{ab}^{(ij)}) \frac{\partial^2 \pi_{ab}^{(ij)}(\gamma)}{\partial \gamma \partial \gamma'} = \sum_{i=2}^{q} \sum_{j=1}^{k_j} \frac{\partial^2 \pi_{ab}^{(ij)}(\gamma)}{\partial \gamma \partial \gamma'} = 0.$$
Let $H_\gamma$ denote the expected Hessian matrix of $F_{ubn}(\gamma)$, i.e., $H_\gamma = \mathcal{E}(\frac{\partial^2 F_{ubn}(\gamma)}{\partial \gamma \partial \gamma'})$,

$$H_\gamma = \sum_{i=1}^{q} \sum_{a=1}^{k_i} \frac{1}{\pi_a^{(i)}(\gamma)} \frac{\partial \pi_a^{(i)}(\gamma)}{\partial \gamma} \frac{\partial \pi_a^{(i)}(\gamma)}{\partial \gamma'} + \sum_{i=2}^{q} \sum_{j=1}^{k_i} \sum_{a=1}^{k_j} \sum_{b=1}^{k_j} \frac{1}{\pi_{ab}^{(ij)}(\gamma)} \frac{\partial \pi_{ab}^{(ij)}(\gamma)}{\partial \gamma} \frac{\partial \pi_{ab}^{(ij)}(\gamma)}{\partial \gamma'},$$

which can be computed directly by substituting $\frac{\partial \pi_a^{(i)}(\gamma)}{\partial \gamma}$ and $\frac{\partial \pi_{ab}^{(ij)}(\gamma)}{\partial \gamma}$ in (2.33).

### 2.3.3 Equality and Inequality Constraints

Minimizing the UBN fit function is subject to several equality and inequality constraints. When the latent factors are correlated, we use the parameterization (2.23) to model $\Phi$ and apply the equality constraints (2.25)

$$[TD_aT']_{ii} = 1, \quad i = 2, \ldots, m,$$

to identify the model and the inequality constraints (2.24)

$$\alpha_i \geq 0, \quad i = 2, \ldots, m,$$

to ensure that $\Phi$ is non-negative definite. Besides, to make sure that the covariance matrix $\Psi=\text{Diag}[I-\Lambda \Phi \Lambda']$ is non-negative definite, we impose the following inequality constraints on model parameters,

$$[\Lambda \Phi \Lambda']_{ii} = \sum_{l=1}^{m} \alpha_l (\sum_{d=1}^{m} \lambda_{il}^2 t_{dl})^2 \leq 1, \quad i = 1, \ldots, q.$$  

When the latent factors are uncorrelated, $\Phi$ is an identity matrix. Inequality constraints (2.24) and equality constraints (2.25) do not apply. We only have the inequality constraints on factor loadings to make sure that $\Psi$ is non-negative definite,

$$[\Lambda \Lambda']_{ii} = \sum_{l=1}^{m} \lambda_{il}^2 \leq 1, \quad i = 1, \ldots, q.$$  

19
2.3.4 Constraint Function Jacobian Matrix

To apply the optimization procedure described in Browne and du Toit (1992), the first derivatives of the equality and inequality constraints with respect to model parameters are also needed. Write equality and inequality constraints as functions of model parameters that have the form

\[ c_i(\gamma) = 0, \quad i = 1, \ldots, r_1, \]
\[ c_i(\gamma) \geq 0, \quad i = r_1 + 1, \ldots, r_1 + r_2, \]

where \( c_i(\gamma) \) is a continuously differentiable function of \( \gamma \), \( r_1 \) denotes the number of equality constraints and \( r_2 \) denotes the number of inequality constraints. Let \( c(\gamma) \) denote an \( (r_1 + r_2) \times 1 \) vector containing all the equality and inequality constraints,

\[ c(\gamma) = (c_i(\gamma))_{i=1}^{r_1}, \ldots, r_1 + r_2, \quad (2.36) \]

and \( L(\gamma) \) denote an \( (r_1 + r_2) \times w \) constraint function Jacobian matrix,

\[ L(\gamma) = \frac{\partial}{\partial \gamma} c(\gamma), \quad (2.37) \]

where \( w \) denotes the number of parameters in \( \gamma \). The constraint function Jacobian matrix \( L(\gamma) \) will be utilized in the constraint optimization process described in the next subsection. It is worth mentioning that, in the current problem, the inequality constraints (2.24) are imposed as simple bounds on parameters \( \alpha \) and are not involved in \( L(\gamma) \).

When the latent factors are correlated, the constraint vector \( c(\gamma) \) contains equality constraints (2.25) and inequality constraints (2.34). The \( r_1 = m - 1 \) equality constraints in (2.25) can be written as

\[ c_i(\gamma) = \sum_{l=1}^{m} \alpha_i t_{il}^2 - 1 = 0, \quad i = 2, \ldots, m. \quad (2.38) \]
For each $c_i(\gamma)$, the nonzero elements in $\frac{\partial}{\partial \gamma} c_i(\gamma)$ are

$$\frac{\partial}{\partial t_{il}} c_i(\gamma) = 2t_{il} \alpha_l, \quad l = 1, \ldots, m,$$
$$\frac{\partial}{\partial \alpha_l} c_i(\gamma) = t^2_{il}, \quad l = 1, \ldots, m.$$

The $r_2 = q$ inequality constraints in (2.34) can be written as

$$c_{r_1+i}(\gamma) = 1 - \sum_{l=1}^{m} \alpha_l (\sum_{d=1}^{m} \lambda_{id} t_{dl})^2 \geq 0, \quad i = 1, \ldots, q. \quad (2.39)$$

The nonzero elements in $\frac{\partial}{\partial \gamma} c_{r_1+i}(\gamma)$ are

$$\frac{\partial}{\partial \lambda_{id}} c_{r_1+i}(\gamma) = -2 \sum_{l=1}^{m} \alpha_l t_{dl} \lambda_{il}^*, \quad d = 1, \ldots, m,$$
$$\frac{\partial}{\partial t_{dl}} c_{r_1+i}(\gamma) = -2 \alpha_l \lambda_{il}^* \lambda_{id}, \quad d = 1, \ldots, m, \quad l = 1, \ldots, d - 1,$$
$$\frac{\partial}{\partial \alpha_l} c_{r_1+i}(\gamma) = -(\sum_{d=1}^{m} \lambda_{id} t_{dl})^2, \quad l = 2, \ldots, m,$$

where $\lambda_{il}^* = [\Lambda^T]_d = \sum_{d=1}^{m} \lambda_{id} t_{dl}$. Combining these first derivatives of equality and inequality constraints with respect to $\gamma$, we can obtain the constraint function Jacobian matrix $L(\gamma)$.

When the latent factors are uncorrelated, the constraint vector $c(\gamma)$ only contains inequality constraints (2.35), which can be written as

$$c_i(\gamma) = 1 - \sum_{l=1}^{m} \lambda_{il}^2 \geq 0, \quad i = 1, \ldots, q. \quad (2.40)$$

For each $c_i(\gamma)$, the nonzero elements in $\frac{\partial}{\partial \gamma} c_i(\gamma)$ are

$$\frac{\partial}{\partial \lambda_{il}} c_i(\gamma) = -2 \lambda_{il}, \quad l = 1, \ldots, m.$$

The constraint function Jacobian matrix $L(\gamma)$ can then be calculated.
2.3.5 Constrained Optimization Procedure

The current research applies a constrained optimization procedure (c.f., Aitchison & Silvey, 1960; Browne & du Toit, 1984) to minimize the UBN fit function subject to equality and inequality constraints. The optimization problem can be framed as follows,

\[
\begin{align*}
\text{minimize} & \quad F_{ubn}(\gamma) \\
\text{subject to} & \quad c_i(\gamma) = 0, \quad i = 1, \ldots, r_1, \\
& \quad c_i(\gamma) \geq 0, \quad i = r_1, \ldots, r_1 + r_2,
\end{align*}
\]

where the equality and inequality constraints are summarized in subsection 2.3.3.

The constrained optimization method finds successively better solution to (2.41) by using the gradient function \(g(\gamma)\) in (2.9), expected Hessian matrix \(H\), in (2.33), constraint vector \(c(\gamma)\) in (2.36), and constraint function Jacobian matrix \(L(\gamma)\) in (2.37). Let \(\hat{\gamma}_t\) represent an approximation to the UBN estimate \(\hat{\gamma}_{ubn}\) obtained at the \(t\)th iteration. Let \(g_t, H_t, c_t, \) and \(L_t\) denote the gradient function, expected Hessian matrix, constraint vector, and constraint function Jacobian matrix evaluated at \(\hat{\gamma}_t\), respectively. Given the current estimate \(\hat{\gamma}_t\), a new approximation \(\hat{\gamma}_{t+1}\) is calculated by

\[
\begin{pmatrix}
\hat{\gamma}_{t+1} \\
l_{t+1}
\end{pmatrix} = 
\begin{pmatrix}
\hat{\gamma}_t \\
l_t
\end{pmatrix} + \alpha 
\begin{pmatrix}
H_t + L_t^t L_t & L_t^t \\
L_t & 0
\end{pmatrix}^{-1} 
\begin{pmatrix}
-g_t \\
-c_t
\end{pmatrix},
\]

where \(l_t\) is an \((r_1 + r_2) \times 1\) vector of Lagrange multipliers, step size \(\alpha=1, 1/2, 1/4, \ldots\) is employed to make sure that

\[
F_{ubn}(\hat{\gamma}_{t+1}) + l_{t+1}'c_{t+1} < F_{ubn}(\hat{\gamma}_t) + l_t'c_t.
\]
The iteration process converges when the gradient is smaller than a predetermined threshold, and the parameter values at the last iteration are treated as the minimizer of the UBN fit function. Refer to Browne and du Toit (1984, pp. 272-277) for more computational details.

2.4 Comparison between UBN and Other Approaches

In contrast to the full-information approach introduced by Lee et al. (1990), which requires calculating the full-information likelihood, the UBN method relies on one- and two-dimensional marginal distributions only and computes the model parameters by minimizing the UBN fit function. The omission of higher order integrations frees the estimation procedure from a heavy computational burden, especially when the number of ordinal variables $q$ is large. Jöreskog and Moustaki (2001) pointed out that the UBN approach has a weaker theoretical foundation as it maximizes the sum of all univariate and bivariate likelihoods and these likelihoods are not independent. As a consequence, the usual asymptotic result with regard to the full-information MLE is not applicable to UBN estimates. Instead, modifications are made on the usual theoretical result to obtain asymptotic results that are appropriate to the UBN estimates. Bearing in mind that full-information likelihood is usually difficult to evaluate for real world problems, limited-information approaches like UBN are more feasible and involve less technical difficulty.

Like UBN, the methodologies discussed in Muthén (1984), de Leon (2005) and Liu (2007) are limited-information procedures relying on one- and two-dimensional margins only. The difference is that UBN is a single-stage estimation procedure, while the others are multi-stage procedures. In particular, Muthén’s three-stage procedure has thresholds, polychoric correlations, and factor loadings involved in three separate
estimating stages, and parameters in the later stage depend on the estimates obtained from the earlier stage. The MPL method applied by de Leon (2005) and Liu (2007) defines a Pairwise Likelihood that is a function of thresholds and polychoric correlations, but factor loadings are still estimated by GLS as a separate second stage. When using the aforementioned multi-stage procedures, one needs to calculate the asymptotic covariance matrix of polychoric correlations since it is used as the weight matrix in GLS.

The UBN approach does not have a separate stage to estimate polychoric correlations. In UBN, the correlation coefficients between latent response variables are derived from the covariance structure of the modified factor analysis model and substituted in (2.3) as functions of model parameters. By minimizing the UBN fit function $F_{\text{ubn}}(\gamma)$, all the model parameters are estimated in one single stage. Besides, the current research employs a constrained Fisher scoring type of algorithm to calculate the UBN estimates. No GLS is involved and the asymptotic covariance matrix of polychoric correlations is not required in this estimation procedure.

One similarity between the MPL approach described in de Leon (2005) and the UBN approach investigated by the current research is that both approaches can be regarded as special cases of the composite likelihood (c.f., Lindsay, 1988). In contrast to the true likelihood, the composite likelihood defines a new likelihood function based on events of interest and usually is much easier to calculate. The next chapter introduces the composite likelihood and explains the relationship between UBN and the composite likelihood.
Chapter 3: Composite Likelihood

3.1 Definition of Composite Likelihood

Composite likelihood, first proposed and called pseudo-likelihood by Besag (1975), is a likelihood formed by adding together individual component likelihoods, each of which corresponds to a marginal or conditional event (Lindsay, 1988). It is employed to overcome the difficulty in evaluating true likelihood function due to complex dependencies (Varin, 2008).

Under the modified factor analysis model, the true likelihood function of a response vector \( \mathbf{x} \) is

\[
 f(\mathbf{x}; \gamma) = Pr(x_1 = a_1, \ldots, x_q = a_q) \\
 = \int_{x_{a_1}^{(1)}}^{x_{a_1}^{(+1)}} \cdots \int_{x_{a_q}^{(q)}}^{x_{a_q}^{(+q)}} \phi_q(x_1^*, \ldots, x_q^* | P) dx_1^* \cdots dx_q^*,
\]

(3.1)

where \( a_i \in \{1, \ldots, k_i\}, \ i = 1, \ldots, q \). The integration is over the standardized \( q \)-dimensional normal density function \( \phi_q(\cdot, \ldots, \cdot | P) \) with a null mean vector and a covariance matrix \( P \). This is the true likelihood function Lee et al. (1990) compute in their full-information ML approach.

Usually there is no closed form of that integral in (3.1), and it is extremely difficult to evaluate when the dimensionality \( q \) is large. One way to alleviate the computational burden is to consider computing likelihoods of certain subsets of the data. Let \( A_1, \ldots, A_D \) denote a set of marginal or conditional events with associated
likelihoods \( f(x \in A_d; \gamma) = Pr(x \in A_d), \ d = 1, \ldots, D \). A composite likelihood is the weighted product of the individual likelihoods,

\[
C(x; \gamma) = \prod_{d=1}^{D} f(x \in A_d; \gamma)^{\omega_d},
\]

where \( \omega_d \) is positive weight associated with event \( A_d \). The corresponding composite log likelihood is

\[
c(x; \gamma) = \log C(x; \gamma) = \sum_{d=1}^{D} \omega_d \log f(x \in A_d; \gamma),
\]

and its maximizer \( \hat{\gamma}_c \), if unique, is called Maximum Composite Likelihood Estimator (MCLE, c.f., Varin, 2008).

There are two main classes of composite likelihood, namely, composite conditional likelihood and composite marginal likelihood. The composite conditional likelihood removes some terms that make full likelihood evaluation complicated by computing conditional likelihood of certain events (Varin, 2008). The other class, the composite marginal likelihood is constructed from low-dimensional margins and therefore avoids calculating the full likelihood. The purpose of constructing a composite likelihood is to have substantial computational saving and at the same time keep the loss of efficiency tolerable. Since it will be seen that the UBN approach is a special case of composite marginal likelihood, we now focus on discussing this class of composite likelihood.

The simplest composite marginal likelihood is the one constructed under working independence assumptions,

\[
C_1(x; \gamma) = \prod_{i=1}^{q} f(x_i; \gamma)^{\omega_i},
\]

where \( f(x_i; \gamma) \) denotes the probability of observing each individual variable \( x_i \). This composite marginal likelihood relies on the univariate marginal events only and is
termed Independence Likelihood (Chandler & Bate, 2007). $C_1(x; \gamma)$ is equal to the true likelihood if all variables are independent from each other. When this independence assumption is not true, the Independent Likelihood is no longer equal to the true likelihood because it does not take the dependency among variables into account. One can make inferences on marginal parameters, like thresholds, from $C_1(x; \gamma)$. If parameters related to dependence between variables are also of interest, it is necessary to model pairs of observations, for example, the Pairwise Likelihood (Cox & Reid, 2004; Varin, 2008; Liu, 2007),

$$C_2(x; \gamma) = \prod_{i=2}^{q} \prod_{j=1}^{i-1} f(x_i, x_j; \gamma)^{\omega_{i,j}},$$

(3.5)

or composite likelihood constructed from larger subsets such as triplets of observations (Varin & Vidoni, 2005; Engler, Mohapatra, Louis, & Betensky, 2006). The Pairwise Likelihood $C_2(x; \gamma)$ calculates the probability of every possible variable pair, and it has been applied to analyze ordinal data by several authors, e.g., de Leon (2005) and Liu (2007).

Under the modified factor analysis model, $f(x_i; \gamma)$ in the Independence Likelihood is equal to $Pr(x_i = a_i)$ involved in (2.2). It only requires one-dimensional integration over each latent response variable and it is a function of model parameters. Similarly, $f(x_i, x_j; \gamma)$ in the Pairwise Likelihood is equal to $Pr(x_i = a_i, x_j = a_j)$ involved in (2.3). Only two-dimensional integration is needed to calculate the Pairwise Likelihood and it is also a function of model parameters. By calculating marginal likelihood on these low-dimensional events, the composite marginal likelihood avoids computing the probability of a full response vector $x$. 
3.2 Asymptotic Properties of MCLE

The maximizer of the composite likelihood function, $\hat{\gamma}_c$, solves the composite score equation,

$$s(x; \gamma) = \frac{\partial}{\partial \gamma} c(x; \gamma) = \sum_{d=1}^{D} \omega_d s_d(x; \gamma) = 0,$$

where the composite log likelihood $c(x; \gamma)$ is defined in (3.3),

$$s_d(x; \gamma) = \frac{\partial}{\partial \gamma} \log f(x \in A_d; \gamma)$$

is called the component score (Lindsay, 1988). The composite score function $s(x; \gamma)$ is similar to the score function used in full-information ML problem, except that $s(x; \gamma)$ is specifically defined for the composite likelihood. Because the composite log likelihood $c(x, \gamma)$ consists of conventional log likelihood functions, it carries some features of the ordinary likelihood. In particular, Lindsay (1988) shows that, under regularity conditions,

$$E_{\gamma}[s_d(X; \gamma)] = 0, \quad d = 1, \ldots, D,$$

where $s_d(X; \gamma)$ is defined in (3.7) and $D$ denotes the number of marginal events. Equation (3.8) implies that

$$E_{\gamma}[s(X; \gamma)] = 0.$$

Therefore the composite score function $s(x; \gamma)$ satisfies the requirement of being an unbiased estimating function.

Because composite likelihood is not a full likelihood, the usual Fisher information is not applicable for obtaining standard error estimates. It may be replaced by
the Godambe information (Godambe, 1960) or sandwich information,

\[ G(\gamma) = H(\gamma)J^{-1}(\gamma)H(\gamma), \]  

(3.10)

where \( G(\gamma) \) is the Godambe information matrix in a single observation, \( H(\gamma) \) is the negative expected Hessian matrix of composite log likelihood,

\[ H(\gamma) = -\mathcal{E}_\gamma \left[ \frac{\partial}{\partial \gamma} s(X; \gamma) \right] = -\mathcal{E}_\gamma \left[ \frac{\partial^2}{\partial \gamma \partial \gamma'} c(X; \gamma) \right], \]  

(3.11)

and \( J(\gamma) \) is the expected outer product of composite score,

\[ J(\gamma) = \mathcal{E}_\gamma [s(X; \gamma)s(X; \gamma)'] = \mathcal{E}_\gamma \left[ \left( \frac{\partial}{\partial \gamma} c(X; \gamma) \right) \left( \frac{\partial}{\partial \gamma} c(X; \gamma) \right)' \right], \]  

(3.12)

where \( c(X; \gamma) \) is the composite log likelihood function defined in (3.3) and \( s(X; \gamma) \) is the composite score function defined in (3.6). For a true likelihood, matrices \( H(\gamma) \) and \( J(\gamma) \) are equal to each other and the Godambe information matrix \( G(\gamma) \) becomes the usual Fisher information matrix. But for composite likelihood, \( H(\gamma) \neq J(\gamma) \), and it indicates loss of efficiency compared to MLE (Varin, 2008).

Consider a random sample \( \{x_1, \ldots, x_N\} \), which are \( N \) independently and identically distributed ordinal observations from a modified factor analysis model. Each observation contains a person’s responses to \( q \) ordinal variables. Let \( C(\gamma) \) and \( c(\gamma) \) denote the composite likelihood and corresponding composite log likelihood of all the data, respectively,

\[ C(\gamma) = \prod_{n=1}^{N} C(x_n; \gamma), \]  

\[ c(\gamma) = \log C(\gamma) = \sum_{n=1}^{N} c(x_n; \gamma), \]  

(3.13)

(3.14)

refer to (3.2) and (3.3) for definitions of \( C(x_n; \gamma) \) and \( c(x_n; \gamma) \). Let \( \hat{\gamma}_c \) denote the maximizer of \( c(\gamma) \). As \( N \) goes to infinity, under certain regularity conditions, it can
be proven (Liu, 2007) that \( \hat{\gamma}_c \) is consistent and asymptotically normal,

\[
\sqrt{N}(\hat{\gamma}_c - \gamma_0) \xrightarrow{d} N(0, G^{-1}(\gamma_0)),
\]

where \( \gamma_0 \) is the true parameter value when the model holds, \( G(\gamma_0) \) is the Godambe information matrix evaluated at \( \gamma_0 \). Refer to Liu (2007, p. 22) for a detailed proof.

### 3.3 UBN as a Composite Likelihood Approach

A closer look of the UBN fit function shows that it combines both Independence Likelihood \( C_1(x; \gamma) \) and Pairwise Likelihood \( C_2(x; \gamma) \),

\[
F_{ubn}(\gamma) = \sum_{i=1}^{q} \sum_{a=1}^{k_i} p_a^{(i)} \ln(p_a^{(i)}) + \sum_{i=2}^{q} \sum_{j=1}^{k_i} \sum_{a=1}^{k_j} \sum_{b=1}^{k_j} p_{ab}^{(ij)} \ln(p_{ab}^{(ij)})
- \sum_{i=1}^{q} \sum_{a=1}^{k_i} p_a^{(i)} \ln(\pi_a^{(i)}(\gamma)) + \sum_{i=2}^{q} \sum_{j=1}^{k_i} \sum_{a=1}^{k_j} \sum_{b=1}^{k_j} p_{ab}^{(ij)} \ln(\pi_{ab}^{(ij)}(\gamma))
\]

\[
= \eta - \frac{1}{N} \left( \sum_{n=1}^{N} \sum_{i=1}^{q} \ln(f(x_{ni}; \gamma)) + \sum_{n=1}^{N} \sum_{i=2}^{q} \sum_{j=1}^{i-1} \ln(f(x_{ni}, x_{nj}; \gamma)) \right)
\]

\[
= \eta - \frac{1}{N} \left( \sum_{n=1}^{N} \ln(\prod_{i=1}^{q} f(x_{ni}; \gamma)) + \sum_{n=1}^{N} \ln(\prod_{i=2}^{q} \prod_{j=1}^{i-1} f(x_{ni}, x_{nj}; \gamma)) \right)
\]

\[
= \eta - \frac{1}{N} \left( \sum_{n=1}^{N} \ln(C_1(x_n; \gamma)) + \sum_{n=1}^{N} \ln(C_2(x_n; \gamma)) \right)
\]

\[
= \eta - \frac{1}{N} \sum_{n=1}^{N} \ln(C_1(x_n; \gamma) C_2(x_n; \gamma)),
\]

where

\[
\eta = \sum_{i=1}^{q} \sum_{a=1}^{k_i} p_a^{(i)} \ln(p_a^{(i)}) + \sum_{i=2}^{q} \sum_{j=1}^{k_i} \sum_{a=1}^{k_j} \sum_{b=1}^{k_j} p_{ab}^{(ij)} \ln(p_{ab}^{(ij)})
\]

is a constant given the data, \( x_n \) denotes the vector-valued responses given by subject \( n \), and the weights \( \omega_i \) and \( \omega_{i,j} \) are all equal to 1.
Define a composite likelihood $C_{ubn}(x, \gamma)$, which is the product of Independence Likelihood and Pairwise Likelihood,

$$C_{ubn}(x, \gamma) = C_1(x, \gamma)C_2(x, \gamma),$$

with the corresponding composite log likelihood $c_{ubn}(x, \gamma) = \log C_{ubn}(x, \gamma)$. Equation (3.16) shows that the UBN fit function $F_{ubn}(\gamma)$ can be regarded as a special case of the composite likelihood, because

$$F_{ubn}(\gamma) = \eta - \frac{1}{N} \sum_{n=1}^{N} c_{ubn}(x_n, \gamma),$$

and minimizing $F_{ubn}(\gamma)$ is equivalent to maximizing the composite log likelihood of all the data, i.e., $\sum_{n=1}^{N} c_{ubn}(x_n, \gamma)$. Therefore, the asymptotic properties reviewed in Section 3.2 are also applicable to the UBN estimate, meaning that the UBN estimate $\hat{\gamma}_{ubn}$ is consistent and asymptotically normal, and the Godambe information matrix gives estimate of the asymptotic covariance matrix of $\hat{\gamma}_{ubn}$. 

31
Chapter 4: Statistical Inference on the UBN Estimates

4.1 Standard Error Estimates

In their 2001 paper, Jöreskog and Moustaki (2001) focus on introducing the UBN fit function to factor analyze ordinal data and comparing UBN with three popular approaches. However, they do not provide information on estimating standard errors of the UBN estimates or goodness-of-fit tests. The current section discusses a general method for obtaining standard errors when using limited-information approaches and gives detail on applying this method to the UBN approach. Section 4.2 will discuss a residual based goodness-of-fit test.

4.1.1 Estimating Godambe Information Matrix

It is shown in Section 3.2 that under certain regularity conditions the Maximum Composite Likelihood Estimate (MCLE) $\hat{\gamma}_c$ is consistent and asymptotically normal. The asymptotic covariance matrix of $\hat{\gamma}_c$ is given by $\frac{1}{N} G^{-1}(\gamma_0)$, with the Godambe information matrix evaluated at the true value. Hence, standard error estimates of $\hat{\gamma}_c$ can be obtained by taking square roots of corresponding diagonal elements in $\frac{1}{N} G^{-1}(\gamma_0)$.

As a special case of the composite likelihood, the UBN approach possesses this property that the UBN estimate $\hat{\gamma}_{ubn}$ is consistent and asymptotically normal,

$$\sqrt{N} (\hat{\gamma}_{ubn} - \gamma_0) \xrightarrow{d} N(0, G^{-1}(\gamma_0)),$$
where $G(\gamma) = H(\gamma)J^{-1}(\gamma)H(\gamma)$. Because the component matrices $H(\gamma)$ and $J(\gamma)$ involve calculating expected values over all possible responses, they are usually unknown. To derive standard error estimates from $G(\gamma)$, one needs consistent estimates of its component matrices $H(\gamma)$ and $J(\gamma)$. Besides, the true value $\gamma_0$ is also unknown in practical problems and needs to be estimated.

We first focus on estimating component matrices $H(\gamma)$ and $J(\gamma)$. Equation (3.18) implies that

$$
\mathcal{E}_\gamma \left[ \frac{\partial^2}{\partial \gamma \partial \gamma'} F_{ubn}(\gamma) \right] = -\mathcal{E}_\gamma \left[ \frac{\partial^2}{\partial \gamma \partial \gamma'} c_{ubn}(X, \gamma) \right].
$$

By definition, the left hand side of (4.1) is $H_\gamma$ and a computational form of $H_\gamma$ is given in (2.33), while the right hand side of (4.1) is the unknown matrix $H(\gamma)$, see (3.11). Therefore, $H(\gamma)$ is directly obtainable from the minimization procedure, by using the computational form (2.33).

Matrix $J(\gamma)$ is the expected outer product of the composite score $\frac{\partial}{\partial \gamma} c_{ubn}(x_n, \gamma)$, refer to (3.12) for its definition. Although it is difficult to derive an analytical form of $J(\gamma)$, by applying the Law of Large Numbers, an empirical estimate is available. Given that the observed responses are independent replicates from the same model, $J(\gamma)$ can be approximated by its sample mean over available responses $x_1, \ldots, x_n$,

$$
\hat{J}(\gamma) = \frac{1}{N} \sum_{n=1}^{N} \left( \frac{\partial}{\partial \gamma} c_{ubn}(x_n, \gamma) \right) \left( \frac{\partial}{\partial \gamma} c_{ubn}(x_n, \gamma) \right)'.
$$

The gradient vector $\frac{\partial}{\partial \gamma} F_{ubn}(\gamma)$ is already discussed in subsection 2.3.1 and can be substituted in (4.2) to calculate $\hat{J}(\gamma)$. A consistent estimate of $G(\gamma)$ is therefore

$$
\hat{G}(\gamma) = H_\gamma \hat{J}^{-1}(\gamma) H_\gamma,
$$
where \( H \) is given in (2.33) and \( \dot{J}(\gamma) \) is computed by (4.2).

The UBN minimizer \( \hat{\gamma}_{ubn} \) is a consistent estimator of the true value \( \gamma_0 \). When \( N \) is large, it is expected that \( \hat{\gamma}_{ubn} \) will be close to \( \gamma_0 \). So we substitute \( \hat{\gamma}_{ubn} \) in \( \hat{G}(\gamma) \) to approximate \( \hat{G}(\gamma_0) \). The standard errors of the parameter estimates are then approximated by square roots of corresponding diagonal elements in \( \frac{1}{N} \hat{G}^{-1}(\hat{\gamma}_{ubn}) \).

### 4.1.2 Calculating Standard Errors

In Chapter 2, detailed information of the estimation procedure is given, under two situations, i.e., correlated factors and uncorrelated factors. When the latent factors are uncorrelated, model parameter vector \( \gamma \) only contains thresholds \( \tau \) and factor loadings \( \lambda \), and their standard errors can be obtained by taking square roots of the diagonal elements in \( \frac{1}{N} \hat{G}^{-1}(\hat{\gamma}_{ubn}) \), where \( \hat{G}^{-1}(\hat{\gamma}_{ubn}) \) is a consistent estimate of the Godambe information matrix, as is described in subsection 4.1.1.

When the latent factors are correlated, we apply the parameterization (2.23) to make sure that the factor correlation matrix \( \Phi \) is non-negative definite. Elements in lower triangular matrix \( T \) and diagonal matrix \( D_\alpha \) are not independent, and they satisfy the equality constraints (2.25). This parameter dependence does not cause problem to the optimization process, because in the constrained optimization process we use \( H_t + L'_t L_t \) instead of \( H_t \) in (2.42). Matrix \( H_t + L'_t L_t \) is positive definite even when \( H_t \) is singular, and this change has no effect on the final solution (Browne & du Toit, 1992). However, the parameter dependence makes the information matrix singular. To obtain standard errors of model parameters under this situation, we need to remove redundant parameters and write the information matrix as functions of independent parameters only.
From equality constraints (2.25), diagonal elements in $D_\alpha$ can be written as a set of recursive functions,

$$\alpha_2 = 1 - t_{21}^2,$$

$$\alpha_3 = 1 - t_{31}^2 - t_{32}^2 \alpha_2,$$

$$\ldots$$

$$\alpha_i = 1 - t_{i1}^2 - t_{i2}^2 \alpha_2 - \cdots - t_{i,i-1}^2 \alpha_{i-1},$$

$$\ldots$$

$$\alpha_m = 1 - t_{m1}^2 - t_{m2}^2 \alpha_2 - \cdots - t_{m,m-1}^2 \alpha_{m-1}.$$

In (4.4), parameter $\alpha_2$ is expressed as a function of $t$, the vector containing lower triangular elements in $T$. By substituting this function in $\alpha_3$, we can write $\alpha_3$ as a function of $t$. Similarly, by substituting $\alpha_2$ to $\alpha_{i-1}$ in $\alpha_i$ we can write $\alpha_i$ as a function of $t$.

Since $\alpha_2$ to $\alpha_m$ can be written as functions of lower triangular elements in $T$, they are redundant parameters. We replace them by corresponding functions of $t$. Let $\tilde{\gamma}$ denote a parameter vector including only the independent parameters, the original parameter vector $\gamma$ can be written as

$$\gamma = (\tilde{\gamma}', \alpha'),$$

where vector $\alpha$ includes diagonal elements $\alpha_2$ to $\alpha_m$ in $D_\alpha$ and

$$\tilde{\gamma} = (\tau', \lambda', t')',$$

vector $\tau$ contains thresholds, vector $\lambda$ contains factor loadings, and vector $t$ contains lower triangular elements in $T$. 
Let $F_{ubn} (\tilde{\gamma})$ denote the UBN fit function expressed in independent parameter vector $\tilde{\gamma}$ and $F_{ubn} (\tilde{\gamma}, \alpha)$ denote the UBN fit function expressed in all model parameters, i.e., $F_{ubn} (\tilde{\gamma}, \alpha) = F_{ubn} (\gamma)$. Note that $F_{ubn} (\tilde{\gamma})$ and $F_{ubn} (\tilde{\gamma}, \alpha)$ are the same function under two parameterization. To obtain the Godambe information matrix of $\tilde{\gamma}$ requires gradient function and Hessian matrix of $F_{ubn} (\tilde{\gamma})$ with respect to $\tilde{\gamma}$. By applying the Chain rule, we have
\[
\frac{\partial}{\partial \tilde{\gamma}} F_{ubn} (\tilde{\gamma}) = \frac{\partial}{\partial \tilde{\gamma}} F_{ubn} (\tilde{\gamma}, \alpha) + \frac{\partial \alpha'}{\partial \tilde{\gamma}} \frac{\partial}{\partial \alpha} F_{ubn} (\tilde{\gamma}, \alpha),
\] (4.5)
where $\frac{\partial \alpha'}{\partial \tilde{\gamma}}$ represents the first derivatives of $\alpha$ with respect to parameter $\tilde{\gamma}$ and it can be calculated from (4.4). By substituting (4.5) in (4.2) we can get component matrix $\hat{J} (\tilde{\gamma})$. Similarly,
\[
\frac{\partial}{\partial \tilde{\gamma}} \pi_a^{(i)} (\tilde{\gamma}) = \frac{\partial}{\partial \tilde{\gamma}} \pi_a^{(i)} (\tilde{\gamma}, \alpha) + \frac{\partial \alpha'}{\partial \tilde{\gamma}} \frac{\partial}{\partial \alpha} \pi_a^{(i)} (\tilde{\gamma}, \alpha),
\] (4.6)
and by substituting them in (2.33) we can obtain the other component matrix $H (\tilde{\gamma})$.

The Godambe information matrix of parameter $\tilde{\gamma}$
\[
\hat{G} (\tilde{\gamma}) = H (\tilde{\gamma}) \hat{J}^{-1} (\tilde{\gamma}) H (\tilde{\gamma})
\] is now invertible.

Given $\hat{G} (\tilde{\gamma})$, standard errors of thresholds and factor loadings are obtained by taking square roots of the corresponding diagonal elements in $\frac{1}{N} \hat{G}^{-1} (\tilde{\gamma})$. Determining the standard errors of correlation coefficients in the factor correlation matrix $\Phi$ requires additional computation. The parameterization (2.23) implies that
\[
\phi_{ij} = [\Phi]_{ij} = \sum_{l=1}^{m} \alpha_l t_{il} t_{jl}, \quad i \neq j,
\] (4.8)
where $t_{il} = [T]_{il}$, $\alpha_l = [D_{\alpha}]_{il}$, and they satisfy the equality and inequality constraints (2.25) and (2.24). From (4.4), we can write $\alpha_2$ to $\alpha_m$ as functions of the lower triangular elements in $T$. By substituting these functions in (4.8), each $\phi_{ij}$ can be written as a function of $t$, i.e., $\phi_{ij}(t)$, where vector $t$ contains the lower triangular elements in $T$. Let $\hat{\Sigma}_t$ denote the asymptotic covariance matrix of $t$ obtained from $\frac{1}{N}\hat{G}^{-1}(\gamma)$. Using the Delta method (Casella & Berger, 2002), standard error estimate of $\phi_{ij}$ can be calculated as follows,

$$
\hat{\sigma}_{ij} = \sqrt{\left(\frac{\partial \phi_{ij}}{\partial t}\right) \hat{\Sigma}_t \left(\frac{\partial \phi_{ij}}{\partial t}\right)},
$$

(4.9)

where $\frac{\partial \phi_{ij}}{\partial t}$ is gradient function of $\phi_{ij}$ with respect to $t$.

In sum, when the latent factors are uncorrelated, standard errors of thresholds and factor loadings are directly obtained from the Godambe information matrix $G(\gamma)$. When the latent factors are correlated, the parameterization (2.23) indicates parameter dependency and the involved information matrix is singular. To obtain standard errors, parameters in $\alpha$ are expressed as functions of $t$. An invertible Godambe information matrix of the independent parameters $\tilde{\gamma}$ is then obtained. Standard errors of thresholds and factor loadings are calculated by using this invertible Godambe information matrix. Calculating standard errors of factor correlation coefficients requires an extra step that uses the Delta method.

In Chapter 5, a series of simulation studies is conducted to evaluate the standard errors estimated by using the approximate Godambe information, under these two situations, i.e., uncorrelated factors and correlated factors.
4.2 Residual Based Goodness-of-Fit Test Statistic

The composite likelihood approaches bring in significant computational savings because they usually rely on low dimensional integrations. But the computational convenience comes with a price — we cannot apply the usual theoretical result developed for full-information MLE procedures. This leads to the development of theories that are appropriate to limited-information estimators. An example is the Godambe information matrix introduced in Section 4.1, which is employed to replace the Fisher information matrix for true MLE. Another challenge confronting researchers is how to assess the overall goodness of fit with limited information.

There are two widely used goodness-of-fit test statistics for assessing the overall model fit in multivariate ordinal data analysis. They are Pearson’s $X^2$,

$$X^2 = N \sum_r \frac{(p_r - \hat{\pi}_r)^2}{\hat{\pi}_r},$$  \hspace{1cm} (4.10)

and the likelihood ratio statistic,

$$G^2 = 2N \sum_r p_r \ln \left( \frac{p_r}{\hat{\pi}_r} \right),$$  \hspace{1cm} (4.11)

where index $r$ goes through all possible response patterns, $p_r$ is the sample proportion of the response pattern $x_r$, and $\hat{\pi}_r$ is a consistent estimate of the corresponding expected probability. When the model holds, the two statistics are asymptotically equivalent (Maydeu-Olivares & Joe, 2006), and both are based on full information.

For $q$ ordinal variables, the total number of possible patterns is $\prod_{i=1}^q k_i$, where $k_i$ is the number of response categories for the ordinal variable $x_i$. A contingency table is said to be sparse if the ratio $N/\prod_{i=1}^q k_i$ is small (Agresti & Yang, 1987). When the number of possible patterns is large or the sample size is small, the contingency table becomes sparse and the empirical Type I error rates of $X^2$ and $G^2$ do not match...
their expected rates under their asymptotic null distribution (Maydeu-Olivares & Joe, 2005). To overcome this problem, a goodness-of-fit test statistic that relies on marginal tables will be introduced in this section. This test statistic assumes a quadratic form of the marginal residuals. Its asymptotic distribution can be derived from the full information residual vector.

4.2.1 Full-Information Residuals

First consider the \( q \)-dimensional response vector \( x \) and the full likelihood \( f(x; \gamma) \) specified in (3.1). The total number of all possible response patterns is

\[
s(q) = \prod_{i=1}^{q} k_i.
\]  

(4.12)

Let \( p \) denote the \( s(q) \)-dimensional vector including sample proportions of all possible response patterns \( x_r, r = 1, \ldots, s(q) \), and \( \pi(\gamma) \) denote the \( s(q) \)-dimensional vector of corresponding model fitted probabilities. Elements in \( \pi(\gamma) \) can be calculated by using the \( q \)-dimensional integration in (3.1). It is proven that

\[
\sqrt{N}(p - \pi(\gamma)) \xrightarrow{d} N(0, \Gamma(\gamma)),
\]  

(4.13)

where \( \Gamma(\gamma) = D - \pi(\gamma)\pi(\gamma)' \) and \( D \) is a diagonal matrix with \( \pi(\gamma) \) being its diagonal elements, i.e., \( D = \text{Diag}(\pi(\gamma)) \) (Agresti, 1990).

Define a full-information residual vector,

\[
\hat{e} = p - \pi(\hat{\gamma}),
\]  

(4.14)

where \( \hat{\gamma} \) is a consistent estimator of model parameters \( \gamma \). Inferring from (4.13), Maydeu-Olivares and Joe (2006, p. 717) show that the residual vector \( \hat{e} \) is asymptotically normal with a null mean vector and an asymptotic covariance matrix \( \hat{\Sigma}_e \), i.e., \( \sqrt{N}\hat{e} \xrightarrow{d} N(0, \hat{\Sigma}_e) \). Therefore, a quadratic form test statistic constructed from
the residual vector, i.e., $N\hat{e}'\hat{\Sigma}_e^{-1}\hat{e}$, is asymptotically Chi-square distributed and can be used to assess model fit.

However, the asymptotic covariance matrix $\hat{\Sigma}_e$ involves computing the full likelihood that requires $q$-dimensional integrations. This is practically infeasible, especially when $q$ is large. The UBN approach avoids computing full-information probabilities by relying on low dimensional marginal probabilities and proportions. Similarly, we will construct a residual vector from the available univariate and bivariate margins and prove that this marginal residual vector is asymptotically normal. Hence, a quadratic form test statistic depending on low dimensional integrations can be derived from this marginal residual vector.

### 4.2.2 Marginal Residual Based Test Statistic $M_2$

Maydeu-Olivares and Joe (2005) provided a family of goodness-of-fit test statistics for structured contingency table proportions to be used in conjunction with consistent but non-efficient parameter estimates. They accomplished this by adapting a test statistic for the analysis of covariance structures, due to Browne (1984, Proposition 4), by substituting contingency table proportions for covariances and respecifying an associated asymptotic covariance matrix. Maydeu-Olivares and Joe (2005) named test statistics in their family $M_r$, where the subscript $r$ stands for the maximum order of marginal contingency tables retained. Some $M_2$ statistics will be developed here as goodness-of-fit tests based on UBN estimates.

Let $\hat{\pi}_1(\gamma)$ denote a vector of all univariate marginal probabilities except the ones involving the first category of any variable, $\hat{\pi}_2(\gamma)$ denote a vector of all bivariate marginal probabilities except the ones involving the first category of any variable, $\hat{p}_1$ and $\hat{p}_2$ denote the corresponding sample proportions, respectively. The four vectors
\[ \tilde{\pi}_1(\gamma) = (\pi_2^{(1)}(\gamma), \pi_3^{(1)}(\gamma), \ldots, \pi_{k_1}^{(1)}(\gamma), \pi_2^{(2)}(\gamma), \ldots, \pi_{k_q}^{(q)}(\gamma))', \]

\[ \tilde{p}_1 = (p_2^{(1)}, p_3^{(1)}, \ldots, p_{k_1}^{(1)}, p_2^{(2)}, \ldots, p_{k_q}^{(q)})', \]

\[ \tilde{\pi}_2(\gamma) = (\pi_{22}^{(12)}(\gamma), \pi_{23}^{(12)}(\gamma), \ldots, \pi_{2k_2}^{(12)}(\gamma), \pi_{32}^{(12)}(\gamma), \ldots, \pi_{k_q-1k_q}^{(q-1q)}(\gamma))', \]

\[ \tilde{p}_2 = (p_{22}^{(12)}, p_{23}^{(12)}, \ldots, p_{2k_2}^{(12)}, p_{32}^{(12)}, \ldots, p_{k_q-1k_q}^{(q-1q)})', \]

where the univariate and bivariate marginal probabilities involved are calculated by (2.2) and (2.3), the superscript \((i)\) indicates the variable index, and the subscript indicates the response category. Notice that no subscript is equal to 1, the lowest category. This is because given an ordinal variable, or a pair of ordinal variables, the sum of all univariate/bivariate marginal proportions is equal to 1, meaning that these proportions are linearly dependent. By excluding the first category of any variable, this dependency problem no longer exists.

Define \( \pi_2(\gamma) = (\tilde{\pi}_1(\gamma)', \tilde{\pi}_2(\gamma)')' \) and \( \mathbf{p}_2 = (\tilde{\mathbf{p}}_1', \tilde{\mathbf{p}}_2')' \). The dimensionality of \( \pi_2(\gamma) \) and \( \mathbf{p}_2 \) is

\[ s(2) = \sum_{i=1}^{q} (k_i - 1) + \sum_{i=2}^{q} \sum_{j=1}^{i-1} (k_i - 1)(k_j - 1). \quad (4.15) \]

Maydeu-Olivares and Joe (2006, p. 717) show that there is a transformation matrix with full row rank that maps the full-information probability vector to the marginal probability vector,

\[ \pi_2(\gamma) = T\pi(\gamma), \quad (4.16) \]

where \( T \) is an \( s(2) \times s(q) \) transformation matrix. The transformation matrix \( T \) also maps the full-information proportion vector to the marginal proportion vector,

\[ \mathbf{p}_2 = T\mathbf{p}. \quad (4.17) \]
Result (4.13) implies that
\[ \sqrt{N}(p_2 - \pi_2(\gamma_0)) = \sqrt{N}T(p - \pi(\gamma_0)) \xrightarrow{d} N(0, \Upsilon), \] (4.18)
where \( \Upsilon = TT'(\gamma_0)T' \).

Consider the marginal residual vector,
\[ \hat{e}_2 = p_2 - \pi_2(\hat{\gamma}) \] (4.19)
where \( \hat{\gamma} \) is a consistent estimator of \( \gamma \). Let
\[ \Delta(\gamma) = \frac{\partial}{\partial \gamma} \pi_2(\gamma) \] (4.20)
be the Jacobian matrix and it is of full column rank. Consider an orthogonal complement matrix to \( \Delta(\gamma) \), i.e., \( \Delta_c(\gamma) \), such that \( \Delta_c(\gamma)'\Delta(\gamma)=0 \). Let \( \hat{\Upsilon} \) be a consistent estimator of \( \Upsilon \), \( \hat{\Delta} = \Delta(\hat{\gamma}) \) and \( \hat{\Delta}_c = \Delta_c(\hat{\gamma}) \). Suppose the number of independent model parameters is \( w \). Under the null hypothesis that the model fits, the asymptotic distribution of the quadratic form test statistic
\[ M_2 = N\hat{e}'_2\hat{\Delta}_c\{\hat{\Delta}_c'\hat{\Upsilon}\hat{\Delta}_c\}^{-1}\hat{\Delta}_c'\hat{\Delta}_c\hat{e}_2 \] (4.21)
as \( N \) tends to infinity is central Chi-square with \( s(2) - w \) degrees of freedom.

The model fitted probabilities and sample proportions required in calculating \( M_2 \) are already available when computing the UBN fit function. Computation of \( \hat{\Delta}_c \) in (4.21) can be avoided if an alternative computational form is used, c.f., equation (2.20b) in Browne (1984). Maydeu-Olivares and Joe\(^1\) (2006, equation (3.1) and Appendix) provide a brief description on obtaining the covariance matrix \( \hat{\Upsilon} \), which involves integrating over three- and four-variate standardized normal distribution.

\(^1\)The author is indebted to A. Maydeu-Olivares and H. Joe (personal communication, October 5, 2010) for providing the C code that is used to calculate the alternative computational form in their program. She converted the C code into FORTRAN and implemented it with other computation tasks to calculate the \( M_2 \) statistic for the UBN approach.
When the specified model is a good representation of the ordinal data, the test statistic $M_2$ will follow a central Chi-square distribution with $s(2) - w$ degrees of freedom. Chapter 5 will show random sampling experiments conducted to demonstrate the asymptotic properties of the test statistic $M_2$ in the UBN approach.

### 4.2.3 A Further Reduction of $M_2$

When calculating $M_2$, the size of asymptotic covariance matrix $\Upsilon = TT(\gamma)T'$ depends on the number of ordinal variables $q$ and the number of response categories per variable $k_i$, see (4.15). Maydeu-Olivares and Joe (2010) comment that there is a computational limit in the size of the models for multidimensional tables that can be tested due to the need to store very large matrices. For example, for a problem with 35 ordinal variables and 7 categories per variable, the number of non-duplicated elements in $\Upsilon$ that need to be stored is about 234 million. This may impose a problem on a PC because it requires too much memory space. To overcome this problem, Maydeu-Olivares and Joe (2010) suggest collapsing of the residual vector and use of a further reduction of $M_2$, which requires same amount of computation but needs less memory space to store the asymptotic covariance matrix $\Upsilon$.

L. Cai (personal communication, January 11, 2011) suggests a simple reduction on $M_2$ that uses the ordered scores. For each ordinal variable, define a weighted sum of the category probabilities to represent this item,

$$
\pi^{(i)}(\gamma) = \sum_{a=2}^{k_i} a \pi^{(i)}_{a}(\gamma).
$$

Notice that $\pi^{(i)}(\gamma)$ may be larger than one, and it is not a probability. For each pair of ordinal variables, perform a similar computation and obtain

$$
\pi^{(ij)}(\gamma) = \sum_{a=2}^{k_i} \sum_{b=2}^{k_j} a b \pi^{(ij)}_{ab}(\gamma).
$$
Define $\tilde{\pi}_2(\gamma) = (\pi^{(1)}(\gamma), \ldots, \pi^{(q)}(\gamma), \pi^{(12)}(\gamma), \ldots, \pi^{(q-1q)}(\gamma))'$. The number of elements in $\tilde{\pi}_2(\gamma)$ is equal to the number of ordinal variables plus the number of all possible pairs, that is $q + \binom{q}{2}$. The new vector $\tilde{\pi}_2(\gamma)$ can also be obtained by pre-multiplying the vector $\pi_2(\gamma)$ by an $s \times s(2)$ transformation matrix $\tilde{T}$ with full row rank, i.e., $\tilde{\pi}_2(\gamma) = \tilde{T}\pi_2(\gamma)$, where $s = q + \binom{q}{2}$. Pre-multiply $p_2$ and $\Delta(\gamma)$ in (4.20) by the transformation matrix $\tilde{T}$ to obtain the corresponding sample proportion vector $\hat{p}_2$ and the Jacobian matrix $\tilde{\Delta}(\gamma)$. Pre- and post-multiply $\Upsilon$ by $\tilde{T}$ to calculate the corresponding asymptotic covariance matrix $\tilde{\Upsilon}$. Then we can obtain a reduction on the original $M_2$ statistic,

$$
\tilde{M}_2 = N\tilde{e}_2(\hat{\gamma})'\tilde{\Delta}^{(c)}(\hat{\gamma})(\tilde{\Delta}^{(c)}(\hat{\gamma})'\tilde{\Upsilon}\tilde{\Delta}^{(c)}(\hat{\gamma}))^{-1}\tilde{\Delta}^{(c)}(\hat{\gamma})'\tilde{e}_2(\hat{\gamma}),
$$

where

$$
\tilde{e}_2(\hat{\gamma}) = \hat{p}_2 - \tilde{\pi}_2(\hat{\gamma}).
$$

Based on the theoretical results given in Maydeu-Olivares and Joe (2010), it can be proven that $\tilde{M}_2$ is asymptotically distributed as $\chi^2_{s-w}$, where $s = q + \binom{q}{2}$. We also conducted simulation studies to test the asymptotic properties of $\tilde{M}_2$.

It should be emphasized that using $\tilde{M}_2$ in place of $M_2$ does not simplify the computation of the asymptotic covariance matrix $\Upsilon$. In fact, it involves the same amount of calculation. However, we do not have to store the $s(2) \times s(2)$ asymptotic covariance matrix. Instead we store an $s \times s$ asymptotic covariance matrix to compute $\tilde{M}_2$. We consider the aforementioned example that has 35 ordinal variables with 7 categories per variable. For the original $M_2$, $s(2)$ is equal to 21,360. After the reduction, the size is $s = 630$ and the number of non-duplicated elements in the collapsed asymptotic covariance matrix is 198,765, which is much more manageable than 234.
million. Alternative reductions are also possible, given different transformation matrix $\tilde{T}$ with full row rank. The reader is referred to Maydeu-Olivares and Joe (2010, Section 2) for more details.
Chapter 5: Random Sampling Experiments

The current dissertation employs a constrained Fisher scoring type of algorithm to minimize the UBN fit function with suitable equality and inequality constraints. Detailed information of the constrained optimization procedure is given in Section 2.3. We implemented the whole estimation approach in FORTRAN 95 and developed a computer program UBN+ (please contact the author for copies of UBN+). This computer program estimates thresholds, factor loadings, and factor correlations (if the latent factors are correlated) using the constrained optimization algorithm. It also provides standard error estimates as well as the $M_2$ statistic with an associated $p$-value. The following simulation studies described in this chapter and the real data examples discussed in Chapter 6 are analyzed by using this computer program, run on a Dell PC with Intel(R) Core(TM)2 CPU at 2.40GHz and 2.00 GB of RAM, with Windows XP operational system.

5.1 One-Factor Model

The first random sampling experiment uses a one-factor model with six indicators. The model setup is very simple. The objective of the first example is threefold: evaluate consistency of the UBN estimates; examine the performance of the Godambe information matrix in providing accurate standard errors; and investigate the asymptotic distribution of the $M_2$ statistic.
5.1.1 Simulation Design

The first random sampling experiment uses simulated data from a modified factor analysis model with one latent factor. The factor has six ordinal variables as its indicators and each of them has four categories. The factor loading matrix is

\[ \Lambda' = \begin{pmatrix} 0.79 & 0.84 & 0.78 & 0.88 & 0.69 & 0.77 \end{pmatrix}. \]

The thresholds are listed in Table A.1, under a \( \tau^{(i)}_a \) column, where \( a \) goes from 1 to 3 and \( i \) goes from 1 to 6. Notice that \( \tau^{(i)}_0 \) is \(-\infty\) and \( \tau^{(i)}_4 \) is \(\infty\) for all \( i \). These are not included in this table.

To generate responses of one typical subject, the latent factor \( z \) was first generated from \( N(0, 1) \). The latent response variable \( x^*_i \) was then calculated,

\[ x^*_i = \lambda_{i1} z + u_i, \quad i = 1, \ldots, 6, \]

where \( u_i \) was generated independently from \( N(0, 1 - \lambda^2_{i1}) \). The variance of \( x^*_i \) is 1. We compared \( x^*_i \) to the population threshold values listed in Table A.1 and decided in which category the corresponding ordinal variable \( x_i \) falls,

\[ x_i = a \leftrightarrow \tau^{(i)}_{a-1} < x^*_i \leq \tau^{(i)}_a, \quad a = 1, \ldots, 4. \]

We independently repeated the above process \( N \) times, where \( N \) is the sample size. The \( N \) independently and identically generated responses formed one data set. The simulation process was also implemented in FORTRAN 95.

5.1.2 Simulation Results — Parameter Estimates and Standard Errors

In this simulation study, 1000 data sets were generated independently, and sample size \( N \) was first set at 200. The UBN estimates were obtained independently
for each data set, by calling the computer program UBN+. Results are summarized in Table A.1.

Table A.1 lists the true parameter value in the first column and the sample average of 1000 independent parameter estimates next to it, to make it easier to compare them. Following is the empirical standard error and the Godambe information based standard error of this parameter. The empirical standard error is calculated by

$$se_E = \sqrt{\frac{1}{1000 - 1} \sum_{l=1}^{1000} (\gamma_l - \overline{\gamma})^2},$$

where $\gamma_l$ denotes the UBN parameter estimate obtained from the $l$th data set and $\overline{\gamma}$ is the sample average of the 1000 UBN estimates on the same parameter,

$$\overline{\gamma} = \frac{1}{1000} \sum_{l=1}^{1000} \gamma_l.$$

The Godambe information based standard error $se_G$ is the sample mean of 1000 independent standard error estimates obtained from $\hat{G}(\hat{\gamma}_{ubn})$ using the corresponding UBN estimates for each data set. Although the true parameter values are known in this study, it is more reasonable to substitute the parameter estimates in $\hat{G}(\gamma)$ and see if the approximate Godambe information $\hat{G}(\hat{\gamma}_{ubn})$ provides accurate standard error estimates.

Comparing the population value to the mean estimate of each parameter in Table A.1 shows that the mean estimate is very close to the true population value, indicating consistency of the UBN estimates. The true population values are successfully recovered by the UBN approach. The closeness between the empirical standard errors and the Godambe information based standard error estimates demonstrates accuracy of the approximate Godambe information in providing standard error estimates.
The sample size was then increased to 1000. At this new sample size, 1000 independent data sets were generated as described before and the UBN estimates were calculated for each data set, the result of which is summarized in Table A.2. Listed in Table A.2 are population parameter value, sample average of 1000 UBN estimates, empirical standard error, and Godambe information based standard error. Table A.2 shows that the UBN approach provides consistent estimates for each parameter and the Godambe information based standard errors are close to the empirical standard errors. Additionally, by comparing Tables A.1 and A.2, it is clear that as sample size $N$ increases, the sample average of the UBN estimates are closer to the true value and the standard errors are decreasing. This implies that as the sample size goes up, the UBN approach provides more precise estimates.

5.1.3 Simulation Results — the $M_2$ Statistic

To evaluate asymptotic properties of the $M_2$ statistic on simulated data, we chose three sample size levels, i.e., $N=200, 1000, 100000$. At each sample size level, 1000 data sets were independently generated. For each data set, one $M_2$ statistic was obtained after the minimization process was completed. Under the current model setting, the dimensionality of the residual vector $\hat{e}_2$ defined in (4.19) is $s(2) = 153$. Given that there are 24 independent parameters in the model, the asymptotic distribution of $M_2$ is $\chi^2_{129}$. To see if the simulated $M_2$ follows such a distribution, we plotted a histogram of the 1000 $M_2$ values against the density curve of $\chi^2_{129}$, at each sample size level. The plots are shown in Figure A.1.

In Figure A.1, the left column shows histograms of the simulated $M_2$ value at sample sizes 200, 1000, and 100000, from top to bottom. The right column shows P-P plots of the corresponding p-value against Uniform(0,1) distribution at three sample
size levels. The $p$-value is obtained by computing the right tail probability of $\chi^2_{129}$ using the $M_2$ value as the cutoff point. The P-P plot, known as probability-probability plot or percent-percent plot, displays two cumulative distribution functions (CDFs) against each other. It is a graphical way to check whether these two continuous distributions are similar to each other. When the $M_2$ statistic is distributed as its asymptotic distribution $\chi^2_{129}$, the $p$-value will distribute evenly between 0 and 1, i.e., the $p$-value is distributed as Uniform(0,1). Hence, the P-P plot of the $p$-value against Uniform(0,1) is expected to form a straight line going through points (0,0) and (1,1), indicating that the two distributions are very close to each other.

Inspection of Figure A.1 shows that the three histograms capture the solid density curves approximately. Some discrepancy is observed between 120 and 140 on the $M_2$ continuum when $N$ is equal to 200. As sample size $N$ increases, the discrepancy diminishes, demonstrating improvement over large sample size. Similar improvement is also observed in the P-P plot. As $N$ goes up, the solid P-P plot is getting closer to the dashed straight line going through points (0,0) and (1,1). One thing to point out is that although the P-P plot converges slowly to the dashed straight line in the middle part, the convergence at the lower left corner is very fast. The lower left corner is the area where the model might be rejected, i.e., the $p$-value is less than 0.05, and it is important that the $p$-value distributes closely to Uniform(0,1) in this area. Hence, Figure A.1 confirms the expected asymptotic properties of $M_2$ and demonstrates its usability as a goodness-of-fit test statistic.

We also applied the Kolmogorov-Smirnov test (K-S test, c.f., Siegel & Castellan, 1988) to further compare distributions of the $p$-value and Uniform(0,1). The K-S test is a nonparametric test for the equality of two continuous one-dimensional probability distributions. It quantifies a distance between the empirical distribution
function of the sample and the CDF of the reference distribution. In this case, the empirical distribution of $p$-value is compared to the CDF of Uniform(0,1) by the K-S test. When $N=200$, the K-S distance is 0.044 with a $p$-value of 0.024, and we reject the null hypothesis that the two distributions are equal to each other. When $N$ is 1000, the K-S distance is 0.031 with a $p$-value of 0.291. Finally, as $N$ goes to 100000, the K-S distance is 0.0272 and the corresponding $p$-value is 0.4479. At two large sample sizes, the null hypothesis is retained. This again demonstrates that as $N$ increases, the distribution of $M_2$ is approaching its asymptotic distribution $\chi^2_{129}$.

5.1.4 Summary

The simulation study of the one-factor model demonstrates that the UBN approach can successfully recover the model parameters. The standard error estimates derived from the Godambe information are consistent with the standard errors we observe in the sample, and they decrease as the sample size goes up. The simulated $M_2$ statistic approximately follows its asymptotic distribution, and the similarity between the empirical distribution and the asymptotic distribution gets better when the sample size is increasing, which is shown in both the histograms of $M_2$ and the P-P plots of the corresponding $p$-value. The K-S test further confirms this conclusion. It shows that $M_2$ is a valid statistic for testing model fit for the one-factor model.

5.2 Two-Factor Model

In the second simulation study, we slightly increase the model complexity by employing a two-factor model with correlated factors. We intend to evaluate the UBN estimate by comparing it to the parameter estimates given by a full-information procedure. We choose the Expectation-Maximization (EM, c.f., Bock, & Aitkin, 1981)
algorithm implemented in the prototype program IRTPRO (Cai, du Toit & Thissen, forthcoming) as a reference method. The EM algorithm is a widely used method for finding MLE of parameters in statistical models. It iterates between an Expectation step, which computes the expectation of the log-likelihood given the current estimates, and a Maximization step, which computes parameter estimates by maximizing the expected log-likelihood function found on the E step. In IRTPRO, the EM algorithm is employed to maximize the true likelihood.

The computer program IRTPRO analyzes ordinal data within the Item Response Theory (IRT) framework. Different from the modified factor analysis model, IRT models the probability of an ordinal variable being in one category directly, without using a latent response variable. More specifically, in IRT, the conditional probability of the ordinal variable \( x_i \) being in category \( c \) given the latent variables \( \theta \) is defined as follows,

\[
P(x_i = c|\theta) = P(x_i \geq c|\theta) - P(x_i \geq c + 1|\theta) \\
= \frac{1}{1 + \exp[-(a_i'\theta + b_{jc})]} - \frac{1}{1 + \exp[-(a_i'\theta + b_{jc+1})]},
\]

where \( P(x_i \geq c|\theta) \) denotes the conditional probability of \( x_i \) being in category \( c \) or higher, \( a_i \) is a vector of slope parameters describing the strength of the relationship between the \( i \)th ordinal variable and the latent variables (notice that in the modified factor analysis model, \( a_i \) is used to denote the response category), and \( b_{jc} \) is a difficulty parameter denoting the point on the latent variable separating category \( c \) from category \( c + 1 \) (Wirth & Edwards, 2007).

Although the IRT framework shown in (5.1) appears very different from the modified factor analysis model described in (1.1) to (1.3), they are two closely related modeling frameworks. In fact, Takane and de Leeuw (1987) show that the
marginal probabilities of $x_i$ in the two modeling systems are approximately equal to each other (the equality holds for the normal ogive model in the IRT framework, which is computationally demanding and approximated by the logistic model (5.1)). There are transformations between slopes and factor loadings, and between difficulties and thresholds. Let $\lambda_i^*$ denote the $i$th row vector in the factor loading matrix that is transformed from the slope vector $a_i$, and $\tau_{ic}$ the threshold associated with response category $c$ of the $i$th ordinal variable that is transformed from the difficulty parameter $b_{jc}$. We have

$$\lambda_i^* = \frac{a_i}{\sqrt{D^2 + a_i'a_i}},$$

$$\tau_{ic} = -\frac{b_{jc}}{\sqrt{D^2 + a_i'a_i}},$$

where $D = 1.702$ is a scaling constant. Transformed parameters $\lambda_i^*$ and $\tau_{ic}$ are within the modified factor analysis model framework. In the following examples, we will compare the UBN estimates of $\lambda_{ij}$ and $\tau_{ic}^{(i)}$ with the estimates of $\lambda_i^*$ and $\tau_{ic}$ transformed from the EM result.

5.2.1 Simulation Design

The two-factor model assumes two correlated latent factors, each with five ordinal variables as its indicators. The correlation between the two factors is 0.4, and the factor loading matrix is

$$A' = \begin{pmatrix}
0.82 & 0.84 & 0.78 & 0.70 & 0.60 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.81 & 0.73 & 0.65 & 0.70 & 0.80
\end{pmatrix}.$$ 

Each ordinal variable has three response categories. The population threshold values are listed in Table A.3.
To simulate data from the two-factor model, we first generated a two-dimensional latent factors $z$ from $N_2(0, \Phi)$, where

$$
\Phi = \begin{pmatrix}
1 & 0.4 \\
0.4 & 1
\end{pmatrix}.
$$

The latent response variable $x^*_i$ was equal to

$$
x^*_i = \lambda_{i1} z_1 + \lambda_{i2} z_2 + u_i, \quad i = 1, \ldots, 10,
$$

where $u_i$ was simulated from $N(0, \psi_i)$, and $\psi_i$ is equal to 1 minus the $i$th diagonal element of $\Lambda \Phi \Lambda'$, making the variance of $x^*_i$ equal to 1. The value of $x_i$ was then decided by comparing $x^*_i$ to the population threshold values in Table A.3,

$$
x_i = a \iff \tau_{a-1}^{(i)} < x^*_i \leq \tau_a^{(i)}, \quad a = 1, 2, 3.
$$

We independently repeated the above process $N$ times, where $N$ is the sample size. The $N$ independently and identically generated responses formed one data set. In this example, the simulation process is implemented in R.

### 5.2.2 Simulation Results — Parameter Estimate Comparison

1000 independent data sets were generated at two sample size levels, i.e., $N=200$ and $N=1000$. Each data set was analyzed by both UBN implemented in FORTRAN and EM in IRTPRO. The parameter estimates obtained by EM were first transformed into the modified factor analysis framework by (5.2) and (5.3) and then compared to the UBN estimates.

Given the current model setting, there are 20 thresholds, 10 factor loadings, and 1 factor correlation. We compared the estimates given by UBN and EM for each parameter. 3 thresholds and 2 factor loadings were randomly selected to present...
the comparison. They are $\tau_{2}^{(2)}$, $\tau_{1}^{(6)}$, $\tau_{1}^{(7)}$, $\lambda_{21}$, and $\lambda_{72}$ (symbol $\tau_{a}^{(i)}$ represents the $a$th threshold on the $i$th ordinal variable and $\lambda_{ij}$ represents the factor loading of the $i$th latent response variable on the $j$th latent factor). The factor correlation is also included for comparison. Refer to subsection 5.2.1 and Table A.3 for the population values of these parameters.

Figures A.2 and A.3 present the parameter estimate comparison between UBN and EM. Each sub-figure plots the 1000 UBN estimates of the same parameter against the corresponding estimates given by EM, at specific sample size. The sub-figures in the left column of Figures A.2 and A.3 have $N=200$, the sub-figures in the right column have $N=1000$. Take the two plots on the first row in Figure A.2 as example. Both of them depict parameter estimates of $\tau_{2}^{(2)}$, the 2nd threshold on the 2nd ordinal variable. Each point in the plot has the transformed EM estimate as its $x$-coordinate and the corresponding UBN estimate, i.e., the UBN estimate based on the same data set, as its $y$-coordinate. When the estimates given by the two methods are close to each other, the plotted point is expected to fall on a straight line going through points (0,0) and (1,1), which is plotted in red as the reference line.

The first two plots in Figure A.2 show that the estimates of $\tau_{2}^{(2)}$ given by UBN and EM are very comparable. The correlation between the two sets of parameter estimates is 0.9928 when $N=200$ and 0.9942 when $N=1000$. The points in each sub-figure scatter along the reference line with some variation, and the center of the 1000 points is close to 0.16, the true value of $\tau_{2}^{(2)}$. This indicates that both methods provide consistent estimates. When the sample size $N$ is 1000, the variation in the parameter estimates is smaller than when $N$ is 200, and the difference between the two methods is also diminished. Therefore, as sample size increases, both the UBN approach and the EM algorithm provide more precise estimate.
The other sub-figures in Figures A.2 and A.3 show the same trends as described above. In fact, we found similar property in all the parameter estimates. We conclude that the UBN approach, as a limited information method, gives parameter estimates comparable to the EM algorithm implemented under the IRT framework.

5.2.3 Simulation Results — Standard Error Comparison

Both UBN and IRTPRO-EM provide standard error estimates. It is interesting to see if the standard error estimates given by the two approaches are similar. Because IRTPRO-EM is under the IRT framework, to compare standard errors given by UBN and EM also requires some additional calculation. The transformation of the slope, see equation (5.2), only involves the slope itself, while the transformation of the difficulty parameter involves both the slope and the difficulty. In this subsection we will only present the comparison result for the factor loadings, since the computation is simpler. An evaluation of standard error estimate for factor correlation is also included.

Apply equation (5.2) to the current two-factor model, we have

\[ \lambda^*_i = \frac{a_{i1}}{\sqrt{D^2 + a_{i1}^2}}, \quad i = 1, \ldots, 5 \]

\[ \lambda^*_i = \frac{a_{i2}}{\sqrt{D^2 + a_{i2}^2}}, \quad i = 6, \ldots, 10 \]

Suppose the standard error estimate of a slope parameter \( a \) is \( \sigma_a \). The Delta method indicates that the standard error of transformed factor loading \( \lambda^* \) will be

\[ \sigma_\lambda = \sigma_a \left| \frac{\partial \lambda^*}{\partial a} \right| = \sigma_a \frac{D^2}{(D^2 + a^2)^{3/2}}, \quad (5.4) \]
where $D=1.702$ is a scaling constant. In the following part, we applied (5.4) to transform the standard errors given by EM and compared the transformed standard errors to the Godambe information based standard errors employed in UBN.

There are 10 factor loadings. We randomly selected 3 factor loadings, $\lambda_{11}$, $\lambda_{62}$, and $\lambda_{82}$, to present the comparison graphically. In Figure A.4, the UBN standard error estimates of these factor loadings are plotted against the corresponding standard errors given by EM, after transformation. The left column shows the result at sample size 200, and the right column shows the result at sample size 1000. Take the two sub-figures on the first row as example. Both of them depict standard error estimates of $\lambda_{11}$, the factor loading of the 1st latent response variable on the 1st factor. Each point in the plot has the transformed standard error estimate from EM as its $x$-coordinate and the corresponding UBN standard error estimate, i.e., the Godambe information based standard error calculated using the same data set, as its $y$-coordinate. When the standard error estimates given by the two methods are close to each other, the plotted point is expected to fall on a straight line going through points (0,0) and (1,1), which is also plotted in red. These plots show that the two standard error estimates are in general very comparable. They scatter along the red reference line in both plots. The correlation between the two sets of standard error estimates is 0.6968 when $N=200$ and 0.7289 when $N=1000$. The standard error given by EM is slightly larger, in the third decimal place. When the sample size $N$ is 1000, variation in standard error estimates is greatly decreasing, and the difference between the two method is also diminished.

The other sub-figures in Figure A.4 and the standard error comparison of the other factor loadings, which is not shown in Figure A.4, have the similar property as
described above. It can be concluded that UBN and EM provide comparable standard error estimates for factor loadings.

Also compared are standard error estimates of factor correlation $\phi_{21}$. Computation details of the Godambe information based standard error estimate of $\phi_{21}$ are given in subsection 4.1.2. No transformation is needed, since IRTPRO-EM provides standard error estimates of $\phi_{21}$ directly. Figure A.5 depicts the comparison between these two methods. The left graph in A.5 shows the comparison at sample size $N=200$, and the right graph shows the comparison at sample size $N=1000$. Each point in the plot has the EM standard error estimate as its $x$-coordinate and the corresponding UBN standard error estimate as its $y$-coordinate. When the standard error estimates given by the two methods are close to each other, the plotted point is expected to fall on a straight line going through points (0,0) and (1,1), which is plotted in red. These plots show that the two standard error estimates are in general very comparable. Comparing the two graphs implies that as sample size increases, variation in standard error estimates is greatly decreasing, and the difference between the two methods is also diminished.

In addition, we also check the accuracy of the Godambe information based standard error estimate of $\phi_{21}$ in this example, since the first simulation study does not have any factor correlation coefficient to estimate. At sample size 200, sample mean of the 1000 standard error estimates of $\phi_{21}$ is 0.08057, while the empirical standard error of $\phi_{21}$ is 0.08250. When the sample size is increased to 1000, sample mean of the 1000 standard error estimates of $\phi_{21}$ is 0.03644, while the empirical standard error of $\phi_{21}$ is 0.03618. This shows that the Godambe information matrix provides accurate standard error estimates for factor correlations. Also, as sample size goes up, variation in parameter estimates is decreasing.
5.2.4 Simulation Results — $M_2$ and $\tilde{M}_2$ Statistics

In this simulation study, we implemented both the original $M_2$ statistic and the further reduced $\tilde{M}_2$ statistic, which is introduced in subsection 4.2.3, under the UBN approach. The dimensionality of the asymptotic covariance matrix involved in calculating $M_2$ is $200 \times 200$, and the dimensionality of the asymptotic covariance matrix involved in calculating $\tilde{M}_2$ is $55 \times 55$. The data sets used to compare UBN and EM estimation results are used here. For each data set, the UBN approach was carried out once. The UBN estimates were recorded for parameter comparison, as is shown in last two subsections, and both $M_2$ and $\tilde{M}_2$ were calculated. The asymptotic properties of $M_2$ and $\tilde{M}_2$ are examined under sample sizes 200 and 1000.

Figure A.6 plots histograms of $M_2$ against density curve of its asymptotic distribution, $\chi^2_{169}$, and P-P plots of the corresponding $p$-value, at two sample sizes. The histograms of $M_2$ generally capture the distributional property of $\chi^2_{169}$, at both sample sizes. The P-P plots are also close to the dashed straight line going through points (0,0) and (1,1). In addition, we carried out the K-S test on the simulated $p$-values. When $N=200$, the K-S test is not significant with the K-S distance being 0.0354 and the $p$-value being 0.1639. When $N=1000$, the K-S test is also insignificant with the K-S distance being 0.0234 and the $p$-value being 0.7637. There is some improvement over a larger sample size.

Figure A.7 plots histograms of the further reduced $\tilde{M}_2$ against density curve of its asymptotic distribution, $\chi^2_{24}$, and P-P plots of the corresponding $p$-value, at two sample sizes. The histograms of $\tilde{M}_2$ generally capture the distributional property of $\chi^2_{24}$, at both sample sizes. The P-P plots are also close to the dashed straight line going through points (0,0) and (1,1). When $N=200$, the K-S distance is 0.0191 with
the $p$-value being 0.8582. When $N=1000$, the K-S distance is 0.0251 with the $p$-value being 0.5532. Based on Figures A.6 and A.7 and the K-S tests, we conclude that both $M_2$ and $\tilde{M}_2$ are appropriate for testing model fit. The computation time of obtaining $M_2$ and $\tilde{M}_2$ for one data set are also about the same, both around 20 seconds. The major advantage of $\tilde{M}_2$ over $M_2$ is that it requires less memory space to store the asymptotic covariance matrix when calculating $\tilde{M}_2$. But the computation complexity of the two test statistics are about the same.

5.2.5 Summary

The random sampling experiment of the two-factor model demonstrates that the UBN approach provides parameter estimates and standard error estimates that are comparable to a full information procedure. The reference method we used in this example is the EM algorithm implemented in IRTPRO. Besides, this simulation study shows that both the original $M_2$ statistic proposed by Maydeu-Olivares and Joe (2006) and a further reduced $\tilde{M}_2$ statistic are appropriate for the purpose of testing model fit. When there are too many ordered categories per variable, $\tilde{M}_2$ requires a smaller memory space to store the asymptotic covariance matrix and is therefore more manageable. However, calculating $\tilde{M}_2$ does need same amount of computation time as $M_2$.

5.3 A Short Study of Computation Time

Jöreskog and Moustaki (2001) mention that the computer time of UBN increases as number of variables, number of categories, or number of factors increases. And the relationship is approximately linear (refer to Table 27 in Jöreskog & Moustaki, 2001). In contrast, the computer time is constant when the sample size changes.
In the current research, the FORTRAN program written for the UBN approach involves three parts: minimizing the UBN fit function; calculating standard error estimates; and computing the $M_2$ statistic. The sampling experiments we have conducted show that the first two parts are generally fast and insensitive to sample size change. Computing $M_2$, on the other hand, consumes most of the computation time. Take the two-factor model described in subsection 5.2 as an example. When the sample size $N$ is 200, the total computation time is 22.4 seconds, among which minimizing the UBN fit function and calculating standard error estimates consumes 0.2 seconds each, and computing $M_2$ uses the remaining 22 seconds. The computation time does not change much when $N$ is increased to 1000. For the same model, when $N$ is 200, the EM algorithm in IRTPRO requires 4.09 seconds to finish E-step, M-step, and standard error computation. When $N$ is 1000, it takes 12.26 seconds to complete the same computation tasks in EM. If we only consider the computation time needed for estimating model and calculating standard errors, UBN is very fast for this simulation example.

To further examine change in computation time with regard to possible factors other than sample size, we conducted the third simulation study. We are mostly interested in computation time of the first two parts, and the $M_2$ statistic is not computed in this study.

### 5.3.1 Models Setup

We employed two model structures in this study. One has five independent latent factors, each of which has five ordinal variables as its indicators. Totally there are 25 items in the first model structure. We call this model the independent-cluster model. The other one assumes a bi-factor structure with one general factor that
exerts effect on every ordinal variable and four specific factors that have effects on independent groups of five ordinal variables. There are 20 items in this model. It is assumed that the general factor is independent from the specific factors, and the specific factors are independent from each other. This model is labeled the bi-factor model. The factor loadings of independent-cluster and bi-factor models are given in Tables A.4 and A.5.

Two levels of category size are considered, i.e., two categories per variable and five categories per variable. Therefore four different models are under investigation. They are the independent-cluster model with two categories per variable, the independent-cluster model with five categories per variable, the bi-factor model with two categories per variable, and the bi-factor model with five categories per variable. The sample size is 200 for all four models.

5.3.2 Simulation Results — Computation Time

Under each of the four models, 10 independent data sets of sample size 200 were generated and analyzed by the UBN approach. No $M_2$ statistic was calculated. The computation time of minimizing the UBN fit function plus the computation time of calculating standard errors was recorded. The average computation time of the 10 runs is given in Table A.6, on the row of “Time in UBN”. The same data sets were also analyzed by the Metropolis-Hastings Robbins-Monro (MH-RM; c.f., Cai, 2010) method implemented in IRTPRO. The MH-RM algorithm is a synthesis of the Metropolis-Hastings (MH; Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller & Teller, 1953) sampler and the Robbins-Monro (RM, Robbins & Monro, 1951) stochastic approximation (SA) algorithm. MH-RM is employed to replace the deterministic Gaussian quadrature method in the EM procedure (Cai, 2010). When
the number of latent factors is large, MH-RM is computationally faster than the existing methods. It is applied as a reference method in the current study to evaluate the computation time of UBN. The average computation time of MH-RM is also listed in Table A.6, on the row of “Time in MH-RM”.

Table A.6 shows that the computation time of UBN is greatly affected by the number of categories in the model. Under the independent-cluster model, the average computation time is only 1.03 seconds when there are two categories per variable. It increases to more than 18 seconds when there are five categories per variable. Such a jump in computation time also happens in the bi-factor model. When the category size is two, the average computation time is 2.06 seconds, which increases to 13.58 seconds when the category size is five. Between the independent-cluster and the bi-factor models, the average computation time is longer for the bi-factor model when there are two categories per variable, but when we increase the number of categories to five, the average computation time is longer for the independent-cluster model. It appears that the computation time is not linearly related to the number of parameters. It also depends on the model structure and the relationship among the parameters.

The computation time used by the MH-RM method does not change very much, all around 20 seconds. Within each model structure, increasing the number of categories does increase the computation time for the MH-RM method, but only slightly. Between the two model structures, the computation time is longer for the bi-factor model when there are two categories per variable. When the number of categories is increased to five, the computation time is longer for the independent-cluster model. This is same to what we found for the UBN approach. Comparing computation time between UBN and MH-RM shows that under the current model
settings, the UBN approach is very efficient in providing parameter estimates and standard errors.

It should be emphasized that the current simulation study is far from a thorough investigation of the computation time issue. In fact, there are too many possibilities in combining number of factors, number of ordinal variables, and number of categories per variable to construct different models. What we found from the two-factor model simulation study in Section 5.2 and the current simulation study is that UBN is generally efficient in minimizing the UBN fit function and calculating estimates. The computation time of UBN is approximately constant when only the sample size changes. It does vary a lot with the number of factor, number of variables, and number of categories per variable. The relationship between the computation time of UBN and the number of model parameters is not linear. The model structure and the relationship among the parameters also affects the computation time.
Chapter 6: Applications

In this chapter, the UBN approach will be illustrated using two real data examples. The first example is a class test of an introductory psychology course. The second example is a psychological measurement of patient’s general well-being status. In both examples, the UBN estimates will be given and compared to corresponding full-information estimates.

6.1 PSYCH101 Class Example

The first real data illustration is used as a demonstration example of TestGraf98 (Ramsay, 2001), a standalone program designed for analyzing tests and questionnaires. The original data set includes 379 students’ responses to 100 dichotomous items of an introductory PSYCH101 course, in the Christmas exam period of 1989 at McGill. Dichotomous item is a special case of ordinal variable and it has two ordered response categories. The manual of TestGraf98 (Ramsay, 2000) points out that 4 items in the original test have severe problems. Besides, a preliminary analysis of the original data set indicates that there are several items associated with extreme threshold estimates. Therefore, 8 items with extreme threshold estimates and the 4 items with severe problems were deleted, leaving 88 items to be analyzed by the UBN approach.

We specified a modified factor analysis model with one factor to this example, and the factor is knowledge in this PSYCH101 course. We employed the UBN
approach implemented in FORTRAN to analyze the data set. For each item, one threshold, $\tau^{(i)}_1$, and one factor loading of the corresponding latent response variable on the latent factor, $\lambda_{i1}$, are estimated.

The UBN estimates of $\tau^{(i)}_1$ and $\lambda_{i1}$ are given in Table A.7, followed by the Godambe information based standard error estimates enveloped by a pair of parentheses. Table A.7 shows that most of the UBN estimates and the standard error estimates are of reasonable sizes. The threshold estimates range from -1.527 to 0.830. The factor loading estimates are all positive, which indicates that knowing required material of this PSYCH101 course has positive effect on endorsing the questions. Some factor loading estimates are small. For example, the latent response variable that corresponds to item 52, i.e., $x^*_52$, has a factor loading of only 0.037 on the latent factor. The corresponding standard error estimate is 0.068. It appears that knowledge in this PSYCH101 class has almost no effect on endorsing this item. Item 62 also has a small factor loading, which is 0.103. These findings suggest that some modification on items 52 and 62 might be necessary.

The PSYCH101 data set was also analyzed by the EM algorithm implemented in IRTPRO. Parameter estimates given by EM were first transformed into the modified factor analysis model framework by applying (5.2) and (5.3) and then compared to the UBN estimates. Figure A.8 plots the parameter estimates given by UBN against those given by EM, in two separate graphs. The left graph in Figure A.8 plots the 88 factor loading estimates, and the right one plots the 88 threshold estimates. The horizontal axis denotes the transformed estimates provided by EM, and the vertical axis denotes the estimates provided by UBN. If the two sets of parameter estimates are close to each other, the dots in the graph should form a straight line going through points (0,0) and (1,1), i.e., the reference line shown in red. Figure
A.8 indicates that the threshold estimates obtained from the two programs are very close. The points in the right graph fall on the red reference line closely. The factor loading estimates, however, display some discrepancy between the two methods. The correlation between the two sets of threshold estimates is 0.9997, and the correlation between the two sets of factor loading estimates is 0.9777. In general, the UBN estimates are reasonably close to the estimates given by the EM algorithm.

The $M_2$ statistic given by UBN is 4121.56, with the degree of freedom being 3740. The corresponding $p$-value is 0.00. The modified factor analysis model with one factor is actually rejected based on the $M_2$ statistic. The calculation of $M_2$ relies on a 3916×3916 weight matrix, and its size is much larger than the sample size we have, $N = 379$. In this case, the weight matrix may be unstable, and it could be one possible reason why the one-factor model is rejected.

### 6.2 Quality of Life Interview Example

The second real data example of the UBN approach for ordinal variables contains 586 patients’ responses to the “Quality of Life Interview for the Chronically Mentally Ill” (Lehman, 1988), a 45-minute psychometric measurement on the general well-being. The data set is used in Gibbons et al. (2007) as an illustration of the bi-factor model. The Quality of Life (QOL) scale analyzed here consists of one item evaluating global life satisfaction and 34 items representing seven subdomains. The subdomains are Family, Finance, Health, Leisure, Living, Safety, and Social (Gibbons et al., 2007). Each item has 7 categories (1=terrible through 7=delighted). Therefore the model structure of the QOL data set is similar to the bi-factor model we used in the simulation study of computation time: one general factor $G$ has effect on every item, and seven specific factors $F_1$ to $F_7$ have effect on independent groups of items.
The UBN estimates of the QOL data, as well as the corresponding standard error estimates (shown in parentheses), are given in Tables A.8 and A.9. Table A.8 displays the factor loading estimates, and Table A.9 displays the threshold estimates. The item descriptions are duplicated from Table 1 in Gibbons et al. (2007). From the item descriptions, it is clear that item 1 is a general evaluation of subject’s well-being and items 2 to 35 represent different aspects of the quality of life. Table A.8 shows that all items have substantial positive factor loadings on the general factor $G$, indicating that all items are positively related to general satisfaction of life. The factor loadings of items 2 to 35 on the specific factors $F_1$ to $F_7$ are also of reasonable size. Items 15 and 21 have factor loadings that are smaller than 0.200. This suggests that some further investigation on these two items might be needed. The threshold estimates listed in Table A.9 provide information on how the ordinal variable value is determined from the corresponding latent response variable. No specific problem is found in this Table.

The QOL data set was also analyzed by the MH-RM method implemented in IRTPRO so that we can compare the UBN estimates to the full-information parameter estimates given by MH-RM. Cai (2010) mentions that MH-RM can be less efficient than the EM algorithm for sufficiently low dimensional problems, but it is far more efficient for high-dimensional problems. Because the dimensionality of the latent construct is eight, the MH-RM method is expected to be much more efficient than the EM approach. It is of interest to see how efficient UBN is when compared to MH-RM.

Similar to the PSYCH101 class example, the parameter estimates obtained by MH-RM were first transformed into the modified factor analysis model framework by (5.2) and (5.3) and then compared to the UBN estimate. Figure A.9 displays the
parameter comparison between UBN and MH-RM implemented in IRTPRO, in two separate graphs. The left graph plots the 69 UBN factor loading estimates against the corresponding estimates obtained from MH-RM. The right graph plots the 210 UBN threshold estimates against the corresponding estimates obtained from MH-RM. In both plots, the horizontal axis denotes the transformed estimates from MH-RM, and the vertical axis denotes the UBN estimates. If the two sets of parameter estimate are close to each other, the dots in the graphs are expected to form a straight line that goes through points (0,0) and (1,1), which is shown in red in Figure A.9. It is shown in the plots that the threshold estimates obtained from the two programs are very close to each other. They fall on the reference line very closely. There is some discrepancy between the UBN factor loading estimates and the MH-RM estimates. It appears that the MH-RM estimates of the factor loading tend to be larger than the UBN estimates. The correlation between the two sets of threshold estimates is 0.9996, and the correlation between the two sets of factor loading estimates is 0.9711. In general, UBN provides parameter estimates that are comparable to the full-information estimation procedure MH-RM.

With regard to computation time, it took 103 seconds for MH-RM to compute the parameter estimates and the standard error estimates, and it took 608 seconds for UBN+ to finish the same computation. The computation time of UBN is acceptable but much longer than that of MH-RM. This shows that the MH-RM approach is much more efficient in providing parameter estimates when the dimensionality of the latent construct is large. No $M_2$ statistic was calculated, because it involves a huge asymptotic covariance matrix that requires considerable computation time. $\tilde{M}_2$ is not calculated either, because it requires same amount of computation as $M_2$. This suggests that the computational burden associated with $M_2$ or the further reduced $\tilde{M}_2$
is quite large when there are many ordinal variables or many categories per variable in the problem. Further research on constructing a limited-information goodness-of-fit test statistic with affordable computation time is still necessary.
Chapter 7: Summary and Future Directions

7.1 Summary

The main contribution of the current study is to provide accurate standard error estimates and a limited-information goodness-of-fit test statistic for the UBN approach. The UBN fit function first proposed by Jöreskog and Moustaki (2001) is proven to be a special case of the composite likelihood, which encompasses various limited-information approaches derived from marginal and/or conditional events of the data. The composite likelihoods are important surrogates for the ordinary full-likelihood when it is too cumbersome or impractical to compute (Varin, 2008). We developed a computer program UBN+ written in FORTRAN 95 to provide estimation results and test statistic. We did more than 8000 runs of the program using simulated data sets, and the computer program works reliably.

In this dissertation, the UBN approach is applied to analyze the modified factor analysis model to multivariate ordinal variables. It computes parameter estimates in one single stage by minimizing the UBN fit function defined in Jöreskog and Moustaki (2001). A constrained optimization procedure (Browne & du Toit, 1992) is employed to carry out the minimization task. The UBN estimates are proven to be consistent and asymptotically normal when the model fits.

To make statistical inference on the UBN estimates and model fit, one needs to take the limited-information feature of UBN into consideration. Therefore, the
usual Fisher information developed for full-likelihood procedure is not appropriate. Instead, the Godambe information is employed to provide standard error estimates for UBN. The Godambe information is a general approach in providing standard errors to composite likelihood procedures. Simulation studies show that the Godambe information gives accurate standard error estimates, and the result becomes more precise when sample size increases.

We apply a residual based test statistic $M_2$ (c.f., Maydeu-Olivares & Joe, 2006) to evaluate the model fit after obtaining the UBN estimates. The $M_2$ statistic utilizes the same model fitted probabilities and observed sample proportions that are involved in computing the UBN fit function. When the hypothesized model fits, $M_2$ is asymptotically central Chi-square distributed. Simulation studies demonstrate $M_2$’s usability as a goodness-of-fit test statistic in evaluating the modified factor analysis model to ordinal data.

In addition, both the UBN parameter estimates and the Godambe information based standard error estimates are compared with corresponding full-information estimates provided by the prototype computer program IRTPRO (Cai et al., forthcoming). The comparison shows that the UBN approach gives estimates comparable to its full-information counterparts and is generally very efficient.

### 7.2 Future Directions

The composite likelihood approaches based on modifications of the true likelihood function have been used in application fields like longitudinal study, time series model, missing data, and so forth. It will be interesting to apply similar composite likelihood procedures on the modified factor analysis model to ordinal data and
compare the estimation result with UBN. Aspects like closeness between the estimates and computation time issue shall be considered when doing such comparison. Besides, it will also be interesting to apply the UBN approach to more general structural equation modeling problem and evaluate its performance by simulation study and real data example.

Another topic worth further investigation is the limited-information test statistic. The current $M_2$ statistic requires calculating an asymptotic covariance matrix of the residual vector, which is of a large dimensionality and computationally demanding when there are too many ordinal variables or too many categories per variable. Maydeu-Olivares and Joe (2006) point out that there is an alternative limited-information test statistic that is specific to ordinal data. L. Cai (personal communication, January 11, 2011) also suggests a further reduction on the original $M_2$, which collapses the ordered categories of one variable into one single number. According to Maydeu-Olivares and Joe (2010), such a reduction on $M_2$ is also asymptotically Chi-square distributed. Simulation studies have shown that this reduced $\tilde{M}_2$ statistic suggested by Cai also follows the Chi-square distribution and can be used as a goodness-of-fit test statistic. However, it does not have advantage over $M_2$ with regard to computation complexity, because it requires the same amount of information to calculate an asymptotic covariance matrix.

Therefore, investigation on the limited-information goodness-of-fit test statistic for ordinal data is still a promising research area. Test statistics that have necessary asymptotic properties with an affordable computation time are in need.
Bibliography


### Table A.1: Simulation results for the one-factor model, at sample size $N=200$. Listed in the table are population parameter values, corresponding sample mean of 1000 UBN estimates, empirical standard error estimates, and Godambe information based standard error estimates.
<table>
<thead>
<tr>
<th>Item #</th>
<th>$\lambda_{i1}$</th>
<th>$\hat{\lambda}_{i1}$</th>
<th>$se_{E}$</th>
<th>$se_{G}$</th>
<th>$\tau_1^{(i)}$</th>
<th>$\hat{\tau}_1^{(i)}$</th>
<th>$se_{E}$</th>
<th>$se_{G}$</th>
<th>$\tau_2^{(i)}$</th>
<th>$\hat{\tau}_2^{(i)}$</th>
<th>$se_{E}$</th>
<th>$se_{G}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i=1$</td>
<td>0.79</td>
<td>0.790</td>
<td>0.017</td>
<td>0.017</td>
<td>-0.83</td>
<td>-0.830</td>
<td>0.047</td>
<td>0.045</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i=2$</td>
<td>0.84</td>
<td>0.840</td>
<td>0.015</td>
<td>0.014</td>
<td>-0.87</td>
<td>-0.872</td>
<td>0.045</td>
<td>0.046</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i=3$</td>
<td>0.78</td>
<td>0.779</td>
<td>0.018</td>
<td>0.018</td>
<td>-1.17</td>
<td>-1.171</td>
<td>0.054</td>
<td>0.051</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i=4$</td>
<td>0.88</td>
<td>0.880</td>
<td>0.013</td>
<td>0.013</td>
<td>-0.45</td>
<td>-0.451</td>
<td>0.042</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i=5$</td>
<td>0.69</td>
<td>0.689</td>
<td>0.022</td>
<td>0.022</td>
<td>-0.75</td>
<td>-0.753</td>
<td>0.046</td>
<td>0.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i=6$</td>
<td>0.77</td>
<td>0.770</td>
<td>0.019</td>
<td>0.019</td>
<td>-0.18</td>
<td>-0.181</td>
<td>0.041</td>
<td>0.040</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.2: Simulation results for the one-factor model, at sample size $N=1000$. Listed in the table are population parameter values, corresponding sample mean of 1000 UBN estimates, empirical standard error estimates, and Godambe information based standard error estimates.

<table>
<thead>
<tr>
<th>Item #</th>
<th>$\tau_1^{(i)}$</th>
<th>$\hat{\tau}_1^{(i)}$</th>
<th>$se_{E}$</th>
<th>$se_{G}$</th>
<th>$\tau_2^{(i)}$</th>
<th>$\hat{\tau}_2^{(i)}$</th>
<th>$se_{E}$</th>
<th>$se_{G}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i=1$</td>
<td>-0.19</td>
<td>-0.190</td>
<td>0.040</td>
<td>0.040</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i=2$</td>
<td>-0.16</td>
<td>-0.159</td>
<td>0.040</td>
<td>0.039</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i=3$</td>
<td>-0.50</td>
<td>-0.500</td>
<td>0.042</td>
<td>0.041</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i=4$</td>
<td>0.19</td>
<td>0.189</td>
<td>0.039</td>
<td>0.039</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i=5$</td>
<td>-0.21</td>
<td>-0.212</td>
<td>0.041</td>
<td>0.040</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i=6$</td>
<td>0.39</td>
<td>0.391</td>
<td>0.040</td>
<td>0.040</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.3: Population threshold values used to generate data for the two-factor model.
Table A.4: Factor loadings of the independent-cluster model. The five factors $F_1$ to $F_5$ are independent. The blank spaces indicate that the factor loadings are zero.

<table>
<thead>
<tr>
<th></th>
<th>$F_1$</th>
<th>$F_2$</th>
<th>$F_3$</th>
<th>$F_4$</th>
<th>$F_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1^*$</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_2^*$</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_3^*$</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_4^*$</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_5^*$</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_6^*$</td>
<td></td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_7^*$</td>
<td></td>
<td>0.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_8^*$</td>
<td></td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_9^*$</td>
<td></td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{10}^*$</td>
<td></td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{11}^*$</td>
<td></td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{12}^*$</td>
<td></td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{13}^*$</td>
<td></td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{14}^*$</td>
<td></td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{15}^*$</td>
<td></td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{16}^*$</td>
<td></td>
<td></td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{17}^*$</td>
<td></td>
<td></td>
<td>0.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{18}^*$</td>
<td></td>
<td></td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{19}^*$</td>
<td></td>
<td></td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{20}^*$</td>
<td></td>
<td></td>
<td></td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>$x_{21}^*$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.82</td>
</tr>
<tr>
<td>$x_{22}^*$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.84</td>
</tr>
<tr>
<td>$x_{23}^*$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.78</td>
</tr>
<tr>
<td>$x_{24}^*$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.70</td>
</tr>
<tr>
<td>$x_{25}^*$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>$G$</td>
<td>$F_1$</td>
<td>$F_2$</td>
<td>$F_3$</td>
<td>$F_4$</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>$x_1^*$</td>
<td>0.51</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_2^*$</td>
<td>0.52</td>
<td>0.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_3^*$</td>
<td>0.50</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_4^*$</td>
<td>0.45</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_5^*$</td>
<td>0.50</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_6^*$</td>
<td>0.51</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_7^*$</td>
<td>0.52</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_8^*$</td>
<td>0.50</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_9^*$</td>
<td>0.45</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{10}^*$</td>
<td>0.50</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{11}^*$</td>
<td>0.51</td>
<td></td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{12}^*$</td>
<td>0.52</td>
<td></td>
<td>0.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{13}^*$</td>
<td>0.50</td>
<td></td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{14}^*$</td>
<td>0.45</td>
<td></td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{15}^*$</td>
<td>0.50</td>
<td></td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_{16}^*$</td>
<td>0.51</td>
<td></td>
<td></td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>$x_{17}^*$</td>
<td>0.52</td>
<td></td>
<td></td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>$x_{18}^*$</td>
<td>0.50</td>
<td></td>
<td></td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>$x_{19}^*$</td>
<td>0.45</td>
<td></td>
<td></td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>$x_{20}^*$</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table A.5: Factor loadings of the bi-factor model. Factor $G$ is the general factor. Factors $F_1$ to $F_4$ are independent specific factors, which are also independent from $G$. The blank spaces indicate that the factor loadings are zero.

<table>
<thead>
<tr>
<th></th>
<th>Independent-cluster model</th>
<th>Bi-factor model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 categories</td>
<td>5 categories</td>
</tr>
<tr>
<td># of factor loadings</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td># of thresholds</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>Total # of parameters</td>
<td>50</td>
<td>125</td>
</tr>
<tr>
<td>Time in UBN</td>
<td>1.03s</td>
<td>18.30s</td>
</tr>
<tr>
<td>Time in MH-RM</td>
<td>19.52s</td>
<td>25.23s</td>
</tr>
</tbody>
</table>

Table A.6: Computation time of independent-cluster model with two categories per variable, independent-cluster model with five categories per variable, bi-factor model with two categories per variable, and bi-factor model with five categories per variable. Numbers of model parameters are listed under each model. The average computation time (in seconds) of ten runs of UBN and MH-RM are listed in the last two rows.
Table A.7: The UBN estimates of the PSYCH101 data set. Numbers in parentheses are corresponding standard error estimates.
<table>
<thead>
<tr>
<th>Item #</th>
<th>Item Description</th>
<th>G</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Life as a whole</td>
<td>0.756(0.024)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Family</td>
<td>0.505(0.035)</td>
<td>0.621(0.037)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Amount of family contact</td>
<td>0.561(0.034)</td>
<td>0.470(0.043)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Family with interaction</td>
<td>0.595(0.033)</td>
<td>0.685(0.034)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>General family stuff</td>
<td>0.641(0.028)</td>
<td>0.618(0.039)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Total money you get</td>
<td>0.506(0.038)</td>
<td>0.680(0.034)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Amount pay for basic needs</td>
<td>0.433(0.041)</td>
<td>0.519(0.042)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Money for fun</td>
<td>0.548(0.035)</td>
<td>0.664(0.033)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Financial well-being</td>
<td>0.573(0.035)</td>
<td>0.657(0.036)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Health in general</td>
<td>0.516(0.039)</td>
<td></td>
<td>0.232(0.058)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Medical Care</td>
<td>0.541(0.037)</td>
<td></td>
<td>0.542(0.052)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>How often see doctor</td>
<td>0.501(0.039)</td>
<td></td>
<td>0.512(0.056)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Talk to therapist</td>
<td>0.550(0.037)</td>
<td></td>
<td>0.517(0.049)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Physical condition</td>
<td>0.618(0.033)</td>
<td></td>
<td>0.265(0.056)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Emotional well-being</td>
<td>0.662(0.028)</td>
<td></td>
<td>0.151(0.054)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Way spend free time</td>
<td>0.680(0.028)</td>
<td></td>
<td>0.291(0.048)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Amount of free time</td>
<td>0.566(0.037)</td>
<td></td>
<td>0.353(0.052)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Chance to enjoy time</td>
<td>0.602(0.034)</td>
<td></td>
<td>0.409(0.056)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Amount of fun</td>
<td>0.633(0.031)</td>
<td></td>
<td>0.562(0.052)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Amount of relaxation</td>
<td>0.565(0.037)</td>
<td></td>
<td>0.410(0.056)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Pleasure from TV</td>
<td>0.464(0.044)</td>
<td></td>
<td>0.106(0.062)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Living arrangements</td>
<td>0.548(0.037)</td>
<td></td>
<td></td>
<td>0.518(0.041)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Food</td>
<td>0.464(0.040)</td>
<td></td>
<td></td>
<td>0.467(0.043)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Privacy</td>
<td>0.499(0.039)</td>
<td></td>
<td></td>
<td>0.632(0.036)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Amount of freedom</td>
<td>0.513(0.040)</td>
<td></td>
<td></td>
<td>0.652(0.040)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Prospect of staying</td>
<td>0.482(0.036)</td>
<td></td>
<td></td>
<td>0.551(0.036)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Neighborhood safety</td>
<td>0.557(0.040)</td>
<td></td>
<td></td>
<td></td>
<td>0.587(0.047)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Safe at home</td>
<td>0.588(0.035)</td>
<td></td>
<td></td>
<td></td>
<td>0.526(0.047)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Police access</td>
<td>0.497(0.042)</td>
<td></td>
<td></td>
<td></td>
<td>0.337(0.053)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Protect robbed/attack</td>
<td>0.551(0.038)</td>
<td></td>
<td></td>
<td></td>
<td>0.515(0.046)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Personal safety</td>
<td>0.589(0.036)</td>
<td></td>
<td></td>
<td></td>
<td>0.438(0.048)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Do things with others</td>
<td>0.572(0.036)</td>
<td></td>
<td></td>
<td></td>
<td>0.654(0.056)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Time with others</td>
<td>0.629(0.031)</td>
<td></td>
<td></td>
<td></td>
<td>0.456(0.057)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Social interactions</td>
<td>0.525(0.040)</td>
<td></td>
<td></td>
<td></td>
<td>0.446(0.048)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>People in general</td>
<td>0.419(0.044)</td>
<td></td>
<td></td>
<td></td>
<td>0.226(0.061)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.8: The UBN factor loading estimates of the Quality of Life data set. Item descriptions are duplicated from Table 1 in Gibbons et al. (2007). G stands for the general factor, F1-F7 represent the 7 subdomains of Family, Finance, Health, Leisure, Living, Safety, and Social. Numbers in parentheses are corresponding standard error estimates.
Table A.9: The UBN threshold estimates of the Quality of Life data set. Item descriptions are duplicated from Table 1 in Gibbons et al. (2007). Numbers in parentheses are corresponding standard error estimates.
Figure A.1: Histogram of simulated $M_2$ statistic and P-P plot of corresponding $p$-value for the one-factor model. Left column: histogram of $M_2$. Right column: P-P plot of $p$-value. From top to bottom: sample size $N=200$, $N=1000$, and $N=100000$. 
Figure A.2: Scatter plot of UBN estimates against estimates given by IRTPRO-EM, after transformation. Left column: sample size $N=200$. Right column: sample size $N=1000$. From top to bottom: parameter estimates of $\tau_{22}^{(2)}$, $\tau_{61}^{(6)}$, and $\tau_{71}^{(7)}$. 

87
Figure A.3: Scatter plot of UBN estimates against estimates given by IRTPRO-EM, after transformation. Left column: sample size $N=200$. Right column: sample size $N=1000$. From top to bottom: parameter estimates of $\lambda_{31}$, $\lambda_{72}$, and $\phi_{12}$. 
Figure A.4: Comparison of Godambe information based standard error estimates against standard error estimates given by IRTPRO-EM, after transformation. Left column: sample size $N=200$. Right column: sample size $N=1000$. From top to bottom: standard error estimates of $\lambda_{11}$, $\lambda_{62}$, and $\lambda_{82}$. 
Figure A.5: Comparison of Godambe information based standard error estimates against standard error estimates given by IRTPRO-EM, for parameter $\phi_{21}$. Left graph: sample size $N=200$. Right graph: sample size $N=1000$. 
Figure A.6: Histogram of simulated $M_2$ statistic and P-P plot of corresponding p-value for the two-factor model, at sample size $N = 200, 1000$. Left column: histogram of $M_2$. Right column: P-P plot of p-value.
Figure A.7: Histogram of simulated $\tilde{M}_2$ statistic and P-P plot of corresponding $p$-value for the two-factor model, at sample size $N = 200, 1000$. Left column: histogram of $\tilde{M}_2$. Right column: P-P plot of $p$-value.
Figure A.8: The PSYCH101 data set. Parameter estimate comparison between UBN and IRTPRO-EM. The left graph pictures the comparison of factor loading estimates, and the right one pictures the comparison of threshold estimates.

Figure A.9: The Quality of Life data set. Parameter estimate comparison between UBN and IRTPRO-MHRM. The left graph pictures the comparison of factor loading estimates, and the right one pictures the comparison of threshold estimates.