Confirmatory Factor Analysis with Continuous and Ordinal Data: An Empirical Study of Stress Level

By Yingruolan Li

Department of Statistics
Uppsala University

Supervisor: Katrin Kraus

2014
Abstract

Ordinal variables are often used in questionnaires in social science. However, most sta-
tistical methods are designed to deal with continuous data. Confirmatory Factor Analysis
(CFA), a method that is frequently used in social science, is applicable for multinormally
distributed continuous data. Theoretical developments have been made so that CFA can be
used for non-multinormal continuous data and ordinal data as well. This thesis explains
some proper ways to deal with non-multinormal continuous data and ordinal data. Em-
pirical data obtained from special designed questionnaires measure stress level of a group
of students in a basic statistical class. These questionnaires contain both continuous and
ordinal scales for every question. A two-factor CFA model is applied to the data and four
estimation methods are used to estimate the model. Model fit statistics and parameter esti-
mates are compared between types of data and numbers of categories in questionnaires. A
simulation study shows that continuous data perform better in parameter estimates, stan-
dard errors and model fit than 7-category and 4-category data. Estimation methods bring
different effects on the above three kinds of data. The results can not be illustrated by the
empirical study. Possible reasons for that are discussed.

Keywords: questionnaire  continuous data  ordinal data  CFA  model fit
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>2 Research Question</strong></td>
<td>2</td>
</tr>
<tr>
<td><strong>3 Literature Review</strong></td>
<td>3</td>
</tr>
<tr>
<td>3.1 Confirmatory factor analysis</td>
<td>3</td>
</tr>
<tr>
<td>3.2 Measurement scales</td>
<td>4</td>
</tr>
<tr>
<td>3.3 Dealing with ordinal data in CFA</td>
<td>5</td>
</tr>
<tr>
<td>3.3.1 Correlation for ordinal data</td>
<td>5</td>
</tr>
<tr>
<td>3.3.2 CFA model for ordinal data</td>
<td>7</td>
</tr>
<tr>
<td>3.4 Estimation method</td>
<td>8</td>
</tr>
<tr>
<td>3.4.1 ML and RML</td>
<td>8</td>
</tr>
<tr>
<td>3.4.2 ULS and RULS</td>
<td>10</td>
</tr>
<tr>
<td>3.5 Comparison between two types of data</td>
<td>12</td>
</tr>
<tr>
<td><strong>4 Simulation Study</strong></td>
<td>13</td>
</tr>
<tr>
<td>4.1 Model</td>
<td>13</td>
</tr>
<tr>
<td>4.2 Conditions and Data Generation</td>
<td>15</td>
</tr>
<tr>
<td>4.3 Simulation Results</td>
<td>15</td>
</tr>
<tr>
<td><strong>5 Empirical Study</strong></td>
<td>18</td>
</tr>
<tr>
<td>5.1 Background of questionnaire</td>
<td>18</td>
</tr>
<tr>
<td>5.2 Missing values</td>
<td>20</td>
</tr>
<tr>
<td>5.3 Descriptive analyses</td>
<td>21</td>
</tr>
<tr>
<td>5.4 CFA results</td>
<td>24</td>
</tr>
<tr>
<td>5.4.1 Version 1—four categories in ordinal data</td>
<td>24</td>
</tr>
<tr>
<td>5.4.2 Version 2—seven categories in ordinal data</td>
<td>25</td>
</tr>
<tr>
<td>5.4.3 Summary of the two versions</td>
<td>26</td>
</tr>
<tr>
<td><strong>6 Conclusion</strong></td>
<td>29</td>
</tr>
<tr>
<td>References</td>
<td>30</td>
</tr>
<tr>
<td>Appendix</td>
<td>34</td>
</tr>
</tbody>
</table>
1 Introduction

Questionnaires are widely used by researchers in social science, psychology and many other fields. These questionnaires can help us understand hypothetical or theoretical constructs which are not directly observable or measurable, such as prejudice, radicalism, trust, self-esteem, discrimination, motivation and ability. Mechanism of investigating such constructs is to have some indicators which can represent these constructs. Indicators, in their turn, are observed through questions in questionnaire. A good questionnaire should be well designed with a deep understanding of what the researchers want to investigate.

Likert-type scale is a popular rating scale used in questionnaires. This scale offers finite choices for interviewees to choose from, where there is a natural order between the choices. For example, if a company wants to investigate its customers’ satisfaction about the service it offers, it can ask a question like ‘How satisfied are you with our service?’ and offer five response categories, including ‘Strongly unsatisfied’, ‘Unsatisfied’, ‘No opinion’, ‘Satisfied’ and ‘Very satisfied’. Results obtained from Likert-type scales are ordinal.

When data gathered from Likert-type scale are to be analyzed, researchers should pay special attention to them for what we regularly deal with, in daily life, is continuous data. Many theories and methods are designed for continuous data, which are not directly applicable for ordinal data (see, e.g., Muthén and Kaplan, 1985). If proper ways are not found to deal with ordinal data, efforts paid to do the survey would be wasted. So how to deal with ordinal data needs to be examined.

Meanwhile, the popularity of ordinal data in questionnaires does not mean that continuous data can not appear in questionnaire. In research where aspects like an interviewee’s height and weight are recorded, continuous data dominate. Continuous data are very familiar to statisticians, and can be handled quite well. Many statistical methods are designed for continuous data, for example, confirmatory factor analysis (CFA). This method is not directly applicable for ordinal data, which must be processed so that CFA can be used as well.

In this thesis, two specially designed questionnaires are gathered in a study concerning students’ stress when learning statistics. These questionnaires have five exactly same questions and each question is answered by visual analogue scale and Likert-type scale simultaneously. The only difference exists in the number of response categories in Likert-type scale. One questionnaire has four response categories and the other has seven response categories. For convenience, questionnaire with four response categories is called ‘Version 1’ questionnaire.
and the one with seven response categories is called 'Version 2' in the rest of the thesis. Data gathered from the empirical study are used to fit a two-factor CFA model. Continuous data in both versions of questionnaires violate the multinormality assumption. Ways to solve violation of multinormality in continuous data and use ordinal data in CFA is discussed. Model results on criteria of goodness-of-fit indices and parameter estimates are presented in this thesis.

The thesis will lay out in the following structure: In research question part, what this thesis wants to discuss are listed. In literature review part, ways to do CFA with both continuous and ordinal data are presented. Polychoric correlation and four estimation methods are explained in detail. Past comparisons between the two types of data are discussed as well. In the simulation study part, impacts of categorization and estimation methods on CFA results are discussed. In the empirical study part, background about the questionnaires we use is stated. Also, missing value issues and a description of our data are presented. Model results of a two-factor CFA model based on two versions of questionnaires are listed in two tables. Differences in model results are stated. At last, a short conclusion about our findings based on both simulation and empirical study completes this thesis.

2 Research Question

In this thesis, analyses are based on the data sets gathered from two versions of questionnaires. These questionnaires contain both continuous and ordinal data for each of the five questions. A pre-determined two-factor CFA model is used to fit our data. Some ways to deal with ordinal data and non-multinormal continuous data when doing CFA are discussed and four estimation methods: maximum likelihood (ML), robust maximum likelihood (RML), unweighted least square (ULS) and robust unweighted least square (RULS) are chosen. Model fit and parameter estimates are checked in different situations. The following questions are addressed in this thesis:

1. What is the impact on the number of categories on parameter estimates, standard errors and model fit?
2. Does the estimation method have an additional impact on these aspects?
3. Does the empirical study in some way support the findings of simulation study?
3 Literature Review

3.1 Confirmatory factor analysis

Confirmatory factor analysis (CFA) is a multivariate statistical method which is frequently used in social science and behavioral research. In these fields, variables of interest may not be directly measurable. For example, there is no natural measurement unit for the construct of satisfaction. In this case, researchers can use some indicators which are believed to be related with the construct (latent variable), along with measurement errors. Researchers have some pre-determined hypotheses about relationships between indicators and latent variables and how many latent variables there should be in the model.

The basic logic of factor analysis is first brought up by Spearman (1904), many versions of factor analysis were then developed since then. CFA has the following basic idea: given a set of observed variables $x_1, x_2, ..., x_p$, there is assumed to exist some underlying factors $\xi_1, \xi_2, ..., \xi_m$, where $m < p$, which account for the inner correlation of observed variables. Written as an equation, it will become:

$$x = \Lambda \xi + \delta,$$

where $x$ is a vector of $p \times 1$ observed variables, $\Lambda_{p \times m}$ is the matrix for factor loadings. $\xi$ is a vector of $m \times 1$ factors (latent variables), $\delta$ is a vector of $p \times 1$ measurement errors corresponding to each of the $p$ observed variables. Elements in $\delta$ are assumed to be uncorrelated.

Based on the basic settings, relationships between the covariance matrices can be formulated. Let $\Phi_{m \times m}$ and $\Theta_{p \times p}$ be the covariance matrices of $\xi$ and $\delta$. It is not surprising that $\Theta_{p \times p}$ is a diagonal matrix due to the uncorrelation we have assumed in elements of $\delta$. The covariance matrix of $x$ is

$$\Sigma(\theta) = \Lambda \Phi \Lambda' + \Theta,$$

where $\theta$ is a set of free parameters in $\Lambda$ and $\Phi$. Basic thoughts of estimation of CFA is that $\Sigma$ (the covariance matrix of the observed variables) is a function of $\theta$. Sample covariance matrix can be reproduced through a correct model. This leads to the main idea of estimating CFA model: minimizing the differences between the sample covariance matrix and the model implied covariance matrix.

Jöreskog (1969) used maximum likelihood (ML) as an estimation method to illustrate CFA model, based on the assumption of multinormality. Since then, it has become popular among
researchers to use ML to do CFA. However, in practice the multinormal distribution is often violated (see e.g., Hu, Bentler, and Kano, 1992). Studies have shown that ML is robust to slight departure from normality (e.g., Chou and Bentler, 1995; West, Finch, and Curran, 1995). Satorra and Bentler (1994), West et al. (1995), Brown (2012) and Rhemtulla, Brosseau-Liard, and Savalei (2012) also point out that severe non-normality will lead to biased standard errors of parameters and cause problem in judging significance of parameter estimates. Solutions of non-multinormality in continuous data are discussed in later section called 'Estimation method’, along with solutions for ordinal data.

3.2 Measurement scales

So far, our discussion has been constrained to continuous data, which is not the type of data that most questionnaires in social science and behavioral research obtain. What they obtain are ordinal data (DiStefano, 2002; Yang-Wallentin, Jöreskog, and Luo, 2010; Rhemtulla et al., 2012). There are many different definitions of continuous and ordinal data, all of which are quite similar. Brief definitions of the two types of data are given as follows. Continuous data are the kind of data which can take on infinite values within some range. For instance, continuous data are used when people measure variables such as income or the size of a house. Ordinal data are the kind of data that intrinsic order exists in them and people can set arbitrary values to them. The values, differences or ratios between ordinal data cannot be interpreted. For example, when talking about education level, it is obvious that there is an increasing ordering in primary school, high school, university and graduate. Another example is Likert-type scale used to present people’s attitude towards something, from 'Very unsatisfied’ to 'Very satisfied’.

Continuous data can be infinitely fine sub-divided, which makes differences between items comparable. For example, the difference between a student with height 184 cm and one with 176 cm is the same with that of a student with height 178 cm and one with 170 cm. Mean and variance are meaningful for continuous data, which is not the case with ordinal data, researchers should think carefully before handling ordinal data. It is not certain whether the spaces between categories are even or not, and thus differences between items are not comparable. The coming instance can give an intuitive feeling about difference between the two types of data. In one data set consisting of Likert-type scale data, the difference between category 2 and category 1 should not be considered the same as that of category 3 and category 2. The unit 'one’ in ordinal data measures unequally corresponding to different number, which means that it is nonsensical
3.3 Dealing with ordinal data in CFA

3.3.1 Correlation for ordinal data

Due to the meaninglessness of mean and covariance for ordinal data, Pearson product-moment correlation is not appropriate to measure the associations between ordinal variables. As the Pearson product-moment correlation is not meaningful (see e.g., Choi, Peters, and Mueller, 2010), thus CFA model in Equation 2 is not suitable for ordinal data.

However, there exist some solutions which can reconcile ordinal data to CFA. For ordinal data, correlation except for Pearson product-moment correlation can be calculated. For example, Kendall’s $\tau - b$ correlations, Spearman’s rho and polychoric correlation can be calculated for ordinal data as well. Babakus, Ferguson Jr, and Jöreskog (1987) show that polychoric correlation leads to the best results on the basis of squared error and factor loading bias. Studies by Jöreskog and Sörbom (1996) show that polychoric correlation is the best correlation among six correlations calculated for ordinal data when the underlying bivariate normality holds. So in this thesis, polychoric correlation is calculated.

To calculate polychoric correlation, an assumption should be made to link ordinal data with 'underlying' continuous data. In this case, relationships between ordinal data can be measured with the help of related underlying continuous data. Múthen (1983) shows the way to link the two types of data. Observed ordinal variables are related to unobserved continuously distributed variables (latent variables). Suppose that $x$ is the observed ordinal variable with $m$ categories and $x^*$ is the underlying continuous variable. A monotonic relation is presented as:

$$x = c \iff \tau_{c-1} < x^* < \tau_c, \quad c = 1, 2, ..., m,$$

where

$$\tau_0 = -\infty, \tau_1 < \tau_2 < ... < \tau_{m-1}, \tau_m = +\infty,$$

$\tau_c$ are thresholds categorizing continuous data into ordinal data.

Based on this relation, polychoric correlation can be found to state relationship between ordinal variables. Polychoric correlation is first brought up by Pearson (1900). He initiates his explanation by a contingency table. Based on the assumption that there exist underlying joint bivariate normally distributed continuous variables, the marginal distribution of the contingency table is therefore normal. Ekström (2011) explained why Pearson would make this
assumption. Two main reasons are: normal distribution was quite popular at that time and has nice properties, and the area Pearson focuses usually has normal distributed variables. However, the violation of normal distribution will bring sever influence on polychoric correlation estimation, not only in the absolute value, but also in sign. The normal distribution assumption of polychoric correlation is rarely checked and is therefore criticized by people (e.g., Yule, 1912).

In LISREL, there is a very easy way to check the assumption. When calculating polychoric correlation, we can choose 'Print tests of underlying bivariate normality'. The null hypothesis of the test is that underlying bivariate normality holds. We conduct the underlying bivariate normality test for ordinal data collected from two versions of questionnaires. Test results show that $p$ values of all pairs of variables are larger than 0.95 (the standard criterion), thus the assumption hold for ordinal data in both versions of questionnaires. Therefore, there should be no worries about calculation of polychoric correlation. The detailed procedure of calculating polychoric correlation is explained in the following paragraphs.

Olsson (1979) gives detailed explanation of how to calculate polychoric correlation for ordinal data. Let $x_1$ and $x_2$ be two ordinal variables with $m_1$ and $m_2$ categories, respectively. Their marginal distributions could be summarized in the following contingency table:

<table>
<thead>
<tr>
<th>Frequency</th>
<th>$x_2$</th>
<th>Category 1</th>
<th>Category 2</th>
<th>⋮</th>
<th>Category $m_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td></td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
<td>⋮</td>
<td>$n_{1m_2}$</td>
</tr>
<tr>
<td>Category 1</td>
<td></td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
<td>⋮</td>
<td>$n_{2m_2}$</td>
</tr>
<tr>
<td>⋮</td>
<td></td>
<td>⋮</td>
<td>⋮</td>
<td></td>
<td>⋮</td>
</tr>
<tr>
<td>Category $m_1$</td>
<td></td>
<td>$n_{m_11}$</td>
<td>$n_{m_12}$</td>
<td>⋮</td>
<td>$n_{m_1m_2}$</td>
</tr>
</tbody>
</table>

where $n_{ij}$ denotes number of observed cases in category $i$ on variable $x_1$ and in category $j$ on variable $x_2$.

Based on theory of Múthen (1983), two underlying variables $x_1^*$ and $x_2^*$ exist for $x_1$ and $x_2$. Thresholds which categorize underlying variables to ordinal variables are denoted as $(\tau_1^1, \tau_2^1, \cdots, \tau_{m_1-1}^1)$ and $(\tau_1^2, \tau_2^2, \cdots, \tau_{m_2-1}^2)$, respectively. Olsson further assumes that $x_1^*$ and $x_2^*$ are bivariate normally distributed. Then likelihood function of the sample presented in the above contingency table can be written in the following way:

$$L = C \cdot \prod_{i=1}^{m_1} \prod_{j=1}^{m_2} \pi_{ij}^{n_{ij}},$$

(3)
where $C$ is a constant and $\pi_{ij}$ is the probability of an observation falling in category $i$ on variable $x_1$ and in category $j$ on variable $x_2$. Based on bivariate normal assumption expression, an formula for $\pi_{ij}$ is available:

$$\pi_{ij} = \Phi_2(\tau_{1i}, \tau_{2j}) - \Phi_2(\tau_{1i-1}, \tau_{2j}) + \Phi_2(\tau_{1i}, \tau_{2j-1}) - \Phi_2(\tau_{1i-1}, \tau_{2j-1}),$$  \hspace{1cm} (4)

where $\Phi_2$ denotes the cumulative distribution function for bivariate normal distributed variables with correlation $\rho$. Detailed expression of $\Phi_2$ are shown in Equation 5.

$$\Phi_2(\tau_{1i}, \tau_{2j}) = \int_{-\infty}^{\tau_{1i}} \int_{-\infty}^{\tau_{2j}} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1x_2 + x_2^2)} dx_1dx_2.$$  \hspace{1cm} (5)

Therefore, the likelihood function $'L'$ is a function of $\tau_{1i}$, $\tau_{2j}$ and $\rho$. Maximizing $L$ can be done in two ways: two-step ML and one-step ML. In two-step ML method $\tau_{1i}$ and $\tau_{2j}$ are first estimated based on the univariate marginal distributions, then polychoric correlations are calculated with respect to maximize $L$, given the thresholds estimated in the first step. In one-step ML method, all parameters are estimated simultaneously. Jöreskog (2005) points out that there is no need to use the one-step ML method for it produces almost the same results as the two-step method and would yield different thresholds for the same variable when paired with different variables. The default method to calculate polychoric correlation by PRELIS is the two-step ML method.

Indeed, polychoric correlation matrix is part of the limited information approaches for dealing with ordinal data. Limited information approaches only use univariate and bivariate frequencies to estimate parameters. There also exists full information approaches which use the whole response pattern to get information about parameter estimates (see Rhemtulla et al. (2012)). Though full information methods have advantage of using more information, practical studies show that this advantage is slight and differences between full and limited information methods are negligible (Maydeu-Olivares and Joe, 2005; Forero and Maydeu-Olivares, 2009). Besides, full information methods are not chosen for they require heavy calculations and are not applicable in all SEM programs (Rhemtulla et al., 2012). Therefore calculating polychoric correlation for ordinal data is used in this thesis.

### 3.3.2 CFA model for ordinal data

Based on the polychoric correlation, the CFA model in Equation 2 can be rewritten. The underlying continuous variable $x^*$ is now modeled instead of the observed ordinal variable. Now
the equation of the CFA becomes:

\[ x^* = \Lambda \xi + \delta. \] (6)

The procedure of estimating CFA for ordinal data is to rebuild the correlation matrix, not the covariance matrix, through the model now. Model implied correlation matrix is in analogous form as that in Equation 2. As the underlying variables now has unit standard error, the special form of \( \Theta \) is stated as:

\[ \Theta = I - \text{diag}(\Lambda \Phi \Lambda'). \] (7)

Substituting Equation 7 in Equation 2 leads to the final version of the model implied correlation matrix of the underlying continuous data:

\[ \Sigma^*(\theta) = \Lambda \Phi \Lambda' + I - \text{diag}(\Lambda \Phi \Lambda'). \] (8)

The aim of estimation process for ordinal data is to find a model in which parameters can make \( \Sigma^*(\theta) \) as close as possible to the sample-implied polychoric correlation matrix.

### 3.4 Estimation method

In this section, some estimation methods for non-multinormal continuous data and ordinal data are reviewed. Four estimation methods compared in this thesis are: ML (maximum likelihood), RML (robust maximum likelihood), ULS (unweighted least square) and RULS (robust unweighted least square). Fit functions of these four methods for both continuous and ordinal data are presented below.

#### 3.4.1 ML and RML

ML method is chosen for its popularity in estimating CFA models. Schermelleh-Engel et al. (2003) and Mîndrilă (2010) make good summaries of ML. They point out that test of overall model fit for overidentified model is possible with ML. Besides, ML is a full information estimator, which allows for statistical inference like significance test and goodness-of-fit evaluation (Brown, 2012, p. 22). One drawback of ML is the requirement of multinormal distribution assumption, which is often not fulfilled in applied data. If the assumption is violated, then model results may not be reliable. Bias appears in standard errors and chi-square. Goodness-of-fit indices GFI and RMR will also be affected, by underestimation in GFI and overestimation in RMR.
Two solutions are available for non-multinormal continuous data: robust maximum likelihood (RML) (Satorra and Bentler, 1994) and weighted least square (WLS) (Browne, 1984). WLS is not recommended for it requires large sample sizes to calculate its weight matrix (Jöreskog and Sörbom, 1996; Brown, 2012). RML is a method based on the use of an asymptotic covariance matrix. It generates less biased standard errors and performs well facing different sample sizes and degrees of non-normality. RML is easy to carry out in LISREL, just by adding the asymptotic covariance matrix in the syntax.

Continuous Data

Jöreskog, Sörbom, Du Toit and Du Toit (2001, p.198), Yang-Wallentin et al. (2010) and Schermelleh-Engel et al. (2003) give a basic ML fit function for continuous data:

$$F_{CML} = \log|\Sigma(\theta)| + tr(SS^{-1}(\theta)) - \log|S| - p,$$

(9)

where $\Sigma(\theta)$ is defined in Equation 2, $S$ is the sample covariance matrix and $p$ is the number of observed variables. Jöreskog et al. (2001, p.198) and Yang-Wallentin et al. (2010) point out that this fit function can be rewritten in a way which is asymptotically equivalent to those of least square methods:

$$F_{CML} = (s - \sigma(\theta))'D(\hat{\Sigma}^{-1}(\theta) \otimes \hat{\Sigma}^{-1}(\theta))D(s - \sigma(\theta)),$$

(10)

where

$$s' = (s_{11}, s_{21}, s_{32}, \cdots, s_{pp}),$$

is a vector of the lower half of sample covariance matrix $S$, with diagonal elements included. $\sigma(\theta)$ is a vector which consists of the corresponding elements of the model implied covariance matrix $\Sigma(\theta)$. $D$ is a duplication matrix (Magnus and Neudecker, 1999) and the symbol ’$\otimes$’ stands for the Kronecket product.

RML has the same estimation function as ML, but it has modified standard errors and chi-square. The robust covariance matrix of $\hat{\theta}$ is in the following form (see Browne, 1984; Yang-Wallentin et al., 2010; Rhemtulla et al., 2012):

$$\frac{1}{N} \times [(\hat{\Delta}'V\hat{\Delta})^{-1}\hat{\Delta}'VWV\hat{\Delta}(\hat{\Delta}'V\hat{\Delta})^{-1}],$$

(11)

where $N$ is the sample size, $\hat{\Delta} = \partial\sigma(\theta)/\partial\theta'|_{\theta=\hat{\theta}}$ is the matrix of model derivatives evaluated at the parameter estimates. $V = D'(\hat{\Sigma}^{-1}(\theta) \otimes \hat{\Sigma}^{-1}(\theta))D$ (as defined in Equation 10). $W$ is the asymptotic covariance matrix of elements in the sample covariance matrix.
Corrected standard errors of parameter estimates are the square root of the diagonal elements in Equation 11. Based on the theory of Satorra and Bentler (1994), scaled $\chi^2$ looks like:

$$
(N - 1)F_{CML} \cdot \frac{df}{tr[V - V\hat{\Delta}(\hat{\Delta}'V\hat{\Delta})^{-1}\hat{\Delta}'V]}
$$

where $F_{CML}$ is defined in Equation 9 and $df$ is the degree of freedom.

**Ordinal Data**

For ordinal data, modifications should be done to the two forms of fit functions, just like for the expression of $\Sigma^*(\theta)$. All that is needed to be done is to replace covariance-related expressions to polychoric-related expressions in Equation 9 and 10. So two forms of fit function for ordinal data will be:

$$
F_{OML} = \log|\Sigma^*(\theta)| + tr(R\Sigma^{-1}(\theta)) - \log|R| - p,
$$

where $\Sigma^*(\theta)$ is defined in Equation 8, $R$ is the polychoric correlation matrix of the sample and $p$ is the number of observed variables. The second form follows as:

$$
F_{OML} = (r - \rho(\theta))'D'(\hat{\Sigma}^{-1}(\theta) \otimes \hat{\Sigma}^{-1}(\theta))D(r - \rho(\theta)),
$$

where

$$
r' = (1, r_{21}, 1, r_{31}, r_{32}, 1, \cdots, r_{p,p-1}, 1),
$$

is a vector of the lower half of sample polychoric correlation matrix $R$, with diagonal elements ‘1s’ included. $\rho(\theta)$ contains the corresponding elements of the model implied correlation matrix $\Sigma^*(\theta)$. $D$ and ‘$\otimes$’ is the same as defined above.

RML for ordinal data shares the same characteristic as RML for continuous data. The robust covariance matrix of $\hat{\theta}$ is in the exact same form in Equation 11. Only $\hat{\Delta}$, $V$, $W$ have a few changes. Again, covariance-related components are replaced by correlation-related components. $\hat{\Delta} = \partial\rho(\theta)/\partial\theta'|_{\theta=\hat{\theta}}$ is the matrix of model derivatives evaluated at the parameters estimates. $V = D'(\hat{\Sigma}^{-1}(\theta) \otimes \hat{\Sigma}^{-1}(\theta))D$, which is defined in Equation 14. $W$ is the asymptotic covariance matrix of elements in polychoric correlation matrix. Expression of scaled $\chi^2$ for ordinal data is just to substitute $F_{CML}$ with $F_{OML}$ in Equation 12.

### 3.4.2 ULS and RULS

Another two methods used in this paper are ULS and RULS. They are chosen to account for ordinal data. Though first WLS (weighted least square) is suggested to deal with ordinal data
(e.g., Jöreskog, 1990; Browne, 1984), it has been criticized for its poor performance when sample size is small and the model is complicated (see Muthén and Kaplan, 1985; Chou and Bentler, 1995; West et al., 1995; Flora and Curran, 2004; Forero et al., 2009; Brown, 2012, p. 388). Methods other than WLS have been brought up.

Two popular methods are unweighted least square (ULS) and diagonally weighted least square (DWLS) (Yang-Wallentin et al., 2010; Brown, 2012). They are quite similar as WLS, just with a little difference in the weight matrix in the fit function. DWLS use a weight matrix which only contains the diagonal elements of the asymptotic covariance matrix, while ULS use the identity matrix as its weight matrix. Forero et al. (2009) have shown that ULS gives more precise estimation (less bias and smaller standard errors) than DWLS. Research of Rigdon and Ferguson Jr (1991) also shows that ULS outperforms DWLS. Besides, Babakus et al. (1987) also recommend that ULS should be used to solve problems when polychoric correlation matrix is used because ULS overcomes the shortage of nonpositive definite weight matrix in WLS and reduces inaccurate solutions and tendency of nonconvergence.

**Continuous Data**

The above discussed three estimation methods have similar fit functions. They share the following fit function:

\[ F = (s - \sigma(\theta))'V(s - \sigma(\theta)). \] (15)

Different choices of \( V \), the weight matrix, leads to different methods. For ULS, \( V \) is the identity matrix: \( I \) with order \( \frac{p(p+1)}{2} \times \frac{p(p+1)}{2} \). Therefore, the fit function of ULS is:

\[ F_{\text{CULS}} = (s - \sigma(\theta))'(s - \sigma(\theta)), \] (16)

\( s \) and \( \sigma(\theta) \) are the same as in Equation 10, the lower part of sample covariance matrix and model-implied covariance matrix, respectively.

RULS, identical with RML, needs alteration on parameters estimates’ standard errors and \( \chi^2 \) value. With \( V = I \), Equation 11 simplifies to:

\[ \frac{1}{N} \times [(\hat{\Delta}'\hat{\Delta})^{-1}\hat{\Delta}'W\hat{\Delta}(\hat{\Delta}'\hat{\Delta})^{-1}]. \] (17)

\( \hat{\Delta} \) and \( W \) have same definition in Equation 11. Besides, scaled \( \chi^2 \) is more brief for RULS as well:

\[ (N - 1)F_{\text{CULS}} \cdot \frac{df}{tr[I - \hat{\Delta}(\hat{\Delta}'\hat{\Delta})^{-1}\hat{\Delta}']}, \] (18)

where \( F_{\text{CULS}} \) is defined in Equation 16.
Ordinal Data

One important thing must first be pointed out when using ULS for ordinal data. Except for substituting the covariance-related elements in Equation 16, the order of vectors has changed as well. Now the diagonal elements of the sample polychoric correlation $R$, '1s', are not included in the fit function. In order to prevent confusion, a '*' is now added, thus the fit function of ordinal data will look like:

$$F_{OULS} = (r^* - \rho^*(\theta))' (r^* - \rho^*(\theta)),$$

where

$$r^* = (r_{21}, r_{31}, r_{32}, \ldots, r_{p,p-1})',$$

is a vector of the elements below the diagonal of sample polychoric correlation matrix $R$. $\rho^*(\theta)$ is the corresponding elements in model implied correlation matrix $\Sigma^*(\theta)$.

To obtain robust covariance matrix of $\hat{\theta}$, corrections analogous to those done in the RML case are made. $\hat{\Delta}$ in Equation 17 and 18 equals $\partial \rho^*(\theta)/\partial \theta'|_{\theta=\theta}$. $W$ is the asymptotic covariance of elements in the polychoric correlation matrix adapted to $r^*$.

3.5 Comparison between two types of data

Comparisons of CFA between continuous and ordinal data have been done mainly on simulation data. Babakus et al. (1987) first generate 24,000 independent random samples of continuous data, then categorize them into five categories. Four correlation measures are calculated for ordinal data, and differences regarding nonconvergence rates, bias and standard errors of parameter estimates, goodness-of-fit indices, $\chi^2$ and its $p$ values are listed. Part of their results show that for ordinal data, polychoric correlation results in the most accurate standard errors of factor loadings, which is the closest to those of continuous cases. DiStefano (2002) also cut normally distributed continuous data to ordinal data with five categories. Two methods: WLS with polychoric correlation matrix and ML with Pearson product-moment matrix are used for ordinal data. For continuous data, ML with Pearson product-moment matrix is used. Then based on bias and the standard errors of parameter estimates, along with five goodness-of-fit indices, they conclude that ordinal data cause bias in parameter estimates and their standard errors, while continuous data do not have this problem. Besides, though all goodness-of-fit indices show good model fits, continuous data have slightly better values on the chosen fit indices. Mîndrilă (2010) uses five artificially generated data sets: one consists of multinormally
distributed continuous data, the others consist of four kinds of ordinal data (two different numbers of categories: 3 and 7 and two distributions: multinormal and non-multinormal). ML and DWLS methods are used to compare parameter estimates and eight fit indices. Results show that ML method leads to more inaccurate estimates when using ordinal data. DWLS provides more accurate parameter estimates and more robust fit indices when dealing with non-normality and different data type. Rhemtulla et al. (2012) use simulated data for 480 conditions to find how many categories are enough to be treated as continuous data. Part of their research is to use RML with different ordinal data, ignoring the categorical nature of the data. Results show that when the number of categories reaches five, then RML with ordinal data produces acceptable results with regard to convergence rates, improper solutions, qualities of parameter estimates and so on.

In summary, methods are available for CFA with violation of multinormality assumption and measurement scale. Comparisons based on simulation data have been made in aspects like: different estimation methods, types of data and non-normality. Bias and standard errors of parameter estimates, nonconvergence and goodness-of-fit indices are popular criteria to be judged.

4 Simulation Study

In this section, simulated data will be used to access the impact of the number of response categories on CFA model. Besides, whether four estimation methods introduced above (ML, RML, ULS and RULS) have impact on CFA, is also checked.

4.1 Model

We use a CFA model which has the same structure as the model for the empirical study. The structure of the model is shown in Figure 1.
When writing the model in matrix form, it becomes:

\[
\begin{pmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4 \\
  x_5 \\
\end{pmatrix} =
\begin{pmatrix}
  \lambda_1 & 0 \\
  \lambda_2 & 0 \\
  \lambda_3 & 0 \\
  0 & \lambda_4 \\
  0 & \lambda_5 \\
\end{pmatrix}
\begin{pmatrix}
  \xi_1 \\
  \xi_2 \\
\end{pmatrix} +
\begin{pmatrix}
  \delta_1 \\
  \delta_2 \\
  \delta_3 \\
  \delta_4 \\
  \delta_5 \\
\end{pmatrix}.
\]

For convenience, the variances of the latent variables are is usually set to one. So the covariance matrix of the latent variables is in the following form:

\[
\Phi =
\begin{pmatrix}
  1 & \\
  \phi_{21} & 1 \\
\end{pmatrix}
\]

and the covariance matrix of the measurement errors \((\delta)\) is represented by \(\Theta=\text{diag}(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)\), where \(\theta\) is the variance of the measurement errors.

The parameter values used to generate data for the model are arbitrarily chosen as:

\[
(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)' = (0.9, 0.8, 0.7, 0.8, 0.7)'
\]

\[
(\phi_{11}, \phi_{21}, \phi_{22})' = (1, 0.3, 1)'
\]

Both \(\Lambda\) and \(\Phi\) are the same for all conditions.
4.2 Conditions and Data Generation

The idea here is to first generate five multivariate normal variables: $x_1^*, x_2^*, x_3^*, x_4^*, x_5^*$, then categorize them into ordinal data. All of them are first categorized into four categories and then into seven categories.

Categorizations are symmetric because previous studies (Babakus et al., 1987; DiStefano, 2002) show that asymmetric category thresholds lead to worst performance for most methods. In the four category condition, thresholds are -1.645, 0 and 1.645, resulting in 5%, 45%, 45% and 5% of normally distributed data in each category. In seven categories case, thresholds are -1.645,-1.036,-0.524, 0.524,1.036 and 1.645, resulting in 5%, 10%, 15%, 40%, 15%, 10% and 5% of normally distributed data in each category.

Sample size is 200 for each data set and the number of replicates is 1000. Taken together, there are twelve conditions for the CFA model: four estimation methods (ML, RML, ULS and RULS) and three different kinds of data (continuous, 4-category and 7-category).

Data generation and analysis follows the following procedures.

Step 1 Generate $x_1^*, x_2^*, x_3^*, x_4^*, x_5^*$ according to the CFA model we set. Draw a random sample with sample size=200.

Step 2 First save the sample we get in Step 1 as a sample which contains continuous data. Second, use the established thresholds to categorize $x^*$ first into four categories and then seven categories. In total, we get three samples.

Step 3 Use the samples from Step 2 to do CFA analysis. Four estimation methods are applied to the three samples. Four each method, record parameter estimates, standard errors and $\chi^2$.

Step 4 Repeat the above steps 1000 times and record all results.

4.3 Simulation Results

The main interest is to access the overall properties of the model results from different kinds of data and estimation methods. Accessing the individual parameter estimates are not the main interest to us. Three aspects to do the assessment are: parameter estimates, standard errors and model fit. In order to get the overall properties, we need to calculate some criteria based on results from the 1000 replicates.
Let $\hat{\lambda}_{ij}$ be the estimated value of the $j$th parameter in the $i$th replicate, where $j = 1, 2, ..., n$ and $i = 1, 2, ..., R$ ($n$ is the number of parameter estimates and $R$ is the number of replicates). $\lambda_{ij}$ is the true value of $\hat{\lambda}_{ij}$, which is settled in Equation 20. In this thesis, we focus only on factor loadings. The following criteria are used for evaluation:

- **Average Relative Bias (ARB)**

\[
ARB = \frac{1}{R} \sum_{i=1}^{R} \frac{1}{n} \sum_{j=1}^{n} \left( \frac{\hat{\lambda}_{ij} - \lambda_{ij}}{\lambda_{ij}} \right).
\] (22)

- **Average Root Mean Square Error (AMSE)**

\[
AMSE = \frac{1}{R} \sum_{i=1}^{R} \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left( \frac{\hat{\lambda}_{ij} - \lambda_{ij}}{\lambda_{ij}} \right)^2}.
\] (23)

- **ARB and AMSE for Standard Errors (ARBSE & AMSESE)**

It is not only the values of parameter estimates we need to pay attention, but also the precision of parameter estimates. The precision is represented by standard errors. To compare the properties of standard errors, Equation 22 and 23 can be applied to estimated standard errors of parameters. The empirical standard errors are obtained from the standard deviation of parameter estimates in the 1000 replicates. ARBSE and AMSESE show how precise the estimated standard errors are.

- **Average $\chi^2$ (AC)**

Every time of 1000 replicates produces a corresponding value of $\chi^2$. Let $\chi^2_i$ be the $\chi^2$ of the $i$th replication. The average $\chi^2$ (AC) is then:

\[
AC = \frac{1}{R} \sum_{i=1}^{R} \chi^2_i
\] (24)

- **Standard Error of $\chi^2$ ($SE_{\chi^2}$)**

The standard error of the $\chi^2$ ($SE_{\chi^2}$) is calculated as:

\[
SE_{\chi^2} = \sqrt{\frac{1}{R - 1} \sum_{i=1}^{R} (\chi^2_i - \overline{\chi^2})^2}
\] (25)

where $\overline{\chi^2}$ is the mean of $\chi^2$. 

16
Table 1: Summary of Model Results from Simulation

<table>
<thead>
<tr>
<th></th>
<th>continuous situation</th>
<th>4-category situation</th>
<th>7-category situation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ML RML ULS RULS</td>
<td>ML RML ULS RULS</td>
<td>ML RML ULS RULS</td>
</tr>
<tr>
<td>ARB</td>
<td>0.0218 0.0218 0.0240 0.0240</td>
<td>0.0550 0.0550 0.0587 0.0587</td>
<td>0.0268 0.0268 0.0297 0.0297</td>
</tr>
<tr>
<td>AMSE</td>
<td>0.2066 0.2066 0.2100 0.2100</td>
<td>0.2985 0.2985 0.2958 0.2958</td>
<td>0.2214 0.2214 0.2249 0.2249</td>
</tr>
<tr>
<td>ARBSE</td>
<td>-0.0452 -0.0561 0.3359 -0.0336</td>
<td>-0.2590 -0.1037 0.0461 -0.0784</td>
<td>-0.0017 -0.0576 0.3493 -0.0689</td>
</tr>
<tr>
<td>AMSESE</td>
<td>0.2823 0.2914 0.5744 0.2949</td>
<td>0.4684 0.4468 0.5044 0.4414</td>
<td>0.3508 0.3674 0.6665 0.3500</td>
</tr>
<tr>
<td>$SE_{\chi^2}$</td>
<td>2.9174 2.9658 2.9483 2.9775</td>
<td>7.0698 3.0646 7.1135 3.0511</td>
<td>3.8395 2.9572 3.8882 2.9625</td>
</tr>
<tr>
<td>$df^*$</td>
<td>4 4 4 4 4 4 4 4 4 4 4 4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$df^*$: stands for degree of freedom.

The above five criteria are calculated for the twelve conditions. Results are listed in Table 1. From the table we can see that continuous data have the smallest biases for parameter estimates and standard errors in most cases in almost all conditions. 7-category performs better, in terms of bias and precision, than 4-category in most of the four estimation methods based on all the criteria, except for the large ARBSE and AMSESE in ULS. Additionally, AC and $SE_{\chi^2}$ of continuous data are also much smaller than those of 4-category and 7-category data, which means continuous data have better and steadier model fit. The differences between continuous data and 4-category data are more apparent than those between continuous data and 7-category data. The smaller values of criteria, the less biased, more precise and better model fit of the model in the specified situation. Therefore, we can conclude that the order of data, which is 'less biased, more precise and better model fit', is continuous data, 7-category data and 4-category data. The result is reasonable for categorization will lead to information loss in data, so the more categories we keep, the more likely that the generated data keep more information from original data. Thus we can see smaller values for criteria of 7-category data than 4-category data.

When looking at the affects brought by estimation methods, it is not hard to see that there is no difference between robust and nonrobust methods in the estimated values of parameters. There is only a very slight difference between ML and ULS in terms of estimated values of parameters in three kinds of data. However, we see obvious differences when we check the standard errors. In the continuous situation, standard errors are more biased and less precise in ULS than with the other three estimation methods. In the 4-category situation, ULS and
RULS have less biased standard errors than ML and RML. However, all the four estimation methods show similar fluctuation in standard errors. In the 7-category situation, the results are similar to the continuous situation: ULS leads to more bias and less precision results in standard errors. Last, AC and $SE_{\chi^2}$ are checked. In continuous case, the differences between the four estimation methods are negligible, which means that the four estimation methods have similar model fit. In 4-category situation, robust estimation methods are much smaller in AC and $SE_{\chi^2}$ than nonrobust estimation methods. Robust estimation methods in the 4-category situation show better model fit than nonrobust estimation methods. Moreover, there are only slight differences between ML and ULS. In the 7-category situation, robust estimation methods have a bit smaller AC and $SE_{\chi^2}$ than nonrobust estimation methods. Again, a very slight difference can be found between ML and ULS. The situation of the 7-category seems like an intermediate between continuous and the 4-category data. To sum up, estimation methods have different effects on different kinds of data. For continuous data, the differences between the four estimation methods are negligible. For the 4-category data, robust methods are much better than nonrobust methods in aspect of bias, precision and model fit. However, the predominance of robust methods is not that apparent for the 7-category data.

5 Empirical Study

In this section, the background of the questionnaires and the two-factor CFA model used in this thesis are described. Further, descriptive statistics are presented and the issue of missing values is discussed. Last, CFA model results using the data gathered from questionnaires are discussed.

5.1 Background of questionnaire

Since data analysis has become more and more important nowadays, statistical courses are compulsory for many students. Learning basic statistical knowledge can prepare students better for their future work with data. Though useful, statistics is perceived as one of the hardest subjects to study. It can be rational to state that social science students feel stressed when learning statistical subjects.

A survey was conducted at Uppsala University by the statistics department. Respondents were students who attended a basic course in statistics, covering basics in descriptive statis-
tics, probability theory and some inference. In the questionnaire of the survey, students were asked to answer five questions regarding how they feel when having the statistics class. These questions were measured by Likert-type scale and visual analogue scale simultaneously. Thus, from every piece of questionnaire, we got both ordinal data (from Likert-type scale) and continuous data (from visual analogue scale). Beside, two versions of questionnaires were designed, with the only difference in the numbers of response categories in Likert-type scale. The number of response categories is four in version 1, while in version 2 it is seven. The difference in Likert-type scale was designed to find out the impact of the number of categories for the Likert-type scale on model results. Two versions of questionnaires were sent out randomly to students in the class. In total, 140 copies of version 1 questionnaire and 142 copies of version 2 questionnaire are collected.

The CFA model used in this thesis is presented below. Based on the five questions listed in the questionnaire, a two-factor CFA model is build. The five questions in the questionnaire are:

**Question 1** I have a hard time to concentrate

**Question 2** I feel bad

**Question 3** I am nervous

**Question 4** I feel tense

**Question 5** I feel stressed

Question 1, Question 2 and Question 3 can be assumed to be affected by a factor $\xi_1$ called ‘Indirect Stress’. It is a factor which will cause bad feelings which are evoked when people feel stressed. For example, people may talk less and eat less when they feel stressed. Question 4 and Question 5 are assumed to be affected by $\xi_2$, which is called ’Direct Stress’. ’Direct Stress’ is the direct measurement of feeling stressed. The matrix form of the model looks like:

$$
\begin{pmatrix}
Q_1 \\
Q_2 \\
Q_3 \\
Q_4 \\
Q_5
\end{pmatrix} =
\begin{pmatrix}
\lambda_1 & 0 \\
\lambda_2 & 0 \\
\lambda_3 & 0 \\
0 & \lambda_4 \\
0 & \lambda_5
\end{pmatrix}
\begin{pmatrix}
\xi_1 \\
\xi_2
\end{pmatrix} +
\begin{pmatrix}
\delta_1 \\
\delta_2 \\
\delta_3 \\
\delta_4 \\
\delta_5
\end{pmatrix}.
$$
Figure 2: Path Diagram for CFA model

A path diagram shown in Figure 2 helps to give an intuitive sight of what the model looks like.

Data obtained from two versions of questionnaires are used to verify our hypothesis about the structure of the CFA model.

5.2 Missing values

Empirical data sets often have missing values. Missing values will diminish the sample size and may cause biased estimation when missing values are correlated with non-missing values. It is crucial to first change incomplete data to complete data and then continue to build models. To handle the missing data problem, two options are available: deletion and imputation. Schafer (1997) has pointed out that deletion is not appropriate when the proportion of missing values exceeds 5%. In the data sets we have, incomplete data account for 15.7% and 19.7% for version 1 questionnaire and version 2 questionnaire, respectively. Besides, the sample sizes are quite small: 140 observations in version 1 and 142 observations in version 2. In order to maintain as many observations as possible, imputation is preferred to solve the missing data problem in our data sets.

A more important matter that needs to be considered when handling the missing data problem is to know the mechanisms of missing data. Whether the appearance of missing data is by design or not? If not, are the data missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR)? Rubin (1976) gives a definition of MAR:
the chance of missingness can not be related to data values which are missing, but may be related to observed data values. This means missing values can be estimated through the relationship with observed data values. A special case, which is a stricter one, of MAR is missing completely at random (MCAR). MCAR requires that the probability of missingness is neither related to observed data values nor data values which are missing. Brown (2012, p.364) states that MAR is more likely to hold in empirical data than MCAR. It is reasonable to believe this since MCAR is such a rigid assumption. Using SPSS to check the data sets we have, it corresponds to Brown’s opinion that MCAR does not hold in our data sets. This violation has enhanced that listwise deletion and pairwise deletion are not suitable for the situation (Schafer and Graham, 2002; Jöreskog, 2005).

Unlike MCAR, MAR is not testable. But here we assume MAR holds after checking the data sets carefully. Missing values follow some pattern in the data sets: within one questionnaire, either the whole continuous data or the whole ordinal data are missing. As every question is described by both continuous and ordinal scales, students may be impatient to answer the same question twice, which leads to one kind of data are totally missing. Therefore we assume that MAR holds for our data sets.

Imputation, rather than deletion, is chosen for the data sets based on the above discussions. Multiple imputation (MI) and Maximum likelihood (ML) are two popular ways of dealing with missing data. ML method is not chosen for the data sets in our case for the following reasons. Schafer and Olsen (1998) indicate that realization of the ML method require complicated calculation and is not generalized (imputation should be done based on type of model). MI, on the other hand, can provide complete data sets for various models. Schafer and Graham (2002) do a comparison between these two methods. From their point of view, MI outperforms ML when sample sizes are small and is robust to violation of the assumption of multinormal distribution. This is the case we have: sample sizes are small and continuous data are not multinormally distributed. Due to its nice properties and ease of use, MI is used to deal with missing data in our case.

5.3 Descriptive analyses

For continuous data, their Pearson product-moment correlation matrix, means and standard errors are given, while polychoric correlation matrix is given for ordinal data. For detailed information, four distribution figures for our data are put in Appendix A.
With the help of LISREL, it is easy to get all descriptive statistics both for continuous and ordinal data. Pearson product-moment correlation matrix, means and standard errors of continuous data in version 1 questionnaire are listed in Table 2.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>1</td>
</tr>
<tr>
<td>Q2</td>
<td>0.328</td>
</tr>
<tr>
<td>Q3</td>
<td>0.420 0.366</td>
</tr>
<tr>
<td>Q4</td>
<td>0.371 0.308 0.636</td>
</tr>
<tr>
<td>Q5</td>
<td>0.309 0.228 0.525 0.640</td>
</tr>
<tr>
<td><strong>Means</strong></td>
<td><strong>2.691 1.863 2.773 3.095 3.798</strong></td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td><strong>1.337 1.201 1.709 1.660 1.728</strong></td>
</tr>
</tbody>
</table>

Pearson product-moment correlations indicate that the first three variables are closely related, with the correlation values are around 0.4. The last two variables have high correlations with each other. Means of first three variables are relatively smaller than the last two. Q2 has the smallest mean 1.863, which is the only one lower than 2. Q1 and Q3 have second lowest means, with 2.691 and 2.773, respectively. In the last two variables, their means are larger than 3. For the last one, it has the largest mean 3.798, approximately approaches 4. The order of standard errors for the five variables is the same as their means.

The relationships for ordinal variables in version 1 questionnaire are presented in the following polychoric correlation matrix:

\[
R_{V1} = \begin{pmatrix}
1.000 & & & & \\
0.176 & 1.000 & & & \\
0.367 & 0.458 & 1.000 & & \\
0.316 & 0.395 & 0.621 & 1.000 & \\
0.195 & 0.342 & 0.517 & 0.597 & 1.000
\end{pmatrix}
\]

The polychoric correlation matrix is quite similar with Pearson product-moment correlation matrix for continuous data in version 1. For Q1, variable which is most related to it is Q3, with polychoric correlation be 0.367. The same pattern works for Q2, it has the highest polychoric
correlation with Q3. This trend continues to Q4. For the last one variable Q5, it shares the largest polychoric correlation with Q4.

Similarly, Table 3 shows descriptive statistics of continuous data in version 2.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>1</td>
</tr>
<tr>
<td>Q2</td>
<td>0.326 1</td>
</tr>
<tr>
<td>Q3</td>
<td>0.313 0.561 1</td>
</tr>
<tr>
<td>Q4</td>
<td>0.472 0.502 0.647 1</td>
</tr>
<tr>
<td>Q5</td>
<td>0.387 0.389 0.663 0.573 1</td>
</tr>
</tbody>
</table>

| Means     | 2.643 1.662 2.848 2.787 3.953 |
| SD        | 1.454 1.257 1.794 1.673 1.673 |

Pearson product-moment correlation matrix is different from that of version 1, as Q1 has the largest correlation with Q4 and Q5 has the largest correlation Q3 now. Other correlations follow almost the same patterns as those in version 1. When means and standard errors are considered, it is not hard to find out that they follow almost the same patterns of those of version 1 questionnaire, with Q2 be the smallest in mean and standard error. Q1 and Q3 are the second smallest with regard to mean. Difference in mean between Q2 and Q1 becomes more apparent compared to data in version 1. The same happens for disparity between pair (Q3, Q1) and pair (Q4, Q5). Standard errors behave exactly the same with those of version 1 questionnaire except for the last two have same values now.

The polychoric correlation matrix for version 2 is presented as follows:

\[ R_{V2} = \begin{pmatrix}
1.000 & \\
0.384 & 1.000 \\
0.378 & 0.543 & 1.000 \\
0.501 & 0.581 & 0.697 & 1.000 \\
0.431 & 0.443 & 0.712 & 0.630 & 1.000
\end{pmatrix} \]

The polychoric correlation matrix is slightly different from the Pearson product-moment correlation matrix, which does not happen in version 1. Variable which has largest correlation
with Q2 has changed from Q3 to Q4. The inconsistency in these two correlation matrix in version 2 causes problems in CFA models. Either continuous data or ordinal data from version 2 questionnaire would have poor model fits.

5.4 CFA results

Data gained from the survey mentioned above are used to build a two-factor CFA model. Model results based on four estimation methods (ML, RML, ULS and RULS), two types of data (continuous, ordinal) and two kinds of questionnaires (version 1, version 2) are presented in table 4 and 5. χ² test and RMSEA are chosen here to access model fits. As χ² test measures the magnitude of discrepancy between the sample and fitted covariance/correlation matrices, RMSEA tells researchers how well their model will fit the data. RMSEA is a supplement for χ² test. (Hu and Bentler, 1999; Schreiber et al., 2006; Hooper et al., 2008). Analyses start first with version 1 questionnaires, then with version 2 questionnaires.

5.4.1 Version 1—four categories in ordinal data

All parameter estimates are significantly different from zero in version 1 questionnaire. The absolute values of the first two factor loadings are always small, compared to the last three. However, the differences of parameter estimates in ordinal data are less obvious than those in continuous data. Generally speaking, all parameters range, roughly, from 0.4 to 0.9 in ordinal cases. In continuous cases, values of the first two parameters are both less than 0.7, while for the last three they are all greater than 1.2.

Standard errors of parameter estimates in ordinal cases are less than those in the continuous case when using ML, RML and RULS as estimation methods, which means that steadier parameter estimates are produced when using ordinal data. When ULS method is used, the converse result appears, that is, standard errors with continuous data are smaller than those with ordinal data. Moreover, standard errors of parameter estimates in ULS are apparently smaller than those in other three methods when continuous data are used. Differences in standard errors of parameter estimates are not very obvious in ordinal data, with ULS having slightly smaller standard errors than other three methods.

All models work well for version 1 for p values of χ² are larger than 0.05 (the standard significant level). It means that for data in version 1 questionnaires, overall model fits are good. Take a close look at the difference between continuous and ordinal data, it is clear that ordinal
data have smaller values in $\chi^2$. Absolute values of ordinal data are quite small compared to those of continuous data, though all of them are smaller than critical value. Browne and Cudeck (1992) gives the cutoff criteria of RMSEA. When RMSEA is under 0.05, it indicates a close fit. When RMSEA ranges from 0.05 to 0.08, it indicates a fair fit. If RMSEA is larger than 0.08, a poor fit is obtained. RMSEA of all models are 0, which indicates that all of them have close fits.

As the model fits the data well in version 1 questionnaires, factor loadings can be interpreted as validity coefficients. The one with the largest factor loading is the most valid indicator for its corresponding latent variable. Thus Q3 is the most valid indicator for 'Indirect Stress' and Q4 is the most valid indicator for 'Direct Stress', under all situations in version 1 questionnaires.

5.4.2 Version 2—seven categories in ordinal data

It can be foreseen that outcomes would be different from version 1 as correlation matrices are not similar. However, there are still some similarities in these two versions. All estimates of factor loadings are significant different from zero, which is the same case in version 1 questionnaires. Absolute values of all factor loadings are between 0.5 and 0.9 in ordinal cases. In continuous cases, the first two factor loadings range from 0.69 to 0.77 and the last three ranges from 1.21 to 1.5. It is, again, that factor loadings are more disparate in continuous case.

Standard errors of data in version 2 questionnaires are similar with those of in version 1. In ML, RML and RULS, ordinal data have smaller standard errors while in ULS reverse situation occurs. Multinormal assumption is violated in this version as well. ULS method gives much smaller standard errors of parameter estimates than other estimation methods for continuous data again in version 2. For ordinal data in version 2, standard errors of parameter estimates are small and similar.

Overall model fits are not as good as version 1, $\chi^2$ values are quite large in contrast to data in version 1. Only two $p$s of $\chi^2$ are larger than 0.05, which means that only two models show overall good model fits. Ordinal data have smaller $\chi^2$ values and larger $p$ values than continuous data in version 2 questionnaire. RMSEA of all continuous data are larger than 0.1 which indicates poor model fits (a rule by Browne and Cudeck, 1992). In ordinal cases, situation is a little bit better. The only two models which show good overall model fit appear when using ordinal data, which are RML and RULS. Corresponding RMSEA are 0.089 and 0.087, which can be treated as barely fair model fits. In these two model, the most valid
indicators for 'Indirect Stress’ and 'Direct Stress’ are the same with version 1 questionnaire, that is Q3 and Q4.

5.4.3 Summary of the two versions

CFA results of the two versions of questionnaires are similar in many aspects. For instance, when use ULS estimation method to deal with continuous data, the smallest standard errors of parameter estimates are obtained. Besides, the values of $\chi^2$ are smaller when using ordinal data. Nevertheless, differences of model results in two versions of questionnaires should be pointed out. Model results for data in version 2 are worse than those in version 1. Only two out of eight models in version 2 show overall good fit, while all models in version 1 show good model fits. The results mean that the CFA model we use in this thesis does not fit the sample we get when using version 2 questionnaire. Data descriptions for data in version 2 confirm the unsuitability of the model for data in version 2. Two correlation matrices in version 2 indicate that Q1, Q2 are not in the same group with Q3 for they do not have high correlations, compared to data in version 1 questionnaire. Therefore, the sample we have in version 2 questionnaire is not suitable for the specified CFA model.

Some of the empirical study results don’t support what we find in simulation study. In simulation study, continuous data act best in terms of bias, precision and model fit, followed by 7-category and 4-category data. While empirical study show that 4-category data have the best model fit, followed by continuous and 7-category data. The strange result appear because we have non-multinormal continuous data in both versions of questionnaire. What we have in simulation study is multinormal continuous data. Violation of multinormality brings server consequence to CFA when continuous data are used. Worse performance of 7-category data than 4-category data can be explained by the samples we get. It can be seen that all the data from version 2 questionnaire don’t suit the CFA model we choose. The two samples we get, using version 1 and 2 questionnaire, are different in their underlying stress.
Table 4: Model results for data in Version 1 Questionnaires

<table>
<thead>
<tr>
<th></th>
<th>ML Continuous Data</th>
<th>ML Ordinal Data</th>
<th>RML Continuous Data</th>
<th>RML Ordinal Data</th>
<th>ULS Continuous Data</th>
<th>ULS Ordinal Data</th>
<th>RULS Continuous Data</th>
<th>RULS Ordinal Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_1 )</td>
<td>0.684 (0.117)</td>
<td>0.412 (0.0894)</td>
<td>0.684 (0.115)</td>
<td>0.412 (0.0869)</td>
<td>0.687 (0.0396)</td>
<td>0.394 (0.0628)</td>
<td>0.687 (0.117)</td>
<td>0.394 (0.121)</td>
</tr>
<tr>
<td>( \lambda_2 )</td>
<td>0.529 (0.107)</td>
<td>0.533 (0.0869)</td>
<td>0.529 (0.0928)</td>
<td>0.533 (0.0896)</td>
<td>0.521 (0.0381)</td>
<td>0.534 (0.0656)</td>
<td>0.521 (0.0971)</td>
<td>0.534 (0.0882)</td>
</tr>
<tr>
<td>( \lambda_3 )</td>
<td>1.446 (0.146)</td>
<td>0.862 (0.0851)</td>
<td>1.446 (0.143)</td>
<td>0.862 (0.0808)</td>
<td>1.469 (0.0710)</td>
<td>0.864 (0.0975)</td>
<td>1.469 (0.145)</td>
<td>0.864 (0.0845)</td>
</tr>
<tr>
<td>( \lambda_4 )</td>
<td>1.466 (0.130)</td>
<td>0.851 (0.0808)</td>
<td>1.466 (0.121)</td>
<td>0.851 (0.0805)</td>
<td>1.467 (0.0573)</td>
<td>0.858 (0.0932)</td>
<td>1.467 (0.121)</td>
<td>0.858 (0.0869)</td>
</tr>
<tr>
<td>( \lambda_5 )</td>
<td>1.252 (0.139)</td>
<td>0.701 (0.0821)</td>
<td>1.252 (0.117)</td>
<td>0.701 (0.0658)</td>
<td>1.251 (0.0489)</td>
<td>0.695 (0.0778)</td>
<td>1.251 (0.116)</td>
<td>0.695 (0.0702)</td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td>3.17 (p=0.531)</td>
<td>1.76 (p=0.780)</td>
<td>2.29 (p=0.683)</td>
<td>0.5 (p=0.973)</td>
<td>3.15 (p=0.534)</td>
<td>1.73 (p=0.785)</td>
<td>2.27 (p=0.687)</td>
<td>0.5 (p=0.974)</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses are standard errors.
## Table 5: Model results for data in Version 2 Questionnaires

<table>
<thead>
<tr>
<th></th>
<th>ML</th>
<th></th>
<th>RML</th>
<th></th>
<th>ULS</th>
<th></th>
<th>RULS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Continuous Data</td>
<td>Ordinal Data</td>
<td>Continuous Data</td>
<td>Ordinal Data</td>
<td>Continuous Data</td>
<td>Ordinal Data</td>
<td>Continuous Data</td>
<td>Ordinal Data</td>
</tr>
<tr>
<td>(\lambda_1)</td>
<td>0.712</td>
<td>0.537</td>
<td>0.712</td>
<td>0.537</td>
<td>0.699</td>
<td>0.536</td>
<td>0.699</td>
<td>0.536</td>
</tr>
<tr>
<td></td>
<td>(0.122)</td>
<td>(0.0814)</td>
<td>(0.125)</td>
<td>(0.0819)</td>
<td>(0.0344)</td>
<td>(0.0566)</td>
<td>(0.126)</td>
<td>(0.0845)</td>
</tr>
<tr>
<td>(\lambda_2)</td>
<td>0.761</td>
<td>0.623</td>
<td>0.761</td>
<td>0.623</td>
<td>0.760</td>
<td>0.637</td>
<td>0.760</td>
<td>0.637</td>
</tr>
<tr>
<td></td>
<td>(0.102)</td>
<td>(0.0792)</td>
<td>(0.119)</td>
<td>(0.0863)</td>
<td>(0.0359)</td>
<td>(0.0602)</td>
<td>(0.122)</td>
<td>(0.0879)</td>
</tr>
<tr>
<td>(\lambda_3)</td>
<td>1.493</td>
<td>0.821</td>
<td>1.493</td>
<td>0.821</td>
<td>1.458</td>
<td>0.818</td>
<td>1.458</td>
<td>0.818</td>
</tr>
<tr>
<td></td>
<td>(0.135)</td>
<td>(0.0738)</td>
<td>(0.108)</td>
<td>(0.0422)</td>
<td>(0.0555)</td>
<td>(0.0701)</td>
<td>(0.113)</td>
<td>(0.0417)</td>
</tr>
<tr>
<td>(\lambda_4)</td>
<td>1.315</td>
<td>0.818</td>
<td>1.315</td>
<td>0.818</td>
<td>1.321</td>
<td>0.834</td>
<td>1.321</td>
<td>0.834</td>
</tr>
<tr>
<td></td>
<td>(0.127)</td>
<td>(0.0730)</td>
<td>(0.118)</td>
<td>(0.0482)</td>
<td>(0.0473)</td>
<td>(0.0751)</td>
<td>(0.118)</td>
<td>(0.0465)</td>
</tr>
<tr>
<td>(\lambda_5)</td>
<td>1.219</td>
<td>0.771</td>
<td>1.219</td>
<td>0.771</td>
<td>1.213</td>
<td>0.756</td>
<td>1.213</td>
<td>0.756</td>
</tr>
<tr>
<td></td>
<td>(0.129)</td>
<td>(0.0744)</td>
<td>(0.134)</td>
<td>(0.0530)</td>
<td>(0.0437)</td>
<td>(0.0685)</td>
<td>(0.134)</td>
<td>(0.0543)</td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>(\chi^2)</td>
<td>14.98</td>
<td>10.56</td>
<td>14.10</td>
<td>8.44</td>
<td>14.67</td>
<td>10.41</td>
<td>13.88</td>
<td>8.27</td>
</tr>
<tr>
<td></td>
<td>((p=0.005))</td>
<td>((p=0.032))</td>
<td>((p=0.007))</td>
<td>((p=0.077))</td>
<td>((p=0.005))</td>
<td>((p=0.034))</td>
<td>((p=0.008))</td>
<td>((p=0.082))</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.140</td>
<td>0.108</td>
<td>0.134</td>
<td>0.089</td>
<td>0.138</td>
<td>0.107</td>
<td>0.132</td>
<td>0.087</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses are standard errors.
6 Conclusion

Differences between continuous and ordinal data are discussed in this thesis. Due to the meaninglessness of mean and covariance matrix for ordinal data, adjustments are needed when using CFA with ordinal data. Former studies have suggested that calculating polychoric correlation matrix is suitable for ordinal data. The way to calculate polychoric correlation is explained, which is the work of Olsson (1979).

Estimation methods to deal with violation in multinormality assumption in continuous data and ordinal data are also reviewed. For non-multinormality in continuous data, Satorra-Bentler rescaled method, the robust method, is recommended, and ULS is suggested for ordinal data.

Simulation study show that continuous data are better than 7-category and 4-category data in terms of parameter estimates, standard errors and $\chi^2$. Categorization brings harm to estimating CFA model, and the more degree of categorization, the more harm it brings. Estimation methods have different impacts on different kinds of data. For continuous data, the chosen four estimation methods are almost the same. For 4-category data, robust methods are much better than nonrobust methods in terms of parameter estimates, standard errors and $\chi^2$. The situation of 7-category data is an intermediate between continuous and 4-category data.

Empirical study show contrary results. 4-category data show the best model fit, followed by continuous data and 7-category data. The difference can be explained by the violation of multinormality in continuous data and the different underlying stress level between our two samples.

Further studies may make deeper research on the following two aspects. First, the categorization of continuous data can be more diverse. In this thesis, we only use symmetric thresholds and two different numbers of categories. Asymmetric thresholds and more numbers of categories can be chosen to access the impact of categorization on CFA results. Second, the continuous data we use in simulation is multinormally distributed. It can be interesting to use continuous data with different level of violation of multinormality to check the influence brought by non-multinormality.
References


Appendix

Distribution for variables in questionnaires

Figure 3: Distribution for the first three variables in Version 1 questionnaire
Figure 4: Distribution for the last two variables in Version 1 questionnaire
Figure 5: Distribution for the first three variables in Version 2 questionnaire
Figure 6: Distribution for the last two variables in Version 2 questionnaire