Confirmatory Factor Analysis of Ordinal Variables With Misspecified Models

Fan Yang-Wallentin and Karl G. Jöreskog
Uppsala University and Norwegian School of Management

Hao Luo
Uppsala University

Ordinal variables are common in many empirical investigations in the social and behavioral sciences. Researchers often apply the maximum likelihood method to fit structural equation models to ordinal data. This assumes that the observed measures have normal distributions, which is not the case when the variables are ordinal. A better approach is to use polychoric correlations and fit the models using methods such as unweighted least squares (ULS), maximum likelihood (ML), weighted least squares (WLS), or diagonally weighted least squares (DWLS). In this simulation evaluation we study the behavior of these methods in combination with polychoric correlations when the models are misspecified. We also study the effect of model size and number of categories on the parameter estimates, their standard errors, and the common chi-square measures of fit when the models are both correct and misspecified. When used routinely, these methods give consistent parameter estimates but ULS, ML, and DWLS give incorrect standard errors. Correct standard errors can be obtained for these methods by robustification using an estimate of the asymptotic covariance matrix $W$ of the polychoric correlations. When used in this way the methods are here called RULS, RML, and RDWLS.

Structural equation modeling (SEM) is widely used in the social and behavioral sciences, and within this area, confirmatory factor analysis (CFA) is the most common type of analysis. CFA was originally developed for continuous variables using the maximum likelihood (ML) method, which assumes that the observed variables have a multivariate normal distribution (see, e.g., Jöreskog, 1969). For continuous nonnormal variables, Browne (1984) developed an asymptotically distribution free (ADF) method, which is a weighted least squares (WLS) method using the inverse of the asymptotic covariance matrix $W$ of the sample variances and covariances as a weight matrix.

The variables used in many empirical studies in the social and behavioral sciences are often ordinal rather than continuous. Observations on an ordinal variable are assumed to represent
responses to a set of ordered categories such as a five-category Likert scale. This is typical when data are collected through questionnaires. Although a question might be designed to measure a theoretical concept, the observed responses are only a discrete realization of a small number of categories.

Methods developed for continuous variables are not suitable for such ordinal variables (see B. O. Muthén & Kaplan, 1985, 1992). Jöreskog (1990) suggested that polychoric correlations and WLS could be used to estimate a CFA model based on ordinal variables. B. O. Muthén (1984) developed a general WLS methodology for continuous and categorical variables (CVM). Jöreskog (1994) derived the asymptotic covariance matrix of the polychoric correlations. In the special case of only ordinal variables the weight matrices in B. O. Muthén (1984) and in Jöreskog (1994) are very similar and applying them to WLS gives almost identical results.

A common experience with WLS both for continuous and categorical variables is the poor performance of the WLS estimators and their associated standard errors and chi-squares (see, e.g., Bentler, 1995; Dolan, 1994; Flora & Curran, 2004; Potthast, 1993; West, Finch, & Curran, 1995). Most likely these problems occur because the estimate of $W^{-1}$ used as a weight matrix is very unstable unless the sample size is very large.

A better method of estimation for ordinal variables was proposed by B. O. Muthén, du Toit, and Spisic (1997). This method is called Robust WLS in L. K. Muthén and Muthén (1998, pp. 357–358). It is similar to the method termed DWLS by Jöreskog and Sörbom (1996a, pp. 23–24). Both methods use only the diagonal elements of $W$ in the fitting of the model and use the full $W$ to obtain correct standard errors and chi-squares. Simulation studies by B. O. Muthén et al. (1997) and by Flora and Curran (2004) show that Robust WLS works much better than full WLS especially for large models and small to moderate sample sizes under correct specification of the model. In this article, we study the behavior of robust unweighted least squares (RULS), robust maximum likelihood (RML), and robust diagonally weighted least squares (RDWLS) under misspecified models and we find that these methods perform better than full WLS. We also find that none of these three methods is uniformly better than the other. However, our results also show that the simpler method of RULS and even RML, if used in a specific way, also produce estimates and standard errors that are equally good.

**RESEARCH QUESTIONS**

The purpose of this study is to compare the performance of different estimators (i.e., RULS, RML, WLS, and RDWLS) for estimating the parameters in confirmatory factor analysis models under conditions of correctly and incorrectly specified models using data on ordinal variables. We also examine the effect of number of categories and their probability distribution. We study one small and one large model. Similar simulation studies have been reported by Potthast (1993) and Dolan (1994). However, as far as we are aware, no one has studied all these methods under misspecified models.

We attempt to answer the following research questions:

1. Is any of these methods uniformly better or worse than the others on criteria such as bias and mean square error?
2. Is any of these methods uniformly better or worse than the others in estimating the standard errors of the estimated parameters, using the same criteria?
3. Do bias and mean square error increase with increasing degree of misspecification or decreasing number of categories?
4. Do problems of nonconvergence increase with increasing degree of misspecification or decrease with increasing number of categories? Nonconvergence is discussed later.
5. How large a sample is needed to avoid problems of nonconvergence? How does this sample size depend on the degree of misspecification and on the number of categories?

We are not aware of any studies that address all these questions.

CONFIRMATORY FACTOR ANALYSIS WITH ORDINAL VARIABLES

Ordinal Variables

Let \( x_1, x_2, \ldots, x_p \) be \( p \) ordinal variables to be analyzed. Following B. O. Muthén (1984), Lee, Poon, and Bentler (1990), Jöreskog (1990), and others, it is assumed that there is a continuous variable \( x_i^* \) underlying the ordinal variable \( x_i \). This continuous variable \( x_i^* \) represents the attitude underlying the ordered responses to \( x_i \) and is assumed to have a range from \(-\infty\) to \( +\infty\). It is the underlying variable \( x_i^* \) that is assumed to follow a confirmatory factor analysis model.

The underlying variable \( x_i^* \) is unobservable. Only the ordinal variable \( x_i \) is observed. For an ordinal variable \( x_i \) with \( m_i \) categories, the connection between the ordinal variable \( x_i \) and the underlying variable \( x_i^* \) is

\[
x_i = c \iff \tau_{c-1}^{(i)} < x_i^* < \tau_c^{(i)}, \ c = 1, 2, \ldots, m_i .
\]  

where

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

are threshold parameters. For variable \( x_i \) with \( m_i \) categories, there are \( m_i - 1 \) strictly increasing threshold parameters \( \tau_1^{(i)}, \tau_2^{(i)}, \ldots, \tau_{m_i-1}^{(i)} \).

Because only ordinal information is available about \( x_i \), the distribution of \( x_i^* \) is determined only up to a monotonic transformation. It is convenient to let \( x_i^* \) have the standard normal distribution with density function \( \phi(\cdot) \) and distribution function \( \Phi(\cdot) \). Then the probability \( \pi_c^{(i)} \) of a response in category \( c \) on variable \( x_i \), is

\[
\pi_c^{(i)} = Pr[x_i = c] = Pr[\tau_{c-1}^{(i)} < x_i^* < \tau_c^{(i)}] = \int_{\tau_{c-1}^{(i)}}^{\tau_c^{(i)}} \phi(u)du = \Phi(\tau_c^{(i)}) - \Phi(\tau_{c-1}^{(i)}) ,
\]  

for \( c = 1, 2, \ldots, m_i - 1 \), so that

\[
\tau_c^{(i)} = \Phi^{-1}(\pi_1^{(i)} + \pi_2^{(i)} + \cdots + \pi_c^{(i)}) .
\]
where $\Phi^{-1}$ is the inverse of the standard normal distribution function. The quantity $(\pi_1^{(i)} + \pi_2^{(i)} + \cdots + \pi_c^{(i)})$ is the probability of a response in category $c$ or lower.

The probabilities $\pi_c^{(i)}$ are unknown population quantities. In practice, $\pi_c^{(i)}$ can be estimated consistently by the corresponding percentage $p_c^{(i)}$ of responses in category $c$ on variable $x_i$. Then, estimates of the thresholds can be obtained as

$$\hat{\tau}_c^{(i)} = \Phi^{-1}(p_1^{(i)} + p_2^{(i)} + \cdots + p_c^{(i)}), \quad c = 1, \ldots, m - 1. \tag{5}$$

The quantity $(p_1^{(i)} + p_2^{(i)} + \cdots + p_c^{(i)})$ is the proportion of cases in the sample responding in category $c$ or lower on variable $x_i$.

**Polychoric Correlations**

Let $x_i$ and $x_j$ be two ordinal variables with $m_i$ and $m_j$ categories, respectively. Their marginal distribution in the sample is represented by a contingency table

$$
\begin{pmatrix}
  n_{11}^{(ij)} & n_{12}^{(ij)} & \cdots & n_{1m_j}^{(ij)} \\
  n_{21}^{(ij)} & n_{22}^{(ij)} & \cdots & n_{2m_j}^{(ij)} \\
  \vdots & \vdots & \ddots & \vdots \\
  n_{m_i1}^{(ij)} & n_{m_i2}^{(ij)} & \cdots & n_{m_im_j}^{(ij)} 
\end{pmatrix}, \tag{6}
$$

where $n_{ab}^{(ij)}$ is the number of cases in the sample in category $a$ on variable $x_i$ and in category $b$ on variable $x_j$. The underlying variables $x_i^*$ and $x_j^*$ are assumed to be bivariate normal with zero means, unit variances, and with correlation $\rho_{ij}$, the polychoric correlation.

Let $\tau_1^{(i)}, \tau_2^{(i)}, \ldots, \tau_{m_i-1}^{(i)}$ be the thresholds for variable $x_i^*$ and let $\tau_1^{(j)}, \tau_2^{(j)}, \ldots, \tau_{m_j-1}^{(j)}$ be the thresholds for variable $x_j^*$. The polychoric correlation can be estimated by maximizing the log-likelihood of the multinomial distribution (see Olsson, 1979)

$$\ln L = \sum_{a=1}^{m_i} \sum_{b=1}^{m_j} n_{ab}^{(ij)} \log \pi_{ab}^{(ij)}, \tag{7}$$

where

$$\pi_{ab}^{(ij)} = Pr[x_i = a, x_j = b] = \int_{\tau_{b-1}^{(j)}}^{\tau_b^{(j)}} \int_{\tau_{a-1}^{(i)}}^{\tau_a^{(i)}} \phi_2(u, v) du dv, \tag{8}$$

and

$$\phi_2(u, v) = \frac{1}{2\pi \sqrt{1 - \rho^2}} e^{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho u v + v^2)}, \tag{9}$$

is the standard bivariate normal density with correlation $\rho_{ij}$. Maximizing $\ln L$ gives the sample polychoric correlation denoted $r_{ij}$. 


The polychoric correlation can be estimated by a two-step procedure (see Olsson, 1979). In the first step, the thresholds are estimated from the univariate marginal distributions by Equation 5. In the second step, the polychoric correlations are estimated from the bivariate marginal distributions by maximizing $\ln L$ for given thresholds. The parameters can also be estimated by a one-step procedure that maximizes $\ln L$ with respect to the thresholds and the polychoric correlation simultaneously but this is not necessary because the estimates are almost the same as with the two-step procedure and it is not practical because it would yield different thresholds for the same variable when paired with different variables. For an example, see Jöreskog (2002–2005, Table 3, p. 13).

Jöreskog (1994) showed that the polychoric correlation $r_{ij}$ is asymptotically linear in the bivariate marginal proportions $P_{ij}$, where $P_{ij}$ is a matrix of order $m_i \times m_j$ whose elements are $p_{ab}^{(ij)} = n_{ab}^{(ij)}/N$, where $N$ is the sample size. Thus, $r_{ij} \simeq tr(\Gamma_{ij}P_{ij})$. The elements of the matrix $\Gamma_{ij}$ are given in Jöreskog (1994, Equation 16). Using this result one can estimate the asymptotic covariance $N \times Acov(r_{gh}, r_{ij})$ for all $g \neq h$ and $i \neq j$ (see Jöreskog, 1994, for details).

**ESTIMATION METHODS**

For continuous variables, several methods are available for estimating structural equation models and confirmatory factor analysis models: ML and various least squares methods. If combined with an estimated asymptotic covariance matrix, these methods can also provide correct standard errors even for nonnormal variables under certain assumptions. None of these methods can be used directly with ordinal variables but can be used in modified forms to fit the models to polychoric correlations.

The model to be estimated is a factor model of the form

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

where $x^*$ is a vector of order $p \times 1$ of underlying variables corresponding to the $p \times 1$ vector of the observed ordinal variables $x$, as defined earlier. The vectors $\xi$ of order $k \times 1$ and $\delta$ of order $p \times 1$ represent the factors and the unique variables that are assumed to be uncorrelated. The matrix $\Lambda$ of order $p \times k$ contains the factor loadings $\lambda_{ij}$. Some elements of $\Lambda$ might be fixed at zero.

Let $\Phi$ and $\Theta$ be the covariance matrices of $\xi$ and $\delta$, respectively. We assume that the unique factors are uncorrelated so that $\Theta$ is a diagonal matrix. For convenience we assume that $\Phi$ is a correlation matrix with ones in the diagonal. The covariance matrix of $x^*$ is

$$\Sigma = \Lambda \Phi \Lambda' + \Theta.$$

Because the underlying variables $x_i^*$ have variances equal to 1, it follows that

$$\Theta = I - \text{diag}(\Lambda \Phi \Lambda'),$$

so that

$$\Sigma(\Lambda, \Phi) = \Lambda \Phi \Lambda' + I - \text{diag}(\Lambda \Phi \Lambda').$$
We write $\Sigma(\Lambda, \Phi)$ to emphasize that $\Sigma$ is a function of $\Lambda$ and $\Phi$. This is the correlation matrix implied by the model that is to be fitted to the matrix of polychoric correlations $R$.

To estimate the model four alternative methods are considered and compared in this article, namely unweighted least squares (ULS), diagonally weighted least squares (DWLS), WLS, and ML. In the estimation, the constraints in Equation 13 are achieved by using the constrained parameter feature in LISREL (see Jöreskog & Sörbom, 1996a, pp. 345–349). An example LISREL syntax file is given in the Appendix.

### Three Least Squares Methods

The three least squares methods are two-step methods. In the first step the polychoric correlations $r$ and their asymptotic covariance matrix $W$ are estimated as described earlier. Note that $r = (r_{21}, r_{31}, r_{32}, \ldots, r_{p,p-1})'$ is a vector of the polychoric correlations below the diagonal of the polychoric correlation matrix $R$. The 1s in the diagonal are not included in the vector $r$. As described earlier, both $r$ and $W$ are estimated from the sample data without the use of the model. Let $s = p(p-1)/2$. The vector $r$ is of order $s \times 1$ and the matrix $W$ is of order $s \times s$. The matrix $W$ contains the elements of the estimated $N \times ACov(r_{gh}, r_{ij})$ arranged to correspond to $r$.

In the second step $\Lambda$ and $\Phi$ are fitted to $r$ by minimizing the fit function

$$F(r, \Lambda, \Phi) = [r - \rho(\Lambda, \Phi)]' V [r - \rho(\Lambda, \Phi)],$$

where $V$ is a positive matrix and $\rho(\Lambda, \Phi)$ is a vector of the elements of $\Lambda \Phi \Lambda'$ below the diagonal. The three least squares methods differ in the choice of weight matrix $V$:

**ULS** : $V = I$ 

**DWLS** : $V = (\text{diag} W)^{-1}$

**WLS** : $V = W^{-1}$

The main difference between these weight matrices is that for ULS and DWLS the weight matrix is diagonal, whereas for WLS the weight matrix is the inverse of the full matrix $W$. For DWLS only the diagonal elements of $W$ are used. In scalar form these fit functions can be written as

**ULS** : $F(r, \Lambda, \Phi) = \sum_i (r_i - \rho_i)^2$

**DWLS** : $F(r, \Lambda, \Phi) = \sum_i (r_i - \rho_i)^2 / w_{ii}$

**WLS** : $F(r, \Lambda, \Phi) = \sum_i \sum_j (r_i - \rho_i)(r_j - \rho_j) w^{ij}$

where $w^{ij}$ is an element of $W^{-1}$.
The Maximum Likelihood Method

The method of ML has no theoretical justification for use with ordinal variables. Nevertheless, it works if used as follows. Let $R$ be the matrix of polychoric correlations with ones in the diagonal and let $\Sigma$ be defined as in Equation 13. The ML fit function is

$$F_1(R, \Lambda, \Phi) = \log|\Sigma| + tr(R\Sigma^{-1}) - \log|R| - p,$$

which is to be minimized with respect to the free elements of $\Lambda$ and $\Phi$.

This is of a totally different form from Equation 14. We show that this can also be written in the form of Equation 14. Let $r^* = (1, r_{21}, 1, r_{31}, 1, \ldots, r_{p,p-1}, 1)^\prime$; that is, $r^*$ is a vector of the elements of $R$ below the diagonal and including the 1s in the diagonal of $R$. Similarly, let $\rho^*(\Lambda, \Phi)$ be a vector of the corresponding elements of $\Sigma(\Lambda, \Phi)$, noting that $\Sigma$ also has 1s in the diagonal. Let $K$ be a matrix of order $s \times p(p + 1)/2$ with elements 0 and 1 such that $r = Kr^*$.

Minimizing $F_1$ in Equation 21 is equivalent to minimizing

$$F_2(r^*, \Lambda, \Phi) = [r^* - \rho^*(\Lambda, \Phi)]V^*[r^* - \rho^*(\Lambda, \Phi)],$$

with

$$V^* = D'(\hat{\Sigma}^{-1} \otimes \hat{\Sigma}^{-1})D,$$

where $\otimes$ denotes a Kronecker product, $D$ is the duplication matrix (Magnus & Neudecker, 1999, pp. 48–53), and $\hat{\Sigma} = \Sigma(\hat{\Lambda}, \hat{\Phi})$. This equation should be understood as follows. Suppose $\hat{\Lambda}$ and $\hat{\Phi}$ are estimates of $\Lambda$ and $\Phi$ in the $i$th iteration. New estimates of $\Lambda$ and $\Phi$ can be obtained by minimizing $F_2$ using $V^*$ in Equation 23. Update $V^*$ in each iteration. This iteration algorithm converges to the ML estimates $\hat{\Lambda}$ and $\hat{\Phi}$, which minimizes $F_1$. This shows that the ML estimates can be obtained by iteratively reweighted least squares. Minimizing Equation 22 is equivalent to minimizing Equation 14 using

$$V = V_{ML} = KV^*K',$$

which shows that ML also fits in the same framework as ULS, DWLS, and WLS. The only difference is that its weight matrix $V$ is a bit more complicated.

Standard Errors and Chi-Squares

For continuous variables, various formulas for asymptotic standard errors and chi-squares have been developed, notably by Browne (1984) and Satorra (1989). As far as we are aware it has not been shown that these formulas are also valid for ordinal variables and polychoric correlations. Because the vector $r$ has an asymptotic normal distribution we conjecture that these formulas can be used in modified form also in the situation considered here.

Let $\theta$ be a $t \times 1$ vector of the free elements of $\Lambda$ and $\Phi$ and let $\hat{\theta}$ be the minimizer of $F(r, \theta)$ in Equation 14 for ULS, DWLS, ML, and WLS.
Then a consistent estimate of $N \times ACov(\hat{\theta})$ is

$$\left(\hat{\Delta}'V\hat{\Delta}\right)^{-1}\hat{\Delta}'VWV\hat{\Delta}(\hat{\Delta}'V\hat{\Delta})^{-1},$$

(25)

where $\hat{\Delta} = \partial \rho / \partial \theta$ evaluated at $\hat{\theta}$. This is Browne’s (1984) formula (2.12a) applied to polychoric correlations. The standard error of $\hat{\theta}_i$ is $1/N$ times the square root of the $i$th diagonal element of this matrix. Thus, the same formula is used to obtain the standard errors for all methods. Here $W$ is the same for all methods, whereas $V$ and $\hat{\Delta}$ vary over methods. Note again that the vector $\rho$ does not include the diagonal elements of $\Sigma$.

At least four so-called chi-squares have been suggested for testing structural equation models with continuous variables. Following the notation in Jöreskog, Sörbom, Du Toit, and Du Toit (2003), these chi-squares are denoted $c_1$, $c_2$, $c_3$, and $c_4$. These are valid under different conditions. If the observed variables have a multivariate normal distribution, $c_1$ and $c_2$ have an asymptotic chi-square distribution if the model is correctly specified. $c_3$ is the Satorra and Bentler (1988) SB statistic, which is $c_1$ or $c_2$ multiplied by a scale factor that is estimated from the sample and involves an estimate of the asymptotic covariance matrix (ACM) of the sample variances and covariances. Although the asymptotic distribution of $c_3$ is not exactly $\chi^2$, it is used as a $\chi^2$ statistic because the scale factor is estimated such that $c_3$ has an asymptotically correct mean. The test statistic $c_3$ is considered as a way of correcting $c_1$ or $c_2$ for the effects of nonnormality. $c_4$ is the ADF statistic in Browne (1984, Equation 2.20a). This involves the inverse of the ACM. Browne (1984) showed that $c_4$ has an asymptotic $\chi^2$ distribution under certain standard conditions.

As far as we are aware these test statistics have not been shown to be valid for ordinal variables and polychoric correlations. For the situation considered in this article, we modify the definitions of $c_1$, $c_2$, $c_3$, and $c_4$ as follows:

- $c_1$ is $N - 1$ times the minimum value of Equation 21.
- $c_2$ is $N - 1$ times the minimum value of Equation 22.
- $c_3$ is $d/h$ times $c_2$, where $d$ is the degrees of freedom and

$$h = tr[(\Delta'_{c}V^{-1}\Delta_{c})^{-1}(\Delta'_{c}W\Delta_{c})].$$

(26)

Here $\hat{\Delta}_{c}$ is an orthogonal complement to $\hat{\Delta}$ such that $\hat{\Delta}_{c}'\hat{\Delta} = 0$.
- $c_4$ is $N - 1$ times the minimum value of the WLS fit function.

We are not claiming that any of these $c$s have an asymptotic $\chi^2$ distribution with $d$ degrees of freedom even if the model holds. The properties of the $c$s are investigated by simulation.

In the simulation study reported in this article, we found that $c_1$, $c_2$, and $c_4$ performed much worse than $c_3$, which performed reasonably well. We therefore report only the results for $c_3$. The standard errors reported in this article have been obtained from Equation 25 and the chi-squares are $c_3$ for RULS, RDWLS, and RML and $c_4$ for WLS.
Nothing is known about the estimated standard errors and chi-squares when the model does not hold. The behavior of the standard errors and chi-squares under misspecified models are studied by simulation.

SIMULATION DESIGN

Models

CFA is one of the most widely used applications in the social and behavioral sciences. The models we use in the simulation study are typical examples of CFA. Two models are used, referred to as Model 1 and Model 2. Model 1 is a small model with \( p = 6 \) observed variables and \( k = 2 \) factors and Model 2 is a large model with \( p = 16 \) observed variables and \( k = 4 \) factors. The models will be studied under correct specification and under three levels of misspecification. By correct specification we mean that the model is properly specified such that the estimated model matches the population model that is used to generate the data. Structurally misspecified models are models where the population model used to generate the data differs from the model actually estimated. Thus, for each model (Model 1 and Model 2) there are four different population models. In addition we study three different number of categories with symmetric and nonsymmetric distributions of the observed ordinal variables. Altogether there are 24 experimental cells for each model.

**Model 1.** A path diagram of Model 1 is shown in Figure 1.

In matrix form the model is:

\[
\begin{pmatrix}
\lambda_1^1 \\
\lambda_1^2 \\
\lambda_1^3 \\
\lambda_1^4 \\
\lambda_1^5 \\
\lambda_1^6 \\
\end{pmatrix} = \begin{pmatrix}
\lambda_{11} & 0 \\
\lambda_{12} & 0 \\
\lambda_{13} & 0 \\
\lambda_{14} & 0 \\
\lambda_{15} & 0 \\
\lambda_{16} & 0 \\
\end{pmatrix}
\begin{pmatrix}
\xi_1 \\
\xi_2 \\
\end{pmatrix} + \begin{pmatrix}
\delta_1 \\
\delta_2 \\
\delta_3 \\
\delta_4 \\
\delta_5 \\
\delta_6 \\
\end{pmatrix},
\]

(27)

where the correlation matrix of \( \xi_1 \) and \( \xi_2 \) is

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

(28)

and the covariance matrix of \((\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6)\) is

\[
\Theta = \text{diag}(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6).
\]

(29)

\footnote{For continuous variables, Satorra (1989, 2003) developed a robustness theory for structural equation models where it is assumed that the degree of misspecification is of the order of magnitude \(1/\sqrt{N} \) as the sample size \( N \) increases. We do not make this assumption in this article.}
When the model is estimated it is assumed that $\lambda_{31} = 0$ but the data is generated using different values of $\lambda_{31}$. Because $x_i^*$ has variance 1 the value of $\lambda_{31}$ affects the variance of $\delta_4$, denoted by the symbol $\Lambda$.

The parameter values used to generate data for Model 1 are

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6) = (0.9, 0.8, 0.7, \lambda_{31}, 0.6, 0.7, 0.8),$$

$$(\Phi_{11}, \Phi_{21}, \Phi_{22}) = (1.0, 0.6, 1.0),$$

$$(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6) = (0.19, 0.36, 0.51, \theta_{31}, 0.51, 0.36),$$

where

$$\lambda_{31} = (0, 0.1, 0.3, 0.5),$$

and

$$\theta_{31} = (0.64, 0.558, 0.334, 0.03).$$

Thus, in addition to the base model where $\lambda_{31} = 0$, there are three levels of misspecifications.
where $\theta = 0.1, 0.3, 0.5$, respectively. Note that the $\theta$ parameters are only used in the data generating process; as explained earlier, they are not involved in the estimation of the model. Thus, there are seven parameters to be estimated in $\Lambda$ and $\Phi$ for Model 1.

**Model 2.** A path diagram for Model 2 is shown in Figure 2.
In matrix form Model 2 is as follows:

\[
\begin{pmatrix}
    x_1^* \\
    x_2^* \\
    x_3^* \\
    x_4^* \\
    x_5^* \\
    x_6^* \\
    x_7^* \\
    x_8^* \\
    x_9^* \\
    x_{10}^* \\
    x_{11}^* \\
    x_{12}^* \\
    x_{13}^* \\
    x_{14}^* \\
    x_{15}^* \\
    x_{16}^*
\end{pmatrix}
= 
\begin{pmatrix}
    \lambda_{11} & 0 & 0 & \lambda_{14} \\
    \lambda_{21} & 0 & 0 & 0 \\
    \lambda_{31} & 0 & 0 & 0 \\
    \lambda_{41} & 0 & 0 & 0 \\
    0 & \lambda_{52} & 0 & 0 \\
    0 & \lambda_{62} & 0 & 0 \\
    0 & \lambda_{72} & 0 & 0 \\
    0 & \lambda_{82} & 0 & 0 \\
    0 & 0 & \lambda_{93} & 0 \\
    0 & 0 & \lambda_{103} & 0 \\
    0 & 0 & \lambda_{113} & 0 \\
    0 & 0 & \lambda_{123} & 0 \\
    0 & 0 & 0 & \lambda_{134} \\
    0 & 0 & 0 & \lambda_{144} \\
    0 & 0 & 0 & \lambda_{154} \\
    \lambda_{16,1} & 0 & 0 & \lambda_{16,4}
\end{pmatrix}
\begin{pmatrix}
    \xi_1 \\
    \xi_2 \\
    \xi_3 \\
    \xi_4
\end{pmatrix}
+ 
\begin{pmatrix}
    \delta_1 \\
    \delta_2 \\
    \delta_3 \\
    \delta_4 \\
    \delta_5 \\
    \delta_6 \\
    \delta_7 \\
    \delta_8 \\
    \delta_9 \\
    \delta_{10} \\
    \delta_{11} \\
    \delta_{12} \\
    \delta_{13} \\
    \delta_{14} \\
    \delta_{15} \\
    \delta_{16}
\end{pmatrix}
\]  

(35)

where the correlation matrix of \((\xi_1, \xi_2, \xi_3, \xi_4)\) is

\[
\Phi = \begin{pmatrix}
1 & \phi_{21} & 1 & \\
\phi_{31} & \phi_{32} & 1 & \\
\phi_{41} & \phi_{42} & \phi_{43} & 1
\end{pmatrix}
\]  

(36)

and the covariance matrix of \((\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7, \delta_8, \delta_9, \delta_{10}, \delta_{11}, \delta_{12}, \delta_{13}, \delta_{14}, \delta_{15}, \delta_{16})\) is

\[
\Theta = diag[\begin{pmatrix}
\theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}, \theta_{15}, \theta_{16}
\end{pmatrix}]
\]  

(37)

The parameter values for Model 2 are chosen as:

\[
(\lambda_{11}, \lambda_{21}, \lambda_{31}, \lambda_{41}, \lambda_{52}, \lambda_{62}, \lambda_{72}, \lambda_{82}, \lambda_{93}, \lambda_{103}, \lambda_{113}, \lambda_{123}, \lambda_{134}, \lambda_{144}, \lambda_{154}, \lambda_{16,4}) = \\
(0.4, 0.5, 0.6, 0.7, 0.8, 0.7, 0.6, 0.5, 0.6, 0.7, 0.8, 0.9, 0.8, 0.7, 0.5, 0.3),
\]

\[
(\phi_{21}, \phi_{31}, \phi_{32}, \phi_{41}, \phi_{42}, \phi_{43}) = (0.2, 0.4, 0.6, 0.8, 0.5, 0.3),
\]

\[
[\begin{pmatrix}
\theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}, \theta_{15}, \theta_{16}
\end{pmatrix}]
= \\
\begin{pmatrix}
0.75, 0.64, 0.51, 0.36, 0.51, 0.64, 0.75, 0.64, 0.51, 0.36, 0.19, 0.36, 0.51, 0.75
\end{pmatrix},
\]

where

\[
\lambda_{14} & \lambda_{16,1} = (0, 0.1, 0.3, 0.5),
\]

(38)

\[
[\theta_2] = (0.84, 0.798, 0.654, 0.43),
\]

(39)

\[
[\theta_2] = (0.91, 0.788, 0.748, 0.54),
\]

(40)
Again, the values of the $\theta$ parameters are only used to generate the data. In the model to be estimated there are 22 independent parameters in $\Lambda$ and $\Phi$.

Number of Categories

Each ordinal variable $x_i$ is assigned to have two, five, or seven categories with and without symmetric distributions as shown in Figure 3. The category probabilities are the same for all variables.

Sample Sizes and Number of Replications

Five sample sizes are used in the simulation study. They represent sample sizes commonly encountered in applied research, ranging from fairly small to fairly large. The sample sizes are 100, 200, 400, 800, and 1,600.

FIGURE 3 Category probabilities.
The percentage of nonconvergence is reported for sample sizes 100 and 200. All other outcome variables are reported for sample sizes 400, 800, and 1,600.

There are 72 experimental conditions for each model (four different specifications, three different number of categories with and without symmetric distributions, and three sample sizes) and each of these conditions was repeated 2,000 times. This number of replicates is chosen to be large enough that differences between methods can be detected if they exist.

Data Generation

Because the true $\hat{\Phi}$ and $\Theta$ are positive definite, there are lower triangular matrices $T_1$ $(k \times k)$ and $T_2$ $(p \times p)$ such that $T_1T'_1 = \Phi$ and $T_2T'_2 = \Theta$. In this case $T_2$ is diagonal. Then

$$x^* = A\xi + \delta = \left( \begin{array}{c} v_1 \\ v_2 \end{array} \right),$$

where the elements of $v_1$ and $v_2$ are independent random variables. The elements of $v_1$ and $v_2$ are chosen to be standard normal deviates so that the generated $x^*$ is multivariate normal.

The data generation and analysis proceeds in steps as follows.

0. From $\Phi$ and $\Theta$, compute $T_1$ and $T_2$. From the $\pi_i^{(c)}$ given in Figure 3 compute $\tau_i^{(c)}$ in Equation 4 for $i = 1, 2, \ldots, p$ and $c = 1, 2, \ldots, m_i$.

1. Generate $v_1$ and $v_2$ and compute $x^*$. If $\tau^{(i)}_{c-1} < x_i^* < \tau^{(i)}_c$, set $x_i = c$, for $c = 1, 2, \ldots, m_i$ and $i = 1, 2, \ldots, p$. This gives a single observation of $x$. Repeat this $N$ times. This gives a random sample represented by a data matrix $X$ of order $N \times p$ where each element is an integer $1, 2, \ldots, m_i$ representing the observed category. In the simulation $m_i = m$ is the same for all variables.

2. From $X$, the matrix of polychoric correlations $R$ and the asymptotic covariance matrix $W$ are estimated using PRELIS (see Jöreskog, 1994). $R$ and $W$ are used to estimate the model parameters and their standard errors with ULS, DWLS, ML, and WLS. Note that the parameters and the standard errors are estimated with all methods using the same data $R$ and $W$. This gives maximum power to detect differences between methods if such exist. For each method compute all outcome variables (see next section).

3. Repeat Steps 1 and 2 $R$ times and collect all results.

OUTCOME VARIABLES

Nonconvergence

Nonconvergence is a real problem in simulation studies. This refers to the situation when the iterative process involved in the estimation of parameters does not converge, and therefore does not provide a correct solution that represents the minimum of the fit function.

In the analysis of a single sample a nonconvergence problem can be dealt with in various ways. For example, one can increase the number of iterations allowed, provide better starting values, or modify the model. In a simulation study with $R = 2,000$ replicates, however, these
alternatives are not available because it is impossible to intervene in the simulation process and give special treatment to those samples that do not converge. The nonconverged samples are not known until after the 2,000 replicates are finished.

There seem to be two approaches to deal this problem in the literature, both of which are unsatisfactory. One approach is to ignore the nonconverged samples. This gives rise to sample selection bias in estimates of bias and mean square error and other outcome variables. The other approach is to continue replications until 2,000 converged samples have been obtained. This approach also gives biased estimates of outcome variables and it gives no information about the frequency of occurrence of nonconvergence.

Nonconverged solutions only occur in small samples. Users of SEM should be aware of these problems and know how large sample size is needed to avoid these problems. In this study we report the frequency of occurrence of nonconverged solutions for the sample sizes $N = 100$ and $N = 200$ and for these sample sizes we do not report results on other outcome variables such as bias and mean square error. For the sample sizes $N = 400$, $N = 800$, and $N = 1,600$, problems of nonconvergence do not occur and the results on bias and mean square error and other outcome variables that we report are therefore unbiased.

Bias and Root Mean Square Error

Our primary interest was in the overall properties of the different estimation methods, and the individual parameter estimates were of secondary importance. Because of the large number of estimated factor loadings both within and across different conditions, we examined the average values of factor loadings to summarize our results to be more efficient. We considered three major outcome variables of interest: parameter estimates (including both factor loadings and factor correlations), standard errors, and chi-square test statistics. We examined the average relative bias or percentage bias of each outcome variables across all study conditions. Based on prior simulation studies (Curran, West, & Finch, 1996; Kaplan, 1989), we considered relative bias values less than 5% indicating a trivial bias, values between 5% and 10% indicating a moderate bias, and values greater than 10% indicating a substantial bias. Because the bias and mean square error of an individual parameter can depend on the size of the true parameter value we use the following estimates of bias and mean square error.

Let $\hat{\theta}_{ij}$ be the estimated parameter value of the $j$th parameter in the $i$th sample (replicate), $i = 1, 2, \ldots, R$, $j = 1, 2, \ldots, n$, (thus $n =$ number of parameters, $R =$ number of replicates), and let $\theta_j$ be the corresponding true parameter value.

Average relative bias (ARB).

$$ARB = 100/(1/R) \sum_i (1/n) \sum_j \left( \frac{\hat{\theta}_{ij} - \theta_j}{\theta_j} \right).$$  \hspace{1cm} (42)

Average root mean square error (AMSE).

$$\text{(AMSE)} = (1/R) \sum_i \sqrt{(1/n) \sum_j [(\hat{\theta}_{ij} - \theta_j)/\theta_j]^2}.$$  \hspace{1cm} (43)
TABLE 1
Model 1: Percentage of Samples with Nonconvergent Solutions

<table>
<thead>
<tr>
<th></th>
<th>Symmetric</th>
<th></th>
<th>Nonconvergent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ULS</td>
<td>DWLS</td>
<td>ML</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{21} = 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 100$</td>
<td>0.0</td>
<td>0.1</td>
<td>1.4</td>
</tr>
<tr>
<td>$N = 200$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$\lambda_{21} = 0.1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 100$</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>$N = 200$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$\lambda_{21} = 0.3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 100$</td>
<td>0.0</td>
<td>0.0</td>
<td>1.2</td>
</tr>
<tr>
<td>$N = 200$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$\lambda_{21} = 0.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 100$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>$N = 200$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Note. Number of categories = 2. ULS = unweighted least squares; DWLS = diagonally weighted least squares; ML = maximum likelihood; WLS = weighted least squares.

Bias and Root Mean Square Error for Standard Errors

Equations 42 and 43 can applied to the estimated standard errors $\hat{\sigma}_{ij}$ if $\theta_j$ is replaced by $\sigma_{ij}$, the standard deviation of the parameter estimates in the distribution of the $R = 2,000$ replicates.

RESULTS

Nonconvergence

The percentage of nonconverged solutions is given in Table 1 for Model 1 and in Table 2 for Model 2. Nonconverged samples occur only for sample sizes $N = 100$ and $N = 200$. Even for those sample sizes, nonconvergence is not a serious problem. The largest percentage of nonconvergent samples is 9.0 for Model 1 and 7.2 for Model 2. The largest percentage occurs for method ML for sample size $N = 100$. For other methods and for $N = 200$ the percentages are very small. The percentages seem to decrease with increasing number of categories and do not seem to increase with increasing degree of misspecification. If the number of categories is five or seven, there are no nonconvergent samples for Model 1 and at $N = 200$ for Model 2, regardless of the degree of misspecification. There are nonconverged samples observed at $N = 100$ for Model 2, but the nonconvergence rates are not larger than 2%.

We suspect that the reason for this is that the matrix of polychoric correlations is not positive definite as required for the ML fit function. At the time when the simulations were performed it was not possible distinguish this reason of nonconvergence from other reasons.
Properties of Parameter Estimates

The ARB for parameter estimates as a function of the degree of misspecification and different category distributions is shown in Figure 4 for sample size $N = 400$ for Model 1. It is seen that the bias increases almost linearly with increasing degree of misspecification. It is also seen that the bias is larger for WLS than for the other three methods that are very similar. The number of categories and their distributions do not seem to have any effect on bias. The corresponding figures for sample sizes $N = 800$ and $N = 1,600$, not shown here, look very similar, although the differences between methods seem to become smaller for the larger sample sizes.

The corresponding figure for Model 2 is shown in Figure 5, which exhibits the same characteristics as Figure 4, except that the distance between WLS and the other three methods seems to be even larger.

Figure 6 shows ARB as a function of sample size for Model 1 with no specification error. Again it is seen that the bias is much larger for WLS than for the other methods and that the relative bias converges to 0 as the sample size increases. Even with $N = 1,600$, however, there is considerable bias for WLS. Again the number of categories and their distributions do not seem to make a difference. We have similar figures, not shown here, for the cases $\lambda_{41} = 0.1, 0.3, 0.5$. They exhibit similar characteristics except that the biases are larger for the larger degrees of misspecification. The corresponding figures for Model 2 show similar characteristics.

The number and shape of the distribution of categories do not seem to make any difference. The reason for this is probably that these characteristics do not have an effect on the estimation of the polychoric correlations and their asymptotic covariance matrix and because these are the same for all methods, the methods are unaffected. In the following we therefore show figures only for the case of seven categories with a nonsymmetric distribution. For this case, we can show results for both models in the same figure as in Figure 7 where ARB is shown as a
function of degree of misspecification and sample size. This figure shows that ARB

- increases as the degree of misspecification increases.
- decreases as the sample size increases.
- increases from Model 1 to Model 2.
- is essentially unaffected by the number and shapes of category probabilities.
- is significantly larger for WLS than the other three methods, regardless of condition.

None of the methods underestimates the parameters on average. With no specification error ULS, DWLS, and ML are essentially unbiased, whereas WLS slightly overestimates the parameters on average. With all positive degree of misspecification, all methods overestimate parameters on average.

The previously shown figures give only estimates of average bias. From these figures it is clear that WLS is worse in terms of bias than the other three methods, which in turn look very
similar. Are the biases of ULS, DWLS, and ML equal? Because we have \( R = 2,000 \) replicates of ARB for each degree of misspecification and each combination of number and shape of categories, we estimated a multivariate analysis of covariance model to test for significant differences in the mean of ARB after controlling for the effects of sample size. Although there were no significant differences in several cases, there were a majority of cases where ULS was significantly better than DWLS and some cases where ULS was significantly better than ML.

Figure 8 shows AMSE as a function of sample size for Model 2 with no specification error and for different number and shapes of category probabilities. Figure 9 shows AMSE as a function of degree of misspecification for Model 2 estimated at \( N = 800 \), for different number and shapes of category probabilities. Similarly to ARB these figures show that AMSE

- increases with the degree of misspecification increases.
- decreases as the sample size increases.
- increases from Model 1 to Model 2.
Properties of Standard Errors

In structural equation modeling (SEM) one is not only interested in estimating the parameters of the model; one would also like to know how precise the estimates are. This is answered by the standard errors provided in most computer programs for SEM. The parameter estimate and standard error are usually transformed into a $t$ value or $z$ value to judge whether the parameter estimate is statistically significant.
The standard error depends on the model, on the method of estimation, and on the sample size. Here we also investigate how it depends on the degree of misspecification and the number and shape of the category probabilities.

As explained earlier, we use the standard deviation of the $R = 2,000$ estimates as the true standard error and compute ARB and AMSE in the same way as for parameter estimates. For the standard errors we use the notation ARBSE and average mean square error for standard errors (AMSESE).

Figure 11 shows ARBSE for Model 2 with no specification error as a function of sample size for the different category probabilities. It is seen that the standard errors are grossly underestimated with WLS at all sample sizes. The bias is about $-35\%$ at $N = 400$ and about $-10\%$ at $N = 1,600$. ML also underestimates standard errors at $N = 400$ but the bias is small and vanishes at larger sample sizes. The negative bias for ML seems to decrease with increasing numbers of categories. This is particularly noticeable at $N = 400$. ULS and DWLS seem to estimate the standard errors best with essentially no bias at any of the sample sizes investigated.
Figure 12 shows ARBSE for Model 2 estimated at $N = 800$ as a function of the degree of misspecification for the different category probabilities. It is seen that for ULS, DWLS, and ML, the bias in standard errors remains fairly constant as a function of degrees of misspecification, whereas the bias for WLS seem to get worse with increasing degree of misspecification. Thus, an important result is that the standard errors for ULS, DWLS, and ML are very good even with misspecified models. Furthermore there seems to be no effect of the number and shape of categories on the bias of standard errors.

Figure 13 shows AMSESE for both models as a function of sample size and degrees of misspecification for the case of seven nonsymmetric category distribution. Here one can see that the root mean square error in the standard errors increases with increasing degree of misspecification. For Model 1 the rate of increase is larger for ML and WLS than for ULS and DWLS. For Model 2 such rate of increase is only noticeable for WLS. For all methods the root mean square error decreases with increasing sample size.

FIGURE 8 Model 2 with no specification error: Average root mean square error (AMSE) as a function of sample size.
Model Fit Chi-Square

One important issue in SEM is the testing of model fit and the assessment of fit. In testing the model certain chi-squares are often used. Sometimes other measures of fit are also applied, but most of these depend on chi-square.

As explained earlier, under certain assumptions the chi-square should have a $\chi^2$-distribution with $d$ degrees of freedom if the model holds. Therefore, we investigate whether the $R = 2,000$ replicates of chi-squares follow a $\chi^2_d$ distribution when the model is correctly specified. Although this can be done more accurately, we focus on only two characteristics of the chi-square distribution, namely the mean, which should be $d$, and the proportion of times chi-square exceeds the 95th percentile of the $\chi^2_d$ distribution, which should be 0.05. Results are shown in Table 3 for Model 1 and in Table 4 for Model 2. These results are for the case of seven nonsymmetric categories.
FIGURE 10 Seven categories nonsymmetric: Average root mean square error (AMSE) as a function of degree of misspecification and sample size.

TABLE 3

<table>
<thead>
<tr>
<th></th>
<th>ULS</th>
<th>DWLS</th>
<th>ML</th>
<th>WLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>8.03</td>
<td>8.05</td>
<td>8.03</td>
<td>8.30</td>
</tr>
<tr>
<td>800</td>
<td>7.88</td>
<td>7.89</td>
<td>7.89</td>
<td>8.04</td>
</tr>
<tr>
<td>1,600</td>
<td>8.13</td>
<td>8.13</td>
<td>8.13</td>
<td>8.24</td>
</tr>
</tbody>
</table>

Estimated probability of p value ≤ 0.05

<table>
<thead>
<tr>
<th></th>
<th>ULS</th>
<th>DWLS</th>
<th>ML</th>
<th>WLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>0.049</td>
<td>0.049</td>
<td>0.049</td>
<td>0.048</td>
</tr>
<tr>
<td>800</td>
<td>0.053</td>
<td>0.053</td>
<td>0.053</td>
<td>0.053</td>
</tr>
<tr>
<td>1,600</td>
<td>0.046</td>
<td>0.046</td>
<td>0.045</td>
<td>0.043</td>
</tr>
</tbody>
</table>

Note. ULS = unweighted least squares; DWLS = diagonally weighted least squares; ML = maximum likelihood; WLS = weighted least squares.
FIGURE 11  Model 2 with no specification error: Average relative bias for standard errors (ARBSE) as a function of sample size.

TABLE 4  Model 2 With No Specification Error

<table>
<thead>
<tr>
<th>N</th>
<th>ULS</th>
<th>DWLS</th>
<th>ML</th>
<th>WLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average chi-square with 98 df</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>98.58</td>
<td>98.65</td>
<td>98.50</td>
<td>141.31</td>
</tr>
<tr>
<td>800</td>
<td>98.28</td>
<td>98.31</td>
<td>98.24</td>
<td>116.57</td>
</tr>
<tr>
<td>1,600</td>
<td>98.43</td>
<td>98.45</td>
<td>98.41</td>
<td>107.28</td>
</tr>
<tr>
<td>Estimated probability of ( p ) value ( \leq 0.05 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.041</td>
<td>0.040</td>
<td>0.042</td>
<td>0.001</td>
</tr>
<tr>
<td>800</td>
<td>0.042</td>
<td>0.041</td>
<td>0.040</td>
<td>0.004</td>
</tr>
<tr>
<td>1,600</td>
<td>0.054</td>
<td>0.054</td>
<td>0.054</td>
<td>0.019</td>
</tr>
</tbody>
</table>

*Note.* ULS = unweighted least squares; DWLS = diagonally weighted least squares; ML = maximum likelihood; WLS = weighted least squares.
FIGURE 12 Model 2 estimated at $N = 800$: Average relative bias for standard errors (ARBSE) as a function of degree of misspecification.

For Model 1 the mean of chi-square should be 8. It seems to be overestimated at $N = 1,600$ and underestimated at $N = 800$ for ULS, DWLS, and ML. $P$ values are slightly underestimated at $N = 400$ and more underestimated at $N = 1,600$ and overestimated at $N = 800$. For Model 1 there are no large differences between methods, although the large mean for WLS at $N = 400$ and $N = 1,600$ stands out as different.

For Model 2 the mean of chi-square should be 98. It is rather well estimated for ULS, DWLS, and ML at all sample sizes but highly overestimated for WLS at all sample sizes. The $p$ values are highly underestimated for WLS at all sample sizes. They are also underestimated for ULS, DWLS, and ML at $N = 400$ and $N = 800$ and slightly overestimated at $N = 1,600$.

From these results it is clear that the chi-square for WLS does not work well for large models. The reason for this is probably the use of $W^{-1}$ in the formula. This requires very large samples to be estimated accurately. For Model 2 this matrix is of the order $120 \times 120$. 
FIGURE 13 Seven categories nonsymmetric: AMSESE as a function of degree of misspecification and sample size.

The previous results are based on the case of seven nonsymmetric categories. To see if the number of categories and the shape of distribution have any effect on these results, consider Figure 14 showing the mean of chi-squares for Model 2 with no specification error as a function of sample size. It is seen that the mean remains fairly constant at about 98 for ULS, DWLS, and ML, whereas it is highly overestimated for WLS. These results were already clear from Table 4 but here we can also see that the number of categories and the shape of distribution have no effect on these results.

Figure 15 shows the mean of chi-square as a function of the degree of misspecification for Model 2 estimated at \( N = 400 \). If the model is misspecified we do not expect chi-square to have a \( \chi^2_d \) distribution. Figure 15 shows that the mean increases with increasing degree of misspecification.

Figure 16 shows the corresponding results for \( p \) values. If the model is misspecified one expects the \( p \) values to be less than .05 so that the model is rejected and one expects the \( p \)
FIGURE 14 Model 2 with no misspecification: Average chi-square as a function of sample size.

values to decrease as the degree of misspecification increases. This is the case in Figure 16, but there is an exception for the case of two categories where the \( p \) values increase slightly when the degree of misspecification goes from 0.0 to 0.1.

CONCLUSIONS

The focus of this study was the comparative behavior of estimation methods of ULS, DWLS, ML, and WLS and their performance on estimating parameters, standard errors, and chi-square and to study the effects of sample size, number of categories, and shape of distribution. In particular, we were interested in how well the methods work for correctly specified models and for different degrees of misspecification. We also studied two different models, one small and one large. In these regards, the results are clear that the ULS, DWLS, and ML outperform the WLS.
In sum, we make the following conclusions based on our experimental design and associated findings.

First, WLS performs poorly under all conditions compared to ULS, DWLS, and ML, although WLS performs better for the small model than for the large. Second, the number of categories and shape of distribution do not seem to matter. All methods work equally for two, five, and seven categories. Third, in general, the differences between the ULS, DWLS, and ML methods are small over all conditions. The striking result is the good performance of ULS.

ACKNOWLEDGMENT

The research reported in this article has been supported by the Swedish Research Council (VR) under the program Structural Equation Modeling With Ordinal Variables.
FIGURE 16 Model 2 estimated at $N = 400$. $P$ values as a function of degree of misspecification.

REFERENCES


APPENDIX

Assuming that the raw data on six ordinal variables are in the text file **ordata.raw**, one can use the following PRELIS syntax file to estimate the polychoric correlations and their asymptotic covariance matrix:

```plaintext
DA NI=6
RA=ORDATA.RAW
OR ALL
OU MA=PM PM=ORDATA.PM AC=ORDATA.ACP
```
Model 1 can then be estimated by RML using the following LISREL syntax file:

```plaintext
DA NI=6 MA=PM NO=400
LA
X1 X2 X3 X4 X5 X6
CM=ORDATA.PM
AC=ORDATA.ACP
MD NX=6 NK=2
FR LX(1,1) LX(2,1) LX(3,1) LX(4,2) LX(5,2) LX(6,2)
CO TD(1)=1-LX(1,1)**2
CO TD(2)=1-LX(2,1)**2
CO TD(3)=1-LX(3,1)**2
CO TD(4)=1-LX(4,2)**2
CO TD(5)=1-LX(5,2)**2
CO TD(6)=1-LX(6,2)**2
GU ME=ML
```

To obtain ULS or DWLS estimates, replace ML by ULS or DWLS in the last line.