PSYCHOMETRIKA—VOL. 71, NO. 1, 57–77
MARCH 2006
DOI: 10.1007/S11336-005-0773-4

LIMITED INFORMATION ESTIMATION AND TESTING OF DISCRETIZED MULTIVARIATE NORMAL STRUCTURAL MODELS

ALBERT MAYDEU-OLIVARES

UNIVERSITY OF BARCELONA

Discretized multivariate normal structural models are often estimated using multistage estimation procedures. The asymptotic properties of parameter estimates, standard errors, and tests of structural restrictions on thresholds and polychoric correlations are well known. It was not clear how to assess the overall discrepancy between the contingency table and the model for these estimators. It is shown that the overall discrepancy can be decomposed into a distributional discrepancy and a structural discrepancy. A test of the overall model specification is proposed, as well as a test of the distributional specification (i.e., discretized multivariate normality). Also, the small sample performance of overall, distributional, and structural tests, as well as of parameter estimates and standard errors is investigated under conditions of correct model specification and also under mild structural and/or distributional misspecification. It is found that relatively small samples are needed for parameter estimates, standard errors, and structural tests. Larger samples are needed for the distributional and overall tests. Furthermore, parameter estimates, standard errors, and structural tests are surprisingly robust to distributional misspecification.

Key words: GLS, WLS estimation, LISREL, categorical data analysis, data sparseness, goodness-of-fit, limited information estimation, pseudo-maximum likelihood estimation, IRT, polychoric correlations, structural equation models.

1. Introduction

A popular model for p-way contingency tables assumes that these arise by categorizing a p-dimensional multivariate standard normal density according to a set of thresholds. The thresholds and polychoric correlations may in turn be assumed to depend on a smaller set of structural parameters. Generally speaking, the estimation of such models is not possible by standard maximum likelihood estimation (e.g., Bock & Aitkin, 1981) due to the difficulty in evaluating high order multivariate normal integrals. However, these models can be easily estimated using the following three-stage limited information procedure:

- *Stage 1*: Estimate by maximum likelihood the thresholds for each variable separately from the univariate marginals of the contingency table.
- *Stage 2*: Estimate by maximum likelihood each of the polychoric correlations separately from the bivariate marginals of the contingency table given the estimated thresholds.
- *Stage 3*: If restrictions are imposed on the thresholds and polychoric correlations, estimate the underlying parameters from the estimated thresholds and polychoric correlations by a weighted least squares procedure.

This estimation method has a long tradition in psychometrics using both grouped and ungrouped data (i.e., sample proportions vs. individual observations). When the objective is to estimate the parameters of a discretized structured multivariate normal density, then it is computationally more efficient to estimate the model parameters using grouped data (Muthén, du

Requests for reprints should be sent to Albert Maydeu-Olivares, Faculty of Psychology, University of Barcelona, P. Valle de Hebrón, 171, 08035 Barcelona, Spain. E-mail: amaydeu@ub.edu.

This research was supported by the Department of Universities, Research and Information Society (DURSI) of the Catalan Government, and by grants BSO2000-0661 and BSO2003-08507 of the Spanish Ministry of Science and Technology.

Toit, & Spisic, 1997). However, when continuous exogenous variables are included in the model it is more convenient to resort to ungrouped data due to data sparseness (Muthén, 1982). The use of this estimation method using grouped data has been considered by Muthén (1978, 1993), Olsson (1979), Christoffersson and Gunsjö (1983, 1996), Gunsjö (1994), Jöreskog (1994), and Maydeu-Olivares (2001). Using ungrouped data it has been considered by Muthén (1984), Muthén and Satorra (1995), Muthén et al. (1997), Küsters (1987), and Bermann (1993). Furthermore, this estimation method is currently available in such popular software as PRELIS/LISREL (Jöreskog & Sörbom, 2001) and MPLUS (Muthén & Muthén, 2001) and also in the lesser known program MECOSA (Arminger, Wittenberg, & Schepers, 1996). Alternative sequential limited information estimators for these models have been proposed by other authors (e.g., Lee, Poon, & Bentler, 1995), but these will not be discussed here.

However, although this estimation method has been in use for several years now no satisfactory solution has been offered as to how to assess the goodness-of-fit of these models to the contingency table. See Muthén (1993) for a detailed discussion of this issue. Assessing the goodness-of-fit of discretized multivariate normal structural models involves assessing the overall discrepancy between the observed contingency table and the specified model. This overall discrepancy can be decomposed into a distributional discrepancy (i.e., the extent to which the data arise from discretizing a multivariate normal density) and a structural discrepancy (i.e., the extent to which the restrictions imposed on the parameters of the underlying normal density are appropriate). Tests for assessing the structural restrictions on the parameters of the discretized multivariate normal model are well known (Muthén, 1978, 1984, 1993) and routinely used in practice. However, these tests may only be meaningful if the distributional restrictions hold (i.e., if the data arise by categorizing a multivariate normal density). Yet, tests of the overall restrictions, or of the distributional restrictions, have not been proposed in the literature. The main aim of the present research is to fill this gap using asymptotic theory for sample proportions. In so doing, the literature on the use of this sequential procedure to estimate discretized multivariate normal structural models will also be reviewed and integrated.

The paper is organized as follows. In section 2 the sequential estimation procedure just described is presented. In section 3 the asymptotic distribution of the first, second, and third stage estimates are provided using standard results. In section 4 goodness-of-fit testing is discussed. In this section, tests of the distributional and of the overall restrictions imposed by the model on the bivariate marginals of the contingency table are proposed. Computational aspects of these tests are provided in section 5. In section 6 a small simulation study is reported to illustrate the small sample behavior of the parameter estimates, standard errors, and tests. Simulations are performed under correct model specification as well as under distributional and/or structural misspecification. Finally, section 7 includes two applications. In these applications factor models are fitted to the five-category items of the LOT (Scheier & Carver, 1985) and to the binary items of the LSAT 6 data (Bock & Lieberman, 1970). Additional material is provided as appendices. In one of the appendices it is shown that our expression for the asymptotic covariance matrix of the sample thresholds and polychoric correlations reduces to the expressions provided by Muthén (1978) for the binary case, by Olsson (1979) for the bivariate case, and by Christoffersson and Gunsjö (1983, 1996) and Jöreskog (1994) for the asymptotic covariance matrix of the polychoric correlations.

2. Sequential Estimation of Discretized Multivariate Normal Structural Models

Let $\mathbf{z}^* \sim N(\mathbf{0}, \mathbf{P})$ where $\mathbf{P}$ denotes a correlation matrix with elements $\boldsymbol{\rho}_{ii'}$. Suppose that each z_i^*, $i = 1, \ldots, p$, has been categorized as $y_i = k_i$ if $\tau_{i_k} < z_i^* < \tau_{i_{k+1}}$, $k_i = 0, \ldots, K - 1$, where $\tau_{i_0} = -\infty$, $\tau_{i_K} = \infty$. That is, for ease of exposition and without loss of generality, we shall assume that all observed categorical variables y_i have the same number of categories, K.

According to the model

$$Pr\left[\bigcap_{i=1}^{n}(y_i = k_i)\right] = \int \underset{\mathbf{R}}{\cdots} \int \phi_p(\mathbf{z}^*\colon \mathbf{0}, \mathbf{P})\, d\mathbf{z}^*, \tag{1}$$

where $\phi_p(\bullet)$ denotes a p-dimensional normal density function, and $\mathbf{R}$ is a p-dimensional area of integration with intervals $R_i = (\tau_{i_k}, \tau_{i_{k+1}})$. In particular,

$$\pi_{i_k} = \Pr(y_i = k_i) = \int_{\tau_{i_k}}^{\tau_{i_{k+1}}} \phi(z_i^* : 0, 1)\, dz_i^*, \tag{2}$$

$$\pi_{i_k i_{k'}} = \Pr[(y_i = k_i) \cap (y_i' = k_i')] = \int_{\tau_{i_k}}^{\tau_{i_{k+1}}} \int_{\tau_{i'_{k'}}}^{\tau_{i'_{k'+1}}} \phi_2(z_i^*, z_{i'}^* : 0, 0, 1, 1, \rho_{ii'})\, dz_i^* dz_{i'}^*. \tag{3}$$

We shall first introduce some notation: Let $\boldsymbol{\pi}_i = (\pi_{i_0}, \ldots, \pi_{i_{K-1}})'$ and let $\boldsymbol{\pi}_{ii'} = (\pi_{i_0 i'_0}, \pi_{i_0 i'_1}, \ldots, \pi_{i_{K-1} i'_{K-1}})'$. $\boldsymbol{\pi}_i$ is the set of univariate probabilities for variable i and $\pi_{ii'}$ is the set of bivariate probabilities for variables i and i'. Also, let $\dot{\boldsymbol{\pi}}_1 = (\boldsymbol{\pi}'_1, \ldots, \boldsymbol{\pi}'_p)'$ be a vector containing all p univariate probability tables, and let $\dot{\boldsymbol{\pi}}_2 = (\boldsymbol{\pi}_{21}', \boldsymbol{\pi}'_{31} \ldots, \boldsymbol{\pi}'_{p,p-1})'$ be a vector containing all the $p(p-1)/2$ bivariate probability tables. The sample counterparts of $\dot{\boldsymbol{\pi}}_1$ and $\dot{\boldsymbol{\pi}}_2$ will be denoted by $\dot{\mathbf{p}}_1$ and $\dot{\mathbf{p}}_2$. Finally, let $\boldsymbol{\tau}_i = (\tau_{i_1}, \ldots, \tau_{i_{K-1}})'$ be a vector containing all thresholds for variable i, $\boldsymbol{\tau} = (\boldsymbol{\tau}'_1, \ldots, \boldsymbol{\tau}'_p)'$, $\boldsymbol{\rho} = (\rho_{21}, \rho_{31}, \ldots, \rho_{p,p-1})'$, and $\boldsymbol{\kappa} = (\boldsymbol{\tau}', \boldsymbol{\rho}')'$.

Now, given a random sample of N observations from (1), we can place the observations in a K^p contingency table. We are interested in the following sequential procedure for estimating (1) from the contingency table.

First stage: Estimate the thresholds for each variable separately by maximizing

$$L(\boldsymbol{\tau}_i) = N \sum_{k=0}^{K-1} p_{i_k} \ln \pi_{i_k}(\tau_i), \tag{4}$$

where p_{i_k} denotes the sample counterpart of π_{i_k}.

Second stage: Given the first stage estimates, estimate separately each polychoric correlation $\rho_{ii'}$ by maximizing

$$L(\rho_{ii'}|\hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_{i'}) = N \sum_{k=0}^{K-1} \sum_{k'=0}^{K-1} p_{i_k i'_{k'}} \ln \pi_{i_k i'_{k'}}(\rho_{ii'}|\hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_{i'}.), \tag{5}$$

where $p_{i_k i'_{k'}}$ denotes the sample counterpart of $\pi_{i_k i'_{k'}}$.

Suppose now that some parametric structure is assumed on the reduced form parameters $\boldsymbol{\kappa}$, say $\boldsymbol{\kappa}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of q mathematically independent parameters. Then, these parameters can be estimated in an additional stage.

Third stage: Estimate $\boldsymbol{\theta}$ by minimizing the weighted least squares function

$$F = (\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\boldsymbol{\theta}))'\hat{\mathbf{W}}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\boldsymbol{\theta})), \tag{6}$$

where $\hat{\mathbf{W}}$ is a matrix converging in probability to $\mathbf{W}$, a positive definite matrix. Denoting the asymptotic covariance matrix of the sample thresholds and polychoric correlations by $\boldsymbol{\Xi}$, obvious choices of $\hat{\mathbf{W}}$ in (6) are $\hat{\mathbf{W}} = \hat{\boldsymbol{\Xi}}^{-1}$ (weighted least squares (WLS), Muthén, 1978),

$\hat{\mathbf{W}} = (\mathrm{Diag}(\hat{\boldsymbol{\Xi}}))^{-1}$ (diagonally weighted least squares (DWLS), Gunsjö, 1994; Muthén et al., 1997), and $\hat{\mathbf{W}} = \mathbf{I}$ (unweighted least squares (ULS), Muthén, 1993).

3. Asymptotic Distribution of the Estimates

First, we notice that since the univariate probabilities are simply sums of bivariate probabilities, $\dot{\boldsymbol{\pi}}_1 = \mathbf{T}\,\dot{\boldsymbol{\pi}}_2$, for some matrix $\mathbf{T}$ of 1's and 0's. Therefore, we can write

$$\sqrt{N}(\dot{\mathbf{p}}_1 - \dot{\boldsymbol{\pi}}_1) = \mathbf{T}\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2). \tag{7}$$

For example, for $p = 3$,

$$\begin{bmatrix} \boldsymbol{\pi}_1 \\ \boldsymbol{\pi}_2 \\ \boldsymbol{\pi}_3 \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{T}_2 & \mathbf{T}_2 & \mathbf{0} \\ \mathbf{T}_1 & \mathbf{0} & \mathbf{T}_2 \\ \mathbf{0} & \mathbf{T}_1 & \mathbf{T}_1 \end{bmatrix}}_{\mathbf{T}} \begin{bmatrix} \boldsymbol{\pi}_{21} \\ \boldsymbol{\pi}_{31} \\ \boldsymbol{\pi}_{32} \end{bmatrix},$$

where, letting $\mathbf{1}_K$ and $\mathbf{0}_K$ denote K-dimensional column vectors of 1's and 0's, respectively, we have that, when $K = 4$,

$$\mathbf{T}_1 = \begin{bmatrix} \mathbf{1}'_4 & \mathbf{0}'_4 & \mathbf{0}'_4 & \mathbf{0}'_4 \\ \mathbf{0}'_4 & \mathbf{1}'_4 & \mathbf{0}'_4 & \mathbf{0}'_4 \\ \mathbf{0}'_4 & \mathbf{0}'_4 & \mathbf{1}'_4 & \mathbf{0}'_4 \\ \mathbf{0}'_4 & \mathbf{0}'_4 & \mathbf{0}'_4 & \mathbf{1}'_4 \end{bmatrix}, \qquad \mathbf{T}_2 = \begin{pmatrix} \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 & \mathbf{I}_4 \end{pmatrix}.$$

We shall now provide the asymptotic properties of the first and second stage estimates. We first notice that $\hat{\boldsymbol{\tau}}_i$ is a maximum likelihood estimate, as (4) is the log-likelihood function for estimating $\boldsymbol{\tau}_i$ from a univariate marginal of the contingency table $\mathbf{p}_i$. Similarly, (5) is the log-likelihood function for estimating $\rho_{ii'}$ from a bivariate marginal of the contingency table $\mathbf{p}_{ii'}$ given the estimated thresholds. That is, $\hat{\rho}_{ii'}$ is a pseudo-maximum likelihood estimate in the terminology of Gong and Samaniego (1981). As a result, the asymptotic properties of these estimates can readily be obtained using standard results for maximum likelihood estimation for categorical models. Before proceeding, we shall review some of the relevant theory (see Agresti, 1990; Jöreskog, 1994).

Let $\boldsymbol{\pi}$ and $\mathbf{p}$ be C-dimensional vectors of multinomial probabilities, and sample proportions, respectively. Then,

$$\sqrt{N}(\mathbf{p} - \boldsymbol{\pi}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}), \qquad \boldsymbol{\Gamma} = \mathrm{Diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}', \tag{8}$$

where $\xrightarrow{d}$ denotes convergence in distribution. Also, consider a parametric structure for $\boldsymbol{\pi}$, $\boldsymbol{\pi}(\boldsymbol{\vartheta})$, with Jacobian matrix $\boldsymbol{\Delta} = \partial\boldsymbol{\pi}/\partial\boldsymbol{\vartheta}'$, and suppose we estimate $\boldsymbol{\vartheta}$ by maximizing $L(\boldsymbol{\vartheta}) = N\sum_{c=0}^{C-1} p_c \ln \pi_c(\boldsymbol{\vartheta})$. Then, under typical regularity conditions, it follows that $\hat{\boldsymbol{\vartheta}}$ is consistent and

$$\sqrt{N}(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \stackrel{a}{=} \mathbf{B}\sqrt{N}(\mathbf{p} - \boldsymbol{\pi}), \tag{9}$$

where $\mathbf{B} = (\boldsymbol{\Delta}'\mathbf{D}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{D}$, $\mathbf{D} = \mathrm{Diag}(\boldsymbol{\pi})^{-1}$, and $\stackrel{a}{=}$ denotes asymptotic equality.

Now, we apply (9) to the first stage estimates obtaining

$$\sqrt{N}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \stackrel{a}{=} \mathbf{B}_{11}\sqrt{N}(\dot{\mathbf{p}}_1 - \dot{\boldsymbol{\pi}}_1), \tag{10}$$

where $\mathbf{B}_{11} = (\boldsymbol{\Delta}'_{11}\mathbf{D}_1\boldsymbol{\Delta}_{11})^{-1}\boldsymbol{\Delta}'_{11}\mathbf{D}_1, \mathbf{D}_1 = \text{Diag}(\dot{\boldsymbol{\pi}}_1)^{-1}$, and $\boldsymbol{\Delta}_{11} = \partial\dot{\boldsymbol{\pi}}_1/\partial\boldsymbol{\tau}'$. Furthermore, from (7),

$$\sqrt{N}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \stackrel{a}{=} \mathbf{B}_{11}\mathbf{T}\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2). \tag{11}$$

Now, to apply (9) to the second stage estimates the asymptotic distribution of $\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \hat{\boldsymbol{\tau}}))$ is needed. It is shown in Appendix 1 that

$$\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \hat{\boldsymbol{\tau}})) \stackrel{a}{=} (\mathbf{I} - \boldsymbol{\Delta}_{21}\mathbf{B}_{11}\mathbf{T})\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2), \tag{12}$$

where $\boldsymbol{\Delta}_{21} = \partial\dot{\boldsymbol{\pi}}_2/\partial\boldsymbol{\tau}'$. Then, applying (9) to (12) we obtain

$$\sqrt{N}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) \stackrel{a}{=} \mathbf{B}_{22}(\mathbf{I} - \boldsymbol{\Delta}_{21}\mathbf{B}_{11}\mathbf{T})\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2), \tag{13}$$

where $\mathbf{B}_{22} = (\boldsymbol{\Delta}'_{22}\mathbf{D}_2\boldsymbol{\Delta}_{22})^{-1}\boldsymbol{\Delta}'_{22}\mathbf{D}_2$, $\mathbf{D}_2 = \text{Diag}(\dot{\boldsymbol{\pi}}_2)^{-1}$, and $\boldsymbol{\Delta}_{22} = \partial\dot{\boldsymbol{\pi}}_2/\partial\boldsymbol{\rho}'$. In Appendix 2 we sketch the derivatives involved in $\boldsymbol{\Delta}_{11}$, $\boldsymbol{\Delta}_{21}$, and $\boldsymbol{\Delta}_{22}$. Further details can be found in Olsson (1979).

Collecting (11) and (13), the first and second stage estimates can be expressed asymptotically as a linear function of the bivariate marginal proportions as follows:

$$\sqrt{N}\begin{bmatrix}\hat{\boldsymbol{\tau}} - \boldsymbol{\tau} \\ \hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\end{bmatrix} \stackrel{a}{=} \underbrace{\begin{bmatrix}\mathbf{B}_{11}\mathbf{T} \\ \mathbf{B}_{22}\left[\mathbf{I} - \boldsymbol{\Delta}_{21}\mathbf{B}_{11}\mathbf{T}\right]\end{bmatrix}}_{\mathbf{G}}\sqrt{N}\left(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2\right). \tag{14}$$

Now, since the marginal proportions $\dot{\mathbf{p}}_2$ are simply sums of multinomial cell proportions

$$\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2) \stackrel{d}{\to} N(\mathbf{0}, \dot{\boldsymbol{\Gamma}}), \qquad \dot{\boldsymbol{\Gamma}} = \tilde{\boldsymbol{\Gamma}} - \dot{\boldsymbol{\pi}}_2\dot{\boldsymbol{\pi}}_2, \tag{15}$$

where provided $p > 3$, the elements of $\tilde{\boldsymbol{\Gamma}}$ are four-way marginal probabilities. Thus, by (14) and (15),

$$\sqrt{N}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}) \stackrel{d}{\to} N(\mathbf{0}, \boldsymbol{\Xi}), \qquad \boldsymbol{\Xi} = \mathbf{G}\dot{\boldsymbol{\Gamma}}\mathbf{G}', \tag{16}$$

where $\mathbf{G}$ and $\dot{\boldsymbol{\Gamma}}$ are to be evaluated at the true parameter values. Also, partitioning $\mathbf{G} = \begin{bmatrix}\mathbf{G}_1 \\ \mathbf{G}_2\end{bmatrix}$ and $\boldsymbol{\Xi} = \begin{bmatrix}\boldsymbol{\Xi}_{11} & \boldsymbol{\Xi}'_{21} \\ \boldsymbol{\Xi}_{21} & \boldsymbol{\Xi}_{22}\end{bmatrix}$ according to the partitioning of $\boldsymbol{\kappa}$ we have that

$$\boldsymbol{\Xi}_{22} = N\text{Acov}(\hat{\boldsymbol{\rho}}) = \mathbf{G}_2\dot{\boldsymbol{\Gamma}}\mathbf{G}'_2, \tag{17}$$

where Acov(•) denotes asymptotic covariance matrix. In Appendix 3 it is shown that (17) equals the expression given by Jöreskog (1994), and it is also shown that (16) reduces to the expression given by Muthén, (1978) for the binary case ($K = 2$) and by Olsson (1979) for the bivariate case ($p = 2$).

The asymptotic properties of the third stage estimates can be obtained from (16) using standard results for weighted least squares estimators (e.g., Browne, 1984; Satorra, 1989; Satorra & Bentler, 1994). Letting $\mathbf{H} = \left(\tilde{\boldsymbol{\Delta}}'\mathbf{W}\tilde{\boldsymbol{\Delta}}\right)^{-1}\tilde{\boldsymbol{\Delta}}'\mathbf{W}$, where $\tilde{\boldsymbol{\Delta}} = \partial\boldsymbol{\kappa}/\partial\boldsymbol{\theta}'$, $\hat{\boldsymbol{\theta}}$ is consistent and

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \stackrel{a}{=} \mathbf{H}\sqrt{N}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}), \tag{18}$$

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \stackrel{d}{\to} N(\mathbf{0}, \mathbf{H}\boldsymbol{\Xi}\mathbf{H}'), \tag{19}$$

where $\tilde{\boldsymbol{\Delta}}$ and $\mathbf{W}$ are to be evaluated at the true parameter values. Now, when $\hat{\mathbf{W}} = \hat{\boldsymbol{\Xi}}^{-1}$, (19) simplifies to

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \stackrel{d}{\to} N(\mathbf{0}, (\tilde{\boldsymbol{\Delta}}'\boldsymbol{\Xi}^{-1}\tilde{\boldsymbol{\Delta}})^{-1}) \tag{20}$$

and we obtain an estimator that asymptotically has minimum variance among the class of estimators based on the first and second stage estimates.

In closing this section we note that throughout the presentation it is assumed that a multivariate standard normal density has been categorized according to a set of thresholds, where some parametric structure is imposed on the thresholds and polychoric correlations. When no restrictions are imposed on the thresholds, then some simplifications are available in the third estimation stage (see Muthén 1978, 1993). On the other hand, when a mean and covariance structure model has been discretized rather than a correlation structure model some complexities arise. These are discussed in Maydeu-Olivares and Hernández (2000).

4. Goodness-of-Fit Assessment

Within this estimation framework currently one tests the structural restrictions $\boldsymbol{\kappa}(\boldsymbol{\theta})$ using standard results for weighted least squares estimators. However, these tests may only be meaningful if the distributional restrictions hold (i.e., if the data arise by categorizing a multivariate normal density). For a detailed discussion of this issue, see Muthén (1993). Currently, the distributional restrictions $\dot{\boldsymbol{\pi}}_2(\boldsymbol{\kappa})$ are assessed piecewise by performing tests of bivariate normality for each pair of variables using the likelihood ratio statistic G^2. These tests are implemented, for instance, in PRELIS/LISREL (Jöreskog & Sörbom, 2001). However, it is not clear what to conclude if the hypothesis of categorized bivariate normality is accepted for some pairs of variables but rejected for others. To overcome this limitation a test of the joint distributional restrictions $\dot{\boldsymbol{\pi}}_2(\boldsymbol{\kappa})$ is proposed here. It is also possible to test the overall restrictions imposed by the model directly, $\dot{\boldsymbol{\pi}}_2(\boldsymbol{\theta})$ and we shall propose a test statistic to this purpose.

4.1. Goodness-of-Fit Testing of the Structural Restrictions

Consider the structural residuals $\mathbf{e}_s = \hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\hat{\boldsymbol{\theta}})$. Using standard results for weighted least squares estimators,

$$\sqrt{N}\mathbf{e}_s \stackrel{a}{=} (\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})\sqrt{N}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}), \tag{21}$$

$$\sqrt{N}\mathbf{e}_s \stackrel{d}{\to} N(\mathbf{0}, \mathbf{V}_s), \qquad \mathbf{V}_s = (\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})\boldsymbol{\Xi}(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})', \tag{22}$$

$$T_s := N\hat{F} = N\mathbf{e}_s'\hat{\mathbf{W}}\mathbf{e}_s \stackrel{a}{=} N(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa})'(\mathbf{W}(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H}))(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}) \stackrel{d}{\to} \sum_{i=1}^{r_s} \alpha_i \chi_1^2, \tag{23}$$

where $r_s = p(K-1) + p(p-1)/2 - q$ are the degrees of freedom available for testing the structural restrictions $\boldsymbol{\kappa}(\boldsymbol{\theta})$.

In (23) the χ_1^2's are independent chi-square variables with one degree of freedom and the $\alpha_i's$ are the nonnull eigenvalues of

$$\mathbf{M}_s = \mathbf{W}(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})\boldsymbol{\Xi}. \tag{24}$$

When $\hat{\mathbf{W}} = \hat{\boldsymbol{\Xi}}^{-1}$, (23) simplifies to $T_s \stackrel{d}{\to} \chi^2_{r_s}$. On the other hand, when $\hat{\mathbf{W}} = (\text{Diag}(\hat{\boldsymbol{\Xi}}))^{-1}$ or $\hat{\mathbf{W}} = \mathbf{I}$, a goodness-of-fit of the model can be obtained following Satorra and Bentler (1994) by scaling T_s by its mean or adjusting it by its mean and variance so that it approximates a chi-square distribution as follows (Muthén, 1993; Muthén et al., 1997)

$$\bar{T}_s = \frac{T_s}{\text{Tr}(\mathbf{M}_s)/r_s}, \qquad \bar{\bar{T}}_s = \frac{T_s}{\text{Tr}(\mathbf{M}_s^2)/r_s}, \tag{25}$$

where $\bar{T}_s$ and $\bar{\bar{T}}_s$ denote the scaled (for mean) and adjusted (for mean and variance) test statistics. The former is referred to a chi-square distribution with r_s degrees of freedom, whereas the latter is referred to a chi-square distribution with $d_s = \mathrm{Tr}(\mathbf{M}_s)^2/(\mathrm{Tr}\mathbf{M}_s^2/r_s)$ degrees of freedom.

4.2. Goodness-of-Fit Testing of the Distributional Restrictions

Consider now the distributional residuals $\mathbf{e}_d = \dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\kappa}})$. Let $\boldsymbol{\Delta} = \partial\dot{\boldsymbol{\pi}}_2/\partial\boldsymbol{\kappa}' = (\boldsymbol{\Delta}_{21}|\boldsymbol{\Delta}_{22})$, in Appendix 1 it is shown that

$$\sqrt{N}\mathbf{e}_d \xrightarrow{d} (\mathbf{I} - \boldsymbol{\Delta}\mathbf{G})\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2). \tag{26}$$

The asymptotic distribution of the distributional residuals then follows from (15) and (26),

$$\sqrt{N}\mathbf{e}_d \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_d), \qquad \mathbf{V}_d = (\mathbf{I} - \boldsymbol{\Delta}\mathbf{G})\dot{\boldsymbol{\Gamma}}(\mathbf{I} - \boldsymbol{\Delta}\mathbf{G})'. \tag{27}$$

Now, to test the distributional restrictions of the model $\dot{\boldsymbol{\pi}}_2(\boldsymbol{\kappa})$ we propose using the test statistic

$$T_d := N\mathbf{e}_d'\mathbf{e}_d \xrightarrow{d} \sum_{i=1}^{r_d} \alpha_i \chi_1^2, \tag{28}$$

where, by Theorem 2.1 of Box (1954), the α_i's are now the nonnull eigenvalues of $\mathbf{V}_d$ and the number of degrees of freedom available for testing is $r_d = (K^2 - 2K)[p(p-1)/2]$. Goodness-of-fit tests of the distributional restrictions imposed by the model can be obtained by scaling T_d by its mean or adjusting it by its mean and variance so that it approximates a chi-square distribution as follows:

$$\bar{T}_d = \frac{T_d}{\mathrm{Tr}(\mathbf{M}_d)/r_d}, \qquad \bar{\bar{T}}_d = \frac{T_d}{\mathrm{Tr}\left(\mathbf{M}_d^2\right)/r_d}, \tag{29}$$

where $\bar{T}_d$ and $\bar{\bar{T}}_d$ denote the scaled (for mean) and adjusted (for mean and variance) test statistics. The former is referred to a chi-square distribution with r_d degrees of freedom, whereas the latter is referred to a chi-square distribution with $d_d = \mathrm{Tr}(\mathbf{M}_d)^2/(\mathrm{Tr}\mathbf{M}_d^2/r_d)$ degrees of freedom.

4.3. Goodness-of-Fit Testing of the Overall Restrictions

Consider now the overall residuals $\mathbf{e}_o = \dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\theta}})$. In Appendix 1 it is shown that

$$\sqrt{N}\mathbf{e}_o \xrightarrow{d} (\mathbf{I} - \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G})\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2). \tag{30}$$

From (15) and (30) we immediately have

$$\sqrt{N}\mathbf{e}_o \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_o), \qquad \mathbf{V}_o = (\mathbf{I} - \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G})\dot{\boldsymbol{\Gamma}}(\mathbf{I} - \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G})'. \tag{31}$$

Akin to (28), to test the overall restrictions of the model, $\dot{\boldsymbol{\pi}}_2(\boldsymbol{\theta})$, we propose using the test statistic

$$T_o := N\mathbf{e}_o'\mathbf{e}_o \xrightarrow{d} \sum_{i=1}^{r_o} \alpha_i \chi_1^2, \tag{32}$$

where the α_i's are now the nonnull eigenvalues of $\mathbf{V}_o$ and the number of degrees of freedom available for testing is $r_o = n(K-1) + (K-1)^2[p(p-1)/2] - q$. Goodness-of-fit tests of the overall restrictions imposed by the model can be obtained by scaling T_o by its mean or adjusting it by its mean and variance so that it approximates a chi-square distribution as follows:

$$\bar{T}_o = \frac{T_o}{\mathrm{Tr}\,(\mathbf{M}_o)/r_o}, \qquad \bar{\bar{T}}_o = \frac{T_o}{\mathrm{Tr}\left(\mathbf{M}_o^2\right)/r_o}, \tag{33}$$

where $\bar{T}_o$ and $\bar{\bar{T}}_o$ denote the scaled (for mean) and adjusted (for mean and variance) test statistics. The former is referred to a chi-square distribution with r_o degrees of freedom, whereas the latter is referred to a chi-square distribution with $d_o = \mathrm{Tr}(\mathbf{M}_o)^2/(\mathrm{Tr}\mathbf{M}_o^2/r_o)$ degrees of freedom.

In closing this section we note that the overall residuals $\mathbf{e}_o$ can be decomposed, asymptotically, as a linear function of the distributional residuals $\mathbf{e}_d$ and of the structural residuals $\mathbf{e}_s$,

$$\mathbf{e}_o \overset{a}{=} \mathbf{e}_d + \boldsymbol{\Delta}\mathbf{e}_s. \tag{34}$$

This is shown in Appendix 1. In Appendix 1 it is also shown that

$$\mathrm{Acov}(T_d, T_s) = 2\mathrm{Tr}[(\mathbf{I} - \boldsymbol{\Delta}\mathbf{G})'(\mathbf{I} - \boldsymbol{\Delta}\mathbf{G})\dot{\boldsymbol{\Gamma}}\,\mathbf{G}'(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})'\mathbf{W}(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})\mathbf{G}\dot{\boldsymbol{\Gamma}}\,], \tag{35}$$

$$\mathrm{Acov}(T_o, T_s) = 2\mathrm{Tr}[(\mathbf{I} - \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G})'(\mathbf{I} - \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G})\dot{\boldsymbol{\Gamma}}\,\mathbf{G}'(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})'\mathbf{W}(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})\mathbf{G}\dot{\boldsymbol{\Gamma}}], \tag{36}$$

$$\mathrm{Acov}(T_d, T_s) = 2\mathrm{Tr}[(\mathbf{I} - \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G})'(\mathbf{I} - \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G})\dot{\boldsymbol{\Gamma}}\,(\mathbf{I} - \boldsymbol{\Delta}\mathbf{G})'(\mathbf{I} - \boldsymbol{\Delta}\mathbf{G})\dot{\boldsymbol{\Gamma}}\,]. \tag{37}$$

Thus, when the model holds the overall, distributional, and structural test statistics are asymptotically correlated because of their common dependency on the asymptotic covariance matrix of the bivariate proportions.

5. Computational Aspects

The asymptotic covariance matrix of the bivariate marginal proportions $\dot{\mathbf{p}}_2$,$\dot{\boldsymbol{\Gamma}}$, is of dimension $K^2[p(p-1)/2]$. Clearly, the size of this matrix grows very rapidly for increasing p and K. Thus, it is important to consider how to compute the asymptotic covariance matrix of the sample thresholds and polychoric correlations and the traces required for the proposed distributional and overall goodness-of-fit tests without having to store into memory $\dot{\boldsymbol{\Gamma}}$. In this section, we show how to estimate the elements of the asymptotic covariance matrix of the sample thresholds and polychoric correlations efficiently for very large models and how to obtain tests of the distributional restrictions as a by-product with very little additional computation. The approach employed here relies heavily on Jöreskog (1994). Additional research is needed to manage the computation of the overall tests within available computer memory for large models.

5.1. Asymptotic Covariance Matrix of Sample Thresholds and Polychoric Correlations

Akin to (10) we have

$$\sqrt{N}(\hat{\boldsymbol{\tau}}_i - \boldsymbol{\tau}_i) \overset{a}{=} \mathbf{B}_{11}^{(i)}\sqrt{N}(\dot{\mathbf{p}}_i - \dot{\boldsymbol{\pi}}_i), \tag{38}$$

where $\mathbf{B}_{11}^{(i)} = (\boldsymbol{\Delta}_{11}^{(i)'}\mathbf{D}_i\boldsymbol{\Delta}_{11}^{(i)})^{-1}\boldsymbol{\Delta}_{11}^{(i)'}\mathbf{D}_i$, $\mathbf{D}_i = \mathrm{Diag}(\dot{\boldsymbol{\pi}}_i)^{-1}$, and $\boldsymbol{\Delta}_{11}^{(i)} = \partial\dot{\boldsymbol{\pi}}_i/\partial\boldsymbol{\tau}_i'$. Also, akin to (13), we have

$$\sqrt{N}(\hat{\rho}_{ii'} - \rho_{ii'}) \overset{a}{=} \mathbf{G}_2^{(ii')}\sqrt{N}(\dot{\mathbf{p}}_{ii'} - \dot{\boldsymbol{\pi}}_{ii'}), \tag{39}$$

$$\mathbf{G}_2^{(ii')} = \mathbf{B}_{22}^{(ii')}\big(\mathbf{I} - \boldsymbol{\Delta}_{21}^{(i)}\mathbf{B}_{11}^{(i)}\mathbf{T}_1 - \boldsymbol{\Delta}_{21}^{(i')}\mathbf{B}_{11}^{(i')}\mathbf{T}_2\big), \tag{40}$$

where $\mathbf{B}_{22}^{(ii')} = (\boldsymbol{\Delta}_{22}^{(ii')'}\mathbf{D}_{ii'}\boldsymbol{\Delta}_{22}^{(ii')})^{-1}\boldsymbol{\Delta}_{22}^{(ii')'}\mathbf{D}_{ii'}$, $\mathbf{D}_{ii'} = \mathrm{Diag}(\dot{\boldsymbol{\pi}}_{ii'})^{-1}$, $\boldsymbol{\Delta}_{22}^{(ii')} = \partial\dot{\boldsymbol{\pi}}_{ii'}/\partial\rho_{ii'}$, and $\boldsymbol{\Delta}_{21}^{(i)} = \partial\dot{\boldsymbol{\pi}}_{ii'}/\partial\boldsymbol{\tau}_i'$.

Then, letting (i, i') be any two variables (not necessarily distinct), the asymptotic variances and covariances among the estimated thresholds can be obtained using

$$N\mathrm{Acov}(\hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_{i'}) = \mathbf{B}_{11}^{(i)}(\mathbf{C}_{ii'} - \boldsymbol{\pi}_i\boldsymbol{\pi}_{i'})\mathbf{B}_{11}^{(i')'}, \tag{41}$$

where $\mathbf{C}_{ii'}$ is a $K \times K$ table of bivariate probabilities. Similarly, letting (i, i', j) be any three variables such that $i \neq i'$, the asymptotic covariances between the estimated thresholds and polychoric correlations can be obtained using

$$N\text{Acov}(\hat{\rho}_{ii'}, \hat{\boldsymbol{\tau}}_j) = \mathbf{G}_2^{(ii')}(\mathbf{C}_{ii'j} - \boldsymbol{\pi}_{ii'}\boldsymbol{\pi}_j)\mathbf{B}_{11}^{(j)'}, \tag{42}$$

where $\mathbf{C}_{ii'j}$ is a $K^2 \times K$ table of trivariate probabilities. Finally, letting (i, i', j, j') be any four variables such that $i \neq i'$ and $j \neq j'$, the asymptotic variances and covariances between the estimated polychoric correlations can be obtained using

$$N\text{Acov}(\hat{\rho}_{ii'}, \hat{\rho}_{jj'}) = \mathbf{G}_2^{(ii')}(\mathbf{C}_{ii'jj'} - \boldsymbol{\pi}_{ii'}\boldsymbol{\pi}_{jj'})\mathbf{G}_2^{(jj')'}, \tag{43}$$

where $\mathbf{C}_{ii'jj'}$ is a $K^2 \times K^2$ table of four-way probabilities.

Note that the two- and three-way probability tables can be obtained from the four-way probability tables by using $\mathbf{T}_1$ and $\mathbf{T}_2$ matrices as needed. Also, in (41) to (43) it is possible to use the following simplification: Since $\boldsymbol{\Delta}_{11}'\mathbf{D}_1\dot{\boldsymbol{\pi}}_1 = \mathbf{0}$, $\boldsymbol{\Delta}_{11}'\mathbf{D}_1\mathbf{T}\dot{\boldsymbol{\pi}}_2 = \mathbf{0}$ and hence $\mathbf{B}_{11}\mathbf{T}\dot{\boldsymbol{\pi}}_2 = \mathbf{0}$. Similarly, $\mathbf{B}_{22}\dot{\boldsymbol{\pi}}_2 = \mathbf{0}$. Hence,

$$\mathbf{G}\dot{\boldsymbol{\pi}}_2 = \mathbf{0}, \tag{44}$$

and $\boldsymbol{\Xi} = \mathbf{G}\tilde{\boldsymbol{\Gamma}}\mathbf{G}'$, $\boldsymbol{\Xi}_{22} = \mathbf{G}_2\tilde{\boldsymbol{\Gamma}}\mathbf{G}_2'$. Thus, for instance, the term $-\boldsymbol{\pi}_{ii'}\boldsymbol{\pi}_{jj'}$ can be dropped from (43).

To compute $\hat{\boldsymbol{\Xi}}$ we store into memory all estimated $(K-1) \times K\mathbf{B}_{11}^{(i)}$ matrices, and all $1 \times K$ vectors $\mathbf{G}_{22}^{(ii')}$. We consistently estimate $\mathbf{B}_{11}^{(i)}$ and $\mathbf{G}_{22}^{(ii')}$ by evaluating all derivative matrices and all univariate and bivariate probabilities at $\hat{\boldsymbol{\kappa}}$. Also, we consistently estimate the four-way probability tables by using four-way sample proportions. The four-way contingency tables need not be stored in memory. We compute them one at a time from the raw data. By using these consistent estimates our asymptotic covariance matrix for the polychoric correlations equals Jöreskog's (1994) as implemented in PRELIS/LISREL (Jöreskog & Sörbom, 2001).

5.2. *Tests of the Distributional Restrictions Imposed by the Model*

Akin to (26) we have

$$\sqrt{N}(\dot{\mathbf{p}}_{ii'} - \dot{\boldsymbol{\pi}}_{ii'}(\hat{\boldsymbol{\kappa}})) \stackrel{a}{=} \left(\mathbf{I} - \boldsymbol{\Delta}^{(ii')}\mathbf{G}^{(ii')}\right)\sqrt{N}(\dot{\mathbf{p}}_{ii'} - \dot{\boldsymbol{\pi}}_{ii'}), \tag{45}$$

where

$$\boldsymbol{\Delta}^{(ii')} = \left(\boldsymbol{\Delta}_{21}^{(ii')}|\boldsymbol{\Delta}_{22}^{(ii')}\right), \quad \mathbf{G}^{(ii')} = \begin{bmatrix}\mathbf{G}_1^{(ii')}\\ \mathbf{G}_2^{(ii')}\end{bmatrix}., \quad \boldsymbol{\Delta}_{21}^{(ii')} = \left(\boldsymbol{\Delta}_{21}^{(i)}|\boldsymbol{\Delta}_{21}^{(i')}\right), \quad \text{and} \quad \mathbf{G}_1^{(ii')} = \begin{bmatrix}\mathbf{B}_{11}^{(i)}\mathbf{T}_1\\ \mathbf{B}_{11}^{(i')}\mathbf{T}_2\end{bmatrix}.$$

Now, to obtain $\bar{T}_d$ and $\bar{\bar{T}}_d$, we need $\text{Tr}(\mathbf{V}_d)$ and $\text{Tr}(\mathbf{V}_d^2)$ where $\mathbf{V}_d$ is a symmetric matrix structured in blocks, each of dimension $K^2 \times K^2$. These blocks can be obtained akin to (27) using (45) as

$$\mathbf{V}_d^{(ll')} = (\mathbf{I} - \boldsymbol{\Delta}^{(l)}\mathbf{G}^{(l)})(\mathbf{C}_{ll'} - \boldsymbol{\pi}_l\boldsymbol{\pi}_{l'})(\mathbf{I} - \boldsymbol{\Delta}^{(l')}\mathbf{G}^{(l')})', \tag{46}$$

where, to simplify the notation, we let $l := (i, j)$; $i = 2, \ldots, p$; $j = 1, \cdots, i-1$. Then,

$$\text{Tr}(\mathbf{V}_d) = \sum_l \text{Tr}\left(\mathbf{V}_d^{(l)}\right), \qquad \text{Tr}(\mathbf{V}_d^2) = \sum_l \text{Tr}\left(\mathbf{V}_d^{(ll)^2}\right) + \sum_{l \neq l'} 2\text{Tr}\left(\mathbf{V}_d^{(ll')'}\mathbf{V}_d^{(ll')}\right), \tag{47}$$

where (46) is consistently estimated by evaluating all derivative matrices and univariate and bivariate probabilities at $\hat{\boldsymbol{\kappa}}$, and by estimating the four-way probability tables by using four-way sample proportions. Very little additional computation is involved to obtain these tests and in our implementation we compute them in a single loop while obtaining the asymptotic covariance matrix of the estimated thresholds and polychoric correlations.

6. Small Sample Behavior

The small sample behavior of the parameter estimates, standard errors, and tests of the structural restrictions has been investigated, for instance, by Muthén (1993) and Muthén et al. (1997) under conditions of correct model specification. These simulation studies reveal that the asymptotically optimal WLS has a poorer small sample behavior than ULS or DWLS due to the instability of the four-way proportions in small samples (Muthén, 1993). Furthermore, when ULS or DWLS is employed no weight matrix needs to be inverted in the third stage of the estimation process. Thus, larger models can be handled by ULS and DWLS than by WLS. In the binary case, Maydeu-Olivares (2001) has shown that the small sample behavior of ULS and DWLS is very similar.

In this section we investigate the small sample behavior of the parameter estimates, standard errors, and goodness-of-fit tests by means of a small simulation study, not only under conditions of correct model specification, but also under conditions of structural or/and distributional misspecification. A standard multivariate normal (MVN) density with $p = 12$ variables was considered, where each variable is categorized into $K = 3$ categories using the thresholds $\boldsymbol{\tau}_i = (-0.5, 0.5)'$. A three factor correlation structure model was assumed, $\mathbf{P}_{z^*} = \text{Off}(\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}')$, with

$$\boldsymbol{\Lambda} = \begin{bmatrix} 0.7 & 0.6 & 0.5 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.7 & 0.6 & 0.5 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0.6 & 0.5 & 0.4 \end{bmatrix}',$$

$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & 0.3 & 0.4 \\ 0.3 & 1 & 0.5 \\ 0.4 & 0.5 & 1 \end{bmatrix}.$$

ULS was used in the third stage of the estimation procedure. Two sample sizes, $N = 200$ and $N = 1000$, and four conditions were considered:

1) *Correctly specified model*. Data was generated from the model just described.
2) *Distributionally misspecified model*. Data was generated as in 1) except that a multivariate t (MVT) distribution was used instead of an MVN distribution. Two degrees of distributional misspecification were used: small (MVT with 10 df) and mild (MVT with 1 df).
3) *Structurally misspecified model*. Data was generated as in 1) but a one factor model was estimated.
4) *Distributionally and structurally misspecified model*. Data was generated as in 2) but a one factor model was estimated. Only small distributional misspecifications were considered in this case.

To illustrate the degree of distributional misspecification a standard normal density is plotted in Figure 1 against a t distribution with 1 and 10 degrees of freedom. As can be seen in this figure the difference between a $N(0, 1)$ and a $t(10)$ density is rather small. Even in the case of the $t(1)$ density, the misspecification is not severe as it is still a symmetric distribution.

6.1. Goodness-of-Fit Tests

We shall first examine the small sample behavior of the goodness-of-fit tests. The mean and variance of the tests statistics under consideration, as well as empirical rejection rates in the critical region $\alpha = \{1\%, 5\%, 10\%, 20\%\}$, are provided in Tables 1 to 3.

Table 1 contains the results for the tests of the structural restrictions. As can be seen in this table, when the structural model is correctly specified, reasonably accurate Type I errors

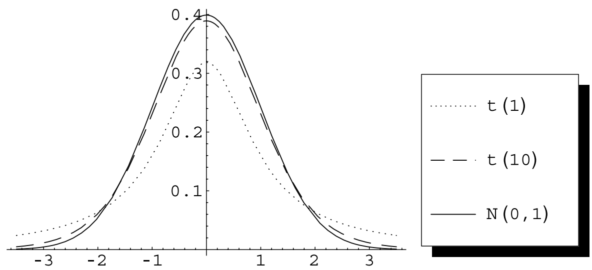

FIGURE 1.
Univariate density plots.

can be obtained with a small sample size ($N = 200$) even in the presence of small distributional misspecifications. With mild distributional misspecifications, MVT(1), the p-values of the structural tests are slightly too small. On the other hand, when the structural model is misspecified, the structural tests are very powerful. Almost invariably they reject the model even with a small sample size.

Table 2 contains the results for the tests of the distributional restrictions. In this case, when the data arise from a categorized MVN density, the Type I errors are accurate only with the mean and variance adjusted statistic and the larger sample size ($N = 1000$). For the smaller sample size ($N = 200$) reasonably accurate Type I errors can be obtained by the heuristic procedure of averaging the p-values of $\bar{T}$ and $\bar{\bar{T}}$. The empirical rejection rates for the heuristic procedure at $\alpha = \{1\%, 5\%, 10\%, 20\%\}$ are $\{0.5, 4.5, 10.5, 22.0\}$. On the other hand, the distributional tests are reasonably powerful in detecting distributionally misspecified models. When the data arise from an MVT(1)—a mild distributional misspecification—the tests proposed here reject the MVN distributional model every time even when $N = 200$. Even when the data arise from a MVT(10)—a very small distributional misspecification—the distributional tests have some power although only with large sample sizes.

Finally, Table 3 contains the results for the tests of the overall restrictions. In this case when the model is correctly specified the Type I errors are accurate only with the mean and variance adjusted statistic and the larger sample size. With the smaller sample size the mean scaled statistic $\bar{T}$ over-rejects the model and the mean and variance statistic $\bar{\bar{T}}$ under-rejects the model. Again, the heuristic procedure of averaging the p-values of $\bar{T}$ and $\bar{\bar{T}}$ can be used to obtain reasonably accurate Type I errors. In this case, the empirical rejection rates for the heuristic procedure at $\alpha = \{1\%, 5\%, 10\%, 20\%\}$ are $\{0.6, 4.1, 10.2, 22.1\}$. Now, when the structural model is correctly specified, the overall tests have reasonable power to detect minor distributional misspecifications, and are very powerful to detect mild distributional misspecifications (they reject the model every time). Also, when the structural model is misspecified, the tests reject the model every time even when there is no distributional misspecification. With the smaller sample size, the overall tests retain some power to detect structural and distributional misspecification. In particular, the overall tests reject the model more often when there is structural and distributional misspecification than when there is only structural misspecification or only distributional misspecification.

6.2. Parameter Estimates and Standard Errors

In Table 4 we provide a summary of the parameter estimates and standard errors obtained in conditions 1) and 2) above. More specifically, results are shown for correctly specified structural models when data arise by categorizing an MVN distribution, an MVT(10) distribution, and an MVT(1) distribution.

TABLE 1.
Simulation results across 1000 replications: Goodness-of-fit tests of the structural restrictions.

Structure	Correctly specified												Misspecified							
distribution	Multivariate normal				Multivariate t (10 df)				Multivariate t (1 df)				Multivariate normal				Multivariate) t (10 df)			
N	200		1000		200		1000		200		1000		200		1000		200		1000	
Stat.	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$
mean	52.1	36.6	51.0	42.9	51.3	34.84	51.9	43.6	55.4	37.09	54.8	46.52	134.5	88.5	458.9	377.0	132.86	86.7	451.5	370.6
variance	129.6	57.4	106.5	74.3	113.2	50.54	114.2	78.72	144.6	58.0	120.8	85.69	718.0	234.2	2986	1573	726.21	230.5	2976	1532
Reject. rates																				
1%	2.5	0.8	1.0	0.4	1.3	0.8	1.3	0.8	4.0	1.9	2.9	2.2	99.1	98.3	100.0	100.0	98.3	97.0	100.0	100.0
5%	8.3	4.9	5.2	4.1	6.1	3.7	6.6	5.2	13.6	8.5	9.6	7.3	99.6	99.4	100.0	100.0	99.7	99.3	100.0	100.0
10%	13.8	9.6	10.0	8.4	9.8	6.6	13.3	11.8	21.8	15.3	18.5	16.5	99.6	99.6	100.0	100.0	99.8	99.7	100.0	100.0
20%	24.0	21.2	20.1	18.6	20.4	15.9	25.7	24.5	35.7	30.1	32.5	29.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: $df = 51$ and 54 for the correctly specified and misspecified models, respectively; $\bar{T}$ and $\bar{\bar{T}}$ denote the mean scaled and mean and variance adjusted statistics, respectively.

TABLE 2.
Simulation results across 1000 replications: Goodness-of-fit tests of the distributional restrictions.

Distribution	Multivariate normal				Multivariate t (10 df)				Multivariate t (1 df)			
N	200		1000		200		1000		200		1000	
Stat.	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$
mean	201.1	93.7	199.7	148.9	207.8	96.3	244.0	180.1	706.6	247.9	2716	1399
variance	489.2	100.0	469.6	250.2	533.1	104.2	818.4	423.8	17009	1227	77688	10879
Reject. rates												
1%	2.5	0.2	1.7	0.7	3.6	0.9	43.6	32.8	100.0	100.0	100.0	100.0
5%	9.4	2.4	8.0	4.8	15.8	3.0	66.2	57.6	100.0	100.0	100.0	100.0
10%	16.1	6.7	13.8	10.3	24.6	11.2	75.4	71.1	100.0	100.0	100.0	100.0
20%	25.6	17.7	24.9	21.7	36.1	25.7	85.2	82.5	100.0	100.0	100.0	100.0

Notes: df = 198; $\bar{T}$ and $\bar{\bar{T}}$ denote the mean scaled and mean and variance adjusted statistics, respectively.

As can be seen in this table, when the distributional assumptions hold, a sample size of 200 observations suffices to obtain accurate parameter estimates as there is no consistent bias in the parameter estimates. The median absolute relative bias is less than 1%. Also, 200 observations suffice to obtain accurate standard errors because, although we observe a consistent downward bias, the median absolute relative bias is 4.7% and it does not exceed 7% in any case. Of course, when $N = 1000$, we obtain more accurate parameter estimates and standard errors. In this case there is no consistent bias either in the parameter estimates or in the standard errors, and the relative bias of the standard errors does not exceed 5% (median = 1.7%). Here we define parameter and standard error relative bias as $(\bar{x}_{\hat{\theta}} - \theta_0)/\theta_0$ and $(\bar{x}_{SE(\hat{\theta})} - sd_{\hat{\theta}})/sd_{\hat{\theta}}$, respectively.

What is most remarkable is how robust is the sequential estimator to small and mild distributional misspecification, particularly in large sample sizes. When the data arise by categorizing an MVT(10) the median absolute relative bias of the parameter estimates and standard errors is only 1.2 and 5.1% when $N = 200$ and less than 1 and 1.7% when $N = 1000$. Even when the data arise by categorizing an MVT(1) the median absolute relative bias of the parameter estimates and standard errors is 1.6 and 8.7% when $N = 200$ and 1 and 4% when $N = 1000$. In fact, the results in terms of parameter estimates and standard errors obtained with an MVT(1) and $N = 1000$ are not much worse than those obtained with an MVN distribution and $N = 200$. In closing, we note that whenever the median absolute bias was larger than 2% the relative bias was consistently negative for parameter estimates and standard errors (underestimation) and positive for factor correlations (overestimation). Whenever the median was smaller than 2% no consistent trend in the bias was observed.

7. Applications

We now provide two applications where a covariance structure is assumed. In the first example, the variables consist of five categories, in the second example the data is binary.

7.1. Life Orientation Test

The Life Orientation Test (LOT) (Scheier & Carver, 1985), is an eight-item questionnaire designed to measure optimism and pessimism where each item consists of five categories. Chang, D'Zurilla, and Maydeu-Olivares (1994) fitted the following covariance structure model to this

TABLE 3.
Simulation results across 1000 replications: Goodness-of-fit tests of the overall restrictions.

Structure	Correctly specified												Misspecified							
distribution	Multivariate normal				Multivariate t (10 df)				Multivariate t (1 df)				Multivariate normal				Multivariate t (10 df)			
N	200		1000		200		1000		200		1000		200		1000		200		1000	
Stat.	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$	$\bar{T}$	$\bar{\bar{T}}$
mean	252.9	103.6	250.8	176.0	260.2	106.6	299.5	209.7	759.5	261.7	2751	1492	312.4	126.7	543.4	379.1	323.6	131.2	607.1	422.5
variance	643.7	95.3	614.7	288.3	642.8	96.6	966.1	449.0	17629	1328	78149	13413	948.0	135.1	2100	950.4	1061	151.7	2810	1248.3
Reject. rates																				
1%	2.7	<0.1	1.8	0.5	4.3	0.4	41.9	28.4	100.0	100.0	100.0	100.0	54.9	18.1	100.0	100.0	68.3	27.9	100.0	100.0
5%	9.9	1.7	8.1	4.5	16.1	2.2	64.9	55.7	100.0	100.0	100.0	100.0	76.6	48.8	100.0	100.0	84.0	62.3	100.0	100.0
10%	16.4	4.9	14.5	9.6	22.8	8.8	74.5	68.0	100.0	100.0	100.0	100.0	84.6	66.1	100.0	100.0	90.2	79.0	100.0	100.0
20%	26.1	16.4	25.3	21.1	36.7	23.2	86.2	82.4	100.0	100.0	100.0	100.0	92.5	84.6	100.0	100.0	96.1	90.2	100.0	100.0

Notes: df = 249 and 252 for the correctly specified and misspecified models, respectively; $\bar{T}$ and $\bar{\bar{T}}$ denote the mean scaled and mean and variance adjusted statistics, respectively.

TABLE 4.
Simulation results across 1000 replications: Parameter estimates and standard errors.

		Multivariate normal distribution						Multivariate t distribution (10 df)						Multivariate t distribution (1 df)					
		$N = 200$			$N = 1000$			$N = 200$			$N = 1000$			$N = 200$			$N = 1000$		
par	true	$\bar{x}$ est	$\bar{x}$ SE	sd SE	$\bar{x}$ est	$\bar{x}$ SE	sd SE	$\bar{x}$ est	$\bar{x}$ SE	sd SE	$\bar{x}$ est	$\bar{x}$ SE	sd SE	$\bar{x}$ est	$\bar{x}$ SE	sd SE	$\bar{x}$ est	$\bar{x}$ SE	sd SE
$\lambda_{1,1}$	0.7	0.70	0.09	0.10	0.70	0.04	0.04	0.69	0.10	0.11	0.70	0.04	0.04	0.69	0.10	0.11	0.69	0.05	0.05
$\lambda_{2,1}$	0.6	0.60	0.09	0.10	0.60	0.04	0.04	0.59	0.09	0.10	0.60	0.04	0.04	0.59	0.10	0.11	0.60	0.05	0.05
$\lambda_{3,1}$	0.5	0.50	0.09	0.09	0.50	0.04	0.04	0.50	0.09	0.10	0.50	0.04	0.04	0.50	0.10	0.11	0.49	0.05	0.05
$\lambda_{4,1}$	0.4	0.40	0.10	0.10	0.40	0.04	0.04	0.40	0.10	0.11	0.40	0.04	0.04	0.39	0.10	0.12	0.40	0.05	0.05
$\lambda_{5,2}$	0.7	0.69	0.09	0.09	0.70	0.04	0.04	0.69	0.09	0.09	0.70	0.04	0.04	0.68	0.10	0.10	0.69	0.04	0.05
$\lambda_{6,2}$	0.6	0.60	0.09	0.09	0.60	0.04	0.04	0.60	0.09	0.09	0.60	0.04	0.04	0.60	0.10	0.10	0.59	0.04	0.05
$\lambda_{7,2}$	0.5	0.50	0.09	0.09	0.50	0.04	0.04	0.49	0.09	0.10	0.50	0.04	0.04	0.50	0.10	0.11	0.49	0.04	0.05
$\lambda_{8,2}$	0.4	0.40	0.09	0.10	0.40	0.04	0.04	0.39	0.10	0.10	0.40	0.04	0.04	0.40	0.10	0.11	0.39	0.05	0.05
$\lambda_{9,3}$	0.7	0.69	0.09	0.09	0.70	0.04	0.04	0.69	0.09	0.10	0.70	0.04	0.04	0.69	0.09	0.10	0.69	0.04	0.05
$\lambda_{10,3}$	0.6	0.60	0.09	0.09	0.60	0.04	0.04	0.60	0.09	0.09	0.60	0.04	0.04	0.59	0.09	0.10	0.59	0.04	0.04
$\lambda_{11,3}$	0.5	0.49	0.09	0.09	0.50	0.04	0.04	0.49	0.09	0.10	0.50	0.04	0.04	0.49	0.10	0.11	0.50	0.04	0.05
$\lambda_{12,3}$	0.4	0.40	0.09	0.10	0.40	0.04	0.04	0.40	0.09	0.10	0.40	0.04	0.04	0.39	0.10	0.11	0.40	0.04	0.05
$\phi_{2,1}$	0.3	0.31	0.11	0.12	0.30	0.05	0.05	0.31	0.12	0.12	0.30	0.05	0.05	0.31	0.12	0.13	0.30	0.05	0.06
$\phi_{3,1}$	0.4	0.41	0.11	0.12	0.40	0.05	0.05	0.41	0.11	0.11	0.40	0.05	0.05	0.41	0.12	0.13	0.41	0.05	0.06
$\phi_{3,2}$	0.5	0.50	0.11	0.11	0.50	0.05	0.05	0.51	0.11	0.11	0.50	0.05	0.05	0.52	0.11	0.12	0.50	0.05	0.05

questionnaire: $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a diagonal matrix,

$$\boldsymbol{\Lambda}' = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{41} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \lambda_{52} & \cdots & \lambda_{82} \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{\Psi} = \begin{bmatrix} 1 & \psi_{21} \\ \psi_{21} & 1 \end{bmatrix}.$$

The clusters correspond to the positively and to the negatively worded items of the questionnaire, respectively. That is, the factors measure optimism and pessimism, respectively. Since this covariance structure is scale invariant and no restrictions are imposed on the thresholds, $\boldsymbol{\theta}'=(\lambda_{11}, \ldots, \lambda_{82}, \psi_{21})$ can be estimated in the third stage by minimizing a discrepancy function of the polychoric correlations only where, for identification purposes, $\boldsymbol{\Theta} = \mathbf{I} - \text{Diag}(\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}')$, see Maydeu-Olivares and Hernández (2000).

Chang et al. (1994) used WLS and found that this model reproduced well the polychoric matrix. We shall reanalyze their data here which consists of 389 observations. Using ULS in the third stage we find that the model reproduces well the polychoric matrix $\bar{T}_s = 25.4$ on 14 df, $p = .15$ and $\bar{\bar{T}}_s = 15.4$ on 11.5 df, $p = .19$. However, using the standard procedure of testing categorized bivariate normality for each pair of variables using a likelihoood ratio statistic, G^2, we find that for 15 out of 28 pairs of variables the null hypotheses of categorized bivariate normality is rejected at $\alpha = .01$. Thus, it is not clear what to conclude about the distributional assumptions of the model, nor about its overall fit.

However, our tests of the distributional assumptions reveal that the hypothesis of joint categorized multivariate normality is to be rejected: $\bar{T}_d = 1070.9$ on 420 df, $p < .01$ and $\bar{\bar{T}}_d = 252.1$ on 98.9 df, $p < .01$. Not surprisingly, overall, the model fails to fit the bivariate tables: $\bar{T}_o = 1112.1$ on 439 df, $p < 0.01$ and $\bar{\bar{T}}_o = 253.8$ on 100.2 df, $p < .01$. Thus, the model does not reproduce well the contingency table because the distributional restrictions do not hold.

7.2. LSAT 6 Data

These data, consisting of 1000 observations on five binary variables, were originally reported in Bock and Lieberman (1970). The data have been reanalyzed repeatedly in the literature using a variety of full and limited information methods (see McDonald & Mok, 1995). A one factor model fits well the 2^5 contingency table. Bock and Lieberman (1970) report a likelihood ratio statistic $G^2 = 21.28$ on 21 df, $p = .44$, and we computed Pearson's statistic using their parameter estimates obtaining $X^2 = 18.03$, $p = .65$.

We fitted a one factor model to these data using ULS in the third stage. The structural tests yielded $\bar{T}_s = 4.67$ on 5 df, $p = .46$ and $\bar{\bar{T}}_s = 4.31$ on 4.6 df, $p = .45$, so the model fits the tetrachoric correlations well. Now, when all the variables are binary, it is not possible to perform the proposed tests of categorized normality as there are no degrees of freedom available for testing. A test of trivariate dichotomized normality has been proposed by Muthén and Hofacker (1988). However, it is not clear what to conclude if the hypothesis of dichotomized normality is rejected for some but not all triplets. To overcome this limitation one can perform a test of the overall restrictions on the bivariate marginals. We obtained $\bar{T}_o = 3.95$ on 5 df, $p = .56$ and $\bar{\bar{T}}_o = 3.55$ on 4.50 df, $p = .56$, which is similar to Bock and Lieberman's results.

8. Conclusions

We have presented a unified framework for the sequential estimation of discretized multivariate normal structural models and their testing using asymptotic theory for sample proportions. In particular, we have proposed tests for the distributional as well as for the overall restrictions imposed by these models on the bivariate margins of the contingency table. Also, we have shown how the overall restrictions imposed by the model on the bivariate margins can be decomposed asymptotically as a linear function of the distributional and the structural restrictions.

The proposed tests are simply mean and mean and variance corrections to the asymptotic distribution of a test statistic consisting of the sum of squared distributional and overall residuals. We have shown that the proposed distributional tests can be computed very efficiently for very large models. On the other hand, further research is needed to compute efficiently the proposed overall tests as n and K become large. As an alternative to the moment corrected statistics proposed here, one could consider the use of a generalized Wald test (Moore, 1977), that is, a quadratic form using a generalized inverse of a consistent estimate of the asymptotic covariance matrix of the distributional or overall residuals as weight matrix. Such statistics would be asymptotically chi-square distributed but they require more computation than the statistics proposed here, and they are likely to suffer from poor small sample performance (see Satorra & Bentler, 1994).

Using a categorized multivariate t framework we have investigated the behavior of the parameter estimates, standard errors, and goodness-of-fit tests under conditions of correct model specification as well as distributional and/or structural misspecification. Our results suggest that good parameter estimates, standard errors, and empirical Type I errors for the structural tests can be obtained with reasonably small sample sizes when the model is correctly specified. Larger sample sizes may be needed to obtain reasonable Type I errors for the distributional and overall tests. This was expected, as the number of degrees of freedom in these tests is very large. Also, the behavior of the structural, distributional, and overall tests under distributional and/or structural misspecification was found to be adequate. Only small samples are needed to correctly reject mildly distributionally misspecified models. Larger samples are needed to detect small distributional misspecifications.

Furthermore, the parameter estimates, standard errors, and structural tests can be very robust in situations of correct structural specification but small and mild in distributional misspecification. The results of the simulation study suggest that meaningful inferences about the dimensional structure of categorical data can be drawn even when the distributional (and overall) tests reject the model. Clearly, further studies are needed to investigate exactly under what type of distributional misspecification these procedures are robust. Also, a test of the joint distributional assumptions, when all the observed variables are dichotomous, is needed.

We have not considered in this paper structured MVN models in which some but not all the variables are categorized. Neither have we considered multivariate ordinal probit models where one assumes categorized multivariate normality conditional on a set of exogonous variables. Estimation and structural inferences for these models have been considered by Muthén (1984), Muthén and Satorra (1995), Muthén et al. (1997), Küsters (1987), and Bermann (1993). It is not clear how one can test the distributional assumptions in these complex situations. Clearly, more work is also needed in this area

Appendix 1. Proofs of Key Results

Proof of Equation (12):

A first-order expansion of $\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \hat{\boldsymbol{\tau}})$ around $\boldsymbol{\tau} = \boldsymbol{\tau}_0$ yields $\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \hat{\boldsymbol{\tau}}) \stackrel{a}{=} \dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \boldsymbol{\tau}) + \boldsymbol{\Delta}_{21}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})$, where $\boldsymbol{\Delta}_{21} = \partial \dot{\boldsymbol{\pi}}_2 / \partial \boldsymbol{\tau}'$. Thus, $\sqrt{N}(\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \hat{\boldsymbol{\tau}}) - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \boldsymbol{\tau})) \stackrel{a}{=} \boldsymbol{\Delta}_{21}\sqrt{N}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})$. Now, by (11), $\sqrt{N}(\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \hat{\boldsymbol{\tau}}) - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \boldsymbol{\tau})) \stackrel{a}{=} \boldsymbol{\Delta}_{21}\mathbf{B}_{11}\mathbf{T}\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \tau))$. Equation (12) follows by noting that

$$\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \hat{\boldsymbol{\tau}})) = \sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \tau)) - \sqrt{N}(\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \hat{\boldsymbol{\tau}}) - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho}, \tau)). \qquad \square$$

Proof of Equation (26):

A first-order expansion of $\dot{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\kappa}})$ around $\boldsymbol{\kappa} = \boldsymbol{\kappa}_0$ yields $\dot{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\kappa}}) \stackrel{a}{=} \dot{\boldsymbol{\pi}}_2(\boldsymbol{\kappa}) + \boldsymbol{\Delta}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa})$, where $\boldsymbol{\Delta} = \partial\dot{\boldsymbol{\pi}}_2/\partial\boldsymbol{\kappa}' = (\,\boldsymbol{\Delta}_{21} \mid \boldsymbol{\Delta}_{22}\,)$. Coupling this with (14), $\sqrt{N}(\dot{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\kappa}}) - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\kappa})) \stackrel{a}{=} \boldsymbol{\Delta}\mathbf{G}\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2)$ Equation (26) follows by noting that

$$\sqrt{N}\mathbf{e}_d := \sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\kappa}})) = \sqrt{N}(\,\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\kappa})) - \sqrt{N}(\dot{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\kappa}}) - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\kappa})). \qquad \square$$

Proof of Equation 30:

A first-order expansion of $\dot{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ yields $\dot{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\theta}}) \stackrel{a}{=} \dot{\boldsymbol{\pi}}_2(\boldsymbol{\theta}) + \partial\dot{\boldsymbol{\pi}}_2/\partial\boldsymbol{\theta}'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, where $\partial\dot{\boldsymbol{\pi}}_2/\partial\boldsymbol{\theta}' = \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}$. Now, again using (14), $\sqrt{N}\,(\dot{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\theta}}) - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\theta})) \stackrel{a}{=} \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{G}\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2)$. Equation (30) follows by noting that

$$\sqrt{N}\mathbf{e}_o := \sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\theta}})) = \sqrt{N}(\,\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\theta})) - \sqrt{N}(\dot{\boldsymbol{\pi}}_2(\hat{\boldsymbol{\theta}}) - \dot{\boldsymbol{\pi}}_2(\boldsymbol{\theta})). \qquad \square$$

Proof of Equation (34):

By (21) and (14),

$$\sqrt{N}\mathbf{e}_s \stackrel{a}{=} (\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})\mathbf{G}\sqrt{N}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2). \tag{48}$$

Now, from (30), $\mathbf{e}_o \stackrel{a}{=} (\mathbf{I} - \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G})(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2)$. Thus, $\mathbf{e}_o \stackrel{a}{=} (\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2) + \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2)$. Now, adding and subtracting $\boldsymbol{\Delta}\mathbf{G}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2)$ to this equation and rearranging terms, $\mathbf{e}_o \stackrel{a}{=} (\mathbf{I} - \boldsymbol{\Delta}\mathbf{G})(\dot{\boldsymbol{\pi}}_2 - \dot{\mathbf{p}}_2) + \boldsymbol{\Delta}(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})\mathbf{G}(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2)$, and (34) follows immediately from (26) and (48). $\square$

Proof of Equations (35), (36) and (37):

Let $\mathbf{e} := (\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2)$ and $\mathbf{A} = (\mathbf{I} - \boldsymbol{\Delta}\mathbf{G})'(\mathbf{I} - \boldsymbol{\Delta}\mathbf{G})$. Then, from (26), $T_d \stackrel{a}{=} N\mathbf{e}'\mathbf{A}\mathbf{e}$. Also, from (48) $\mathbf{W}^{1/2}\sqrt{N}\mathbf{e}_s \stackrel{a}{=} \mathbf{W}^{1/2}(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})\mathbf{G}\sqrt{N}\mathbf{e}$. Then, letting $\mathbf{B} = \mathbf{G}'(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})'\mathbf{W}(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H})\mathbf{G}$, $T_s \stackrel{a}{=} N\mathbf{e}'\mathbf{B}\mathbf{e}$. Finally, letting $\mathbf{C} = (\mathbf{I} - \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G})'(\mathbf{I} - \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G})$, $T_o \stackrel{a}{=} N\mathbf{e}'\mathbf{C}\mathbf{e}$. Equations (35), (36), and (37) then follow from Theorem 3.2d.4 in Mathai and Provost (1992). $\square$

Appendix 2. Derivatives Involved in $\boldsymbol{\Delta}_{11}$, $\boldsymbol{\Delta}_{21}$, and $\boldsymbol{\Delta}_{22}$

To obtain the derivatives involved in $\boldsymbol{\Delta}_{11} = \partial\boldsymbol{\pi}_1/\partial\boldsymbol{\tau}'$ we note that (2) can be rewritten as

$$\pi_{i_k} = \Phi_1(\tau_{i_{k+1}}) - \Phi_1(\tau_{i_k}), \tag{49}$$

where $\Phi_p(\bullet)$ denotes a p-variate standard normal distribution function and, since $\tau_{i_0} = -\infty$, $\tau_{i_K} = \infty$,

$$\Phi_1(\tau_{i_0}) = 0, \qquad \Phi_1(\tau_{i_K}) = 1. \tag{50}$$

Then, the elements of $\boldsymbol{\Delta}_{11}$ can be obtained using (49) and (50) with

$$\frac{\partial\Phi_1\left(\tau_{i_k}\right)}{\partial\tau_{i_k}} = \phi_1(\tau_{i_k}). \tag{51}$$

We also note that because of (49), (4) has a closed form solution

$$\hat{\boldsymbol{\tau}}_{i_k} = \Phi^{-1}\left(\sum_{c=1}^{k} p_{i_c}\right), \qquad k = 1, ..., K - 1.$$

Now, to obtain the derivatives involved in $\mathbf{\Delta}_{21} = \partial \ddot{\boldsymbol{\pi}}_2 / \partial \boldsymbol{\tau}'$ and $\mathbf{\Delta}_{22} = \partial \ddot{\boldsymbol{\pi}}_2 / \partial \boldsymbol{\rho}'$ we first note that (3) can be rewritten as

$$\pi_{i_k i'_{k'}} = \Phi_2\left(\tau_{i_{k+1}}, \tau_{i'_{k'+1}}, \rho_{ii'}\right) - \Phi_2\left(\tau_{i_k}, \tau_{i'_{k'+1}}, \rho_{ii'}\right) - \Phi_2\left(\tau_{i_{k+1}}, \tau_{i'_{k'}}, \rho_{ii'}\right) + \Phi_2\left(\tau_{i_k}, \tau_{i'_{k'}}, \rho_{ii'}\right) \quad (52)$$

(Olsson, 1979, Equation 4), where $\Phi_2(\bullet)$ is a bivariate standard normal distribution function with parameter $\rho_{ii'}$. Again, since $\tau_{i_0} = -\infty$, $\tau_{i_K} = \infty$,

$$\begin{aligned} &\Phi_2\left(\tau_{i_k}, \tau_{i'_K}, \rho_{ii'}\right) = \Phi_1\left(\tau_{i_k}\right), \quad &&\Phi_2\left(\tau_{i_K}, \tau_{i'_K}, \rho_{ii'}\right) = 1, \quad &&\Phi_2\left(\tau_{i_0}, \tau_{i'_{k'}}, \rho_{ii'}\right) = 0, \\ &\Phi_2\left(\tau_{i_K}, \tau_{i'_{k'}}, \rho_{ii'}\right) = \Phi_1\left(\tau_{i'_k}\right), &&\Phi_2\left(\tau_{i_0}, \tau_{i'_0}, \rho_{ii'}\right) = 0, &&\Phi_2\left(\tau_{i_k}, \tau_{i'_0}, \rho_{ii'}\right) = 0. \end{aligned} \quad (53)$$

Then, the elements of $\mathbf{\Delta}_{21}$ can be obtained using (50) through (53), and

$$\frac{\partial \Phi_2\left(\tau_{i_k}, \tau_{i'_{k'}}, \rho_{ii'}\right)}{\partial \tau_{i_k}} = \phi_1\left(\tau_{i_k}\right) \Phi_1\left(\frac{\tau_{i'_{k'}} - \rho_{ii'}\tau_{i_k}}{\sqrt{1 - \rho_{ii'}^2}}\right) \quad (54)$$

(Olsson, 1979, Equation 12). Finally, the elements of $\mathbf{\Delta}_{22}$ can also be obtained using (50) through (53), and

$$\frac{\partial \Phi_2\left(\tau_{i_k}, \tau_{i'_{k'}}, \rho_{ii'}\right)}{\partial \rho_{ii'}} = \phi_2\left(\tau_{i_k}, \tau_{i'_{k'}} : \rho_{ii'}\right) \quad (55)$$

(Muthén, 1978, Equation 18), a bivariate standard normal density function with parameter $\rho_{ii'}$ evaluated at $\left(\tau_{i_k}, \tau_{i'_{k'}}\right)$.

Appendix 3. The Asymptotic Covariance Matrix of Sample Thresholds and Polychoric Correlations in Some Special Cases

We shall first show that when $K = 2$ the expression of the asymptotic covariance of sample thresholds and tetrachoric correlations (16) reduces to that given by Muthén (1978). First we note that for each pair of categorical variables $(y_i, y_{i'})$ there are three mathematically independent probabilities, say $\ddot{\pi}_i = \Pr(y_i = 1)$, $\ddot{\pi}_{i'} = \Pr(y_{i'} = 1)$, and $\ddot{\pi}_{ii'} = \Pr[(y_i = 1) \cap \Pr(y_{i'} = 1)]$. Let $\ddot{\boldsymbol{\pi}}_1 = (\ddot{\pi}_1, \ldots, \ddot{\pi}_n)'$, $\ddot{\boldsymbol{\pi}}_2 = (\ddot{\pi}_{21}, \ddot{\pi}_{31}, \ldots, \ddot{\pi}_{nn-1})'$, and $\ddot{\boldsymbol{\pi}} = (\ddot{\pi}_1, \ddot{\pi}_2)'$, with sample counterparts $\ddot{\mathbf{p}} = (\ddot{\mathbf{p}}_1, \ddot{\mathbf{p}}_2)'$. Muthén (1978) estimates each threshold and tetrachoric correlation separately using

$$\hat{\tau}_i = -\Phi_1^{-1}\left(\ddot{p}_i\right), \quad (56)$$

$$\hat{\rho}_{ii'} = \Phi_2^{-1}\left(\ddot{p}_{ii'} \mid -\hat{\tau}_i, -\hat{\tau}_{i'}\right), \quad (57)$$

where $\Phi_1(\bullet)$ and $\Phi_2(\bullet)$ denote univariate and bivariate standard normal distribution functions. Since the relationship between $(\tau_i, \tau_{i'}, \rho_{ii'})$ and $(\ddot{\pi}_i, \ddot{\pi}_{i'}, \ddot{\pi}_{ii'})$ is one-to-one, using (56) and (57) is equivalent to employing (4) and (5) (Hamdan, 1970). Now,

$$\dot{\boldsymbol{\pi}}_2 = \mathbf{c} + \mathbf{C}\ddot{\boldsymbol{\pi}} = \left(\mathbf{C}_1 \mid \mathbf{C}_2\right)\begin{pmatrix}\ddot{\boldsymbol{\pi}}_1 \\ \ddot{\boldsymbol{\pi}}_2\end{pmatrix}, \quad (58)$$

illustrated here, for $n = 2$,

$$\begin{pmatrix}\pi_{00} \\ \pi_{01} \\ \pi_{10} \\ \pi_{11}\end{pmatrix} = \begin{pmatrix}1 \\ 0 \\ 0 \\ 0\end{pmatrix} + \left(\begin{array}{rr|r} -1 & -1 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \\ 0 & 0 & 1 \end{array}\right)\begin{pmatrix}\ddot{\pi}_i \\ \ddot{\pi}_{i'} \\ \ddot{\pi}_{ii'}\end{pmatrix} = \begin{pmatrix}1 - \ddot{\pi}_i - \ddot{\pi}_{i'} + \ddot{\pi}_{ii'} \\ \ddot{\pi}_{i'} - \ddot{\pi}_{ii'} \\ \ddot{\pi}_i - \ddot{\pi}_{ii'} \\ \ddot{\pi}_{ii'}\end{pmatrix}.$$

Now, by (14) and (58), $\sqrt{N}\,(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}) \overset{a}{=} \mathbf{GC}\sqrt{N}\,(\ddot{\mathbf{p}} - \ddot{\boldsymbol{\pi}})$. Furthermore, it is easy to verify that

$$\mathbf{GC} = \begin{pmatrix} \left(\dfrac{\partial \ddot{\boldsymbol{\pi}}_1}{\partial \boldsymbol{\tau}'}\right)^{-1} & \mathbf{0} \\ -\left(\dfrac{\partial \ddot{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\rho}'}\right)^{-1} \dfrac{\partial \ddot{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\tau}'} \dfrac{\partial \ddot{\boldsymbol{\pi}}_1}{\partial \boldsymbol{\tau}'} & \left(\dfrac{\partial \ddot{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\rho}'}\right)^{-1} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \ddot{\boldsymbol{\pi}}_1}{\partial \boldsymbol{\tau}'} & \mathbf{0} \\ \dfrac{\partial \ddot{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\tau}'} & \dfrac{\partial \ddot{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\rho}'} \end{pmatrix}^{-1} = \left(\dfrac{\partial \ddot{\boldsymbol{\pi}}}{\partial \boldsymbol{\kappa}'}\right)^{-1} = \ddot{\mathbf{G}}^{-1}. \tag{59}$$

Hence, in the binary case (16) reduces to Muthén's (1978) expression for the covariance matrix of the sample thresholds and tetrachoric correlations

$$\boldsymbol{\Xi} = \ddot{\mathbf{G}}^{-1}\ddot{\boldsymbol{\Gamma}}\ddot{\mathbf{G}}^{-1\prime} \tag{60}$$

where $\ddot{\boldsymbol{\Gamma}}$ denotes the covariance matrix of $\sqrt{N}\,(\ddot{\mathbf{p}} - \ddot{\boldsymbol{\pi}})$.

Christoffersson and Gunsjö (1983) and Jöreskog (1994) have provided expressions for the asymptotic covariance matrix of the sample polychoric correlations which are algebraically equivalent (Jöreskog, 1994, p. 386; Christoffersson & Gunsjö, 1996, p. 173). We shall now show that (17) equals their expression for the asymptotic covariance matrix of the sample polychoric correlations. To do so, we simply apply Jöreskog's (1994) Proposition 5 to the vector of all estimated polychoric correlations instead of to a single correlation as in Jöreskog's Equation 12, obtaining

$$\begin{aligned}\sqrt{N}\,(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\boldsymbol{\theta})) &\overset{a}{=} \left(\boldsymbol{\Delta}_{22}'\mathbf{D}_2\boldsymbol{\Delta}_{22}\right)^{-1}\boldsymbol{\Delta}_{22}'\mathbf{D}_2\sqrt{N}\,(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2) \\ &\quad - \left(\boldsymbol{\Delta}_{22}'\mathbf{D}_2\boldsymbol{\Delta}_{22}\right)^{-1}\boldsymbol{\Delta}_{22}'\mathbf{D}_2\boldsymbol{\Delta}_{21}\sqrt{N}\,(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}).\end{aligned} \tag{61}$$

Thus, $\sqrt{N}\,(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\boldsymbol{\theta})) \overset{a}{=} \mathbf{B}_{22}\sqrt{N}\,(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2) - \mathbf{B}_{22}\boldsymbol{\Delta}_{21}\sqrt{N}\,(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})$ and, using (11), we readily obtain (13). Finally, Christoffersson and Gunsjö's (1983) formulas are a direct application to the case $n > 2$ of Olsson's (1979) results. Hence, (16) reduces to Olsson's in the bivariate case.

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Arminger, G., Wittenberg, J., & Schepers, A. (1996). *MECOSA 3*. Friedrichsdorf: Additive GmbH.

Bermann, G. (1993). *Estimation and inference in bivariate and multivariate ordinal probit models*. Acta Universitatis Upsaliensis. Studia Statistica Upsaliensia 1. Uppsala, Sweden.

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihodd estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459.

Bock, R.D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika, 35*, 179–197.

Box, G.E.P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics, 25*, 290–302.

Browne, M.W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62–83.

Chang, E.C., D'Zurilla, T.J., & Maydeu-Olivares, A. (1994). Assessing the dimensionality of optimism and pessimism using a multimeasure approach. *Cognitive Therapy and Research, 18*, 143–160.

Christoffersson, A., & Gunsjö, A. (1983). *Analysis of structures for ordinal data* (Research Report 83-2). Uppsala, Sweden: University of Uppsala, Department of Statistics.

Christoffersson, A., & Gunsjö, A. (1996). A short note on the estimation of the asymptotic covariance matrix for polychoric correlations. *Psychometrika, 61*, 173–175.

Gong, G., & Samaniego, F.J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *Annals of Statistics, 9*, 861–869.

Gunsjö, A. (1994). *Faktoranalys av ordinala variabler*. Acta Universitas Upsaliensis. Studia Statistica Upsaliensia 2. Stockholm, Sweden: Almqvist & Wiksell.

Hamdan, M.A. (1970). The equivalence of tetrachoric and maximum likelihood estimates in 2 × 2 tables. *Biometrika, 57*, 212–215.

Jöreskog, K.G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika, 59*, 381–390.
Jöreskog, K.G., & Sörbom, D. (2001). *LISREL 8*. Chicago, IL: Scientific Software.
Küsters, U.L. (1987). *Hierarchische Mittelwert- und Kovarianztrukturmodelle mit nichtmetrischen endogenen Variablen*. Heidelberg: Physica-Verlag.
Lee, S.Y., Poon, W.Y., & Bentler, P.M. (1995). A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology, 48*, 339–358.
Mathai, A.M., & Provost, S.B. (1992). *Quadratic forms in random variables. Theory and applications*. New York: Marcel Dekker.
Maydeu-Olivares, A. (2001). Limited information estimation and testing of Thurstonian models for paired comparison data under multiple judgment sampling. *Psychometrika, 66*, 209–228.
Maydeu-Olivares, A., & Hernández, A. (2000). *Some remarks on estimating a covariance structure from a sample correlation matrix*. Working Paper. Statistics and Econometrics Series 00-62 (27). Universidad Carlos III de Madrid.
McDonald, R.P., & Mok, M.C. (1995). Goodness of fit in item response models. *Multivariate Behavioral Research, 30*, 23–40.
Moore, D.S. (1977). Generalized inverses, Wald's method, and the construction of chi-squared tests of fit. *Journal of the American Statistical Association, 72*, 131–137.
Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43*, 551–560.
Muthén, B. (1982). Some categorical response models with continuous latent variables. In K.G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation* (Vol. 1, pp. 65–79). Amsterdam: North-Holland.
Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115–132.
Muthén, B. (1993). Goodness of fit with categorical and other non normal variables. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 205–234). Newbury Park, CA: Sage.
Muthén, B., & Hofacker, C. (1988). Testing the assumptions underlying tetrachoric correlations. *Psychometrika, 53*, 563–578.
Muthén, L., & Muthén, B. (2001). *MPLUS*. Los Angeles, CA: Muthén & Muthén.
Muthén, B., & Satorra, A. (1995). Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika, 60*, 489–503.
Muthén, B., du Toit, S.H.C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Paper accepted for publication in Psychometrika.
Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*, 443–460.
Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika, 54*, 131–151.
Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von & and C.C. Clogg (Eds.), *Latent variable analysis: Applications to developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.
Scheier, M.F., & Carver, C.S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology, 4*, 219–247.

Manuscript received 1 NOV 2000
Final version received 3 SEP 2004
Published Online Date: 10 APR 2006