LIMITED INFORMATION ESTIMATION AND TESTING OF DISCRETIZED MULTIVARIATE NORMAL STRUCTURAL MODELS

IE Working Paper **MK8-103-I** **02 / 04 / 2003**

Alberto Maydeu Olivares

Instituto de Empresa
Marketing Dept.
C / María de Molina 11-15,
28006 – Madrid
España
Alberto.Maydeu@ie.edu

Abstract
We consider the estimation of multivariate normal structural models that have been discretized according to a set of thresholds. A popular estimation procedure for this restricted multinomial model consists in the following three stage estimator: First, estimate by maximum likelihood the thresholds for each variable separately from the univariate marginals of the contingency table. Then, estimate by maximum likelihood each of the polychoric correlations separately from the bivariate marginals of the contingency table given the estimated thresholds. Finally, 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. An unresolved issue is how to perform goodness of fit tests in this context.

We show that the first, second and third stage estimates can be expressed asymptotically as a linear function of the bivariate marginal proportions. Using this result, we propose limited information tests of discretized multivariate normality, as well as of the overall restrictions imposed by the model.

Keywords
GLS, WLS estimation, LISREL, categorical data analysis, data sparseness, goodness of fit, limited information estimation, pseudo-maximum likelihood estimation, paired comparisons data

Depósito Legal: M-20073 I.S.S.N.: 1579-4873

1. Introduction

A popular model for n-way contingency tables assumes that these arise by categorizing a n-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 is 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 Toit & Spisic, 1997). However, when continuous exogenous are included in the model, then 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 & Satorra, 1995; Muthén, du Toit & Spisic, 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 used 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 arises 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 are only meaningful if the distributional restrictions hold (i.e., if the data arises by categorizing a multivariate normal density). The main aim of the present research is to fill this gap using asymptotic theory for sample proportions. In so doing, we shall also review and integrate the literature on the use of this sequential procedure to estimate discretized multivariate normal structural models.

The paper is organized as follows. In Section 2 the sequential estimation procedure just described is presented. In Section 3 we provide the asymptotic distribution of the first, second, and third stage estimates using standard results from maximum likelihood estimation using grouped data and standard results from weighted least estimation of moment structures. In Section 4 we discuss goodness of fit testing. In this section after reviewing existing tests for the structural restrictions we propose tests of the distributional and of the overall restrictions imposed by the model on the bivariate marginals of the contingency table. Computational aspects of these tests are provided in Section 5. In Section 6 we provide a small simulation study to illustrate the small sample behavior of the sequential estimator under consideration and of the goodness of fit tests proposed. Finally, Section 7 includes three applications. In the first two applications we fit a covariance structure model to the 5-category items of the LOT (Scheier & Carver, 1985) and to the LSAT 6 binary data (Bock & Lieberman, 1970). In the third application we fit a mean and covariance structure model to Agresti's (1992) soft drink data (graded paired comparisons) and compare our results with those obtained by Böckenholt and Dillon (1997) using full information maximum likelihood.

Additional material is provided as appendices. In one of the appendices we show 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. In another appendix we review the estimation of the parameters of the correlation structure by minimizing a function of the polychoric correlations alone in the third stage. Finally, in another appendix we review the estimation of mean and covariance structure models.

2. Sequential estimation of discretized multivariate normal structural models

Let where $\boldsymbol{P}$ denotes a correlation matrix with elements $\rho_{ii'}$. Suppose that each z_i^*, i = 1, ..., n, has been categorized as $y_i = k_i$ if $\tau_{i_k} < z_i^* < \tau_{i_{k+1}}$, k_i = 0, ..., 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

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

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

$$\pi_{i_k} = \Pr\left(y_i = k_i\right) = \int_{\tau_{i_k}}^{\tau_{i_{k+1}}} \phi_1\left(z_i^* : 0, 1\right) dz_i^* \tag{2}$$

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

We shall first introduce some notation: Let $\boldsymbol{\pi}_i = \left(\pi_{i_0}, \cdots, \pi_{i_{K-1}}\right)'$, and

$\boldsymbol{\pi}_{ii'} = \left(\pi_{i_0 i'_0}, \pi_{i_0 i'_1}, \cdots, \pi_{i_{K-1} i'_{K-1}}\right)'$. Also, we let $\dot{\boldsymbol{\pi}}_1 = \left(\boldsymbol{\pi}_1', \cdots, \boldsymbol{\pi}_n'\right)'$ and $\dot{\boldsymbol{\pi}}_2 = \left(\boldsymbol{\pi}_{21}', \boldsymbol{\pi}_{31}' \cdots, \boldsymbol{\pi}_{nn-1}'\right)'$. where the sample counterparts of these univariate and bivariate marginal probabilities will be denoted by $\dot{\mathbf{p}}_1$ and $\dot{\mathbf{p}}_2$. Finally, let $\boldsymbol{\tau}_i = \left(\tau_{i_1}, \cdots, \tau_{i_{K-1}}\right)'$, $\boldsymbol{\tau} = \left(\boldsymbol{\tau}_1', \cdots, \boldsymbol{\tau}_n'\right)'$,

$\boldsymbol{\rho} = \left(\rho_{21}, \rho_{31}, \cdots, \rho_{n,n-1}\right)'$ and $\boldsymbol{\kappa} = \left(\boldsymbol{\tau}', \boldsymbol{\rho}'\right)'$.

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

<u>First stage</u>: Estimate the thresholds for each variable separately by maximizing

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

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

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

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

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.

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

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

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), $\hat{\mathbf{W}} = \left(\mathrm{Diag}\left(\hat{\boldsymbol{\Xi}}\right)\right)^{-1}$ (diagonally weighted least squares, DWLS: Gunsjö, 1994; Muthén, du Toit & Spisic, 1997), and $\hat{\mathbf{W}} = \mathbf{I}$ (unweighted least squares, ULS: Muthén, 1993).

3. Asymptotic distribution of the estimates

Before proceeding we notice that the univariate probabilities are simply sums of bivariate probabilities, exemplified here for $n = 3$,

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

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

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

Therefore,

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

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 kernel of the log-likelihood function for estimating $\boldsymbol{\tau}_i$ from a univariate marginal of the contingency table $\mathbf{p}_i$. Similarly, (5) is the kernel of 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 be readily obtained using standard results for maximum likelihood estimation for categorical models. Before proceeding, we shall review some of the relevant theory.

Let $\boldsymbol{\pi}$ and $\mathbf{p}$ be vectors of multinomial probabilities and sample proportions respectively. Consider a parametric structure for $\boldsymbol{\pi}$, $\boldsymbol{\pi}(\boldsymbol{\vartheta})$, with Jacobian matrix $\boldsymbol{\Delta} = \frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\vartheta}'}$, 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 (e.g., Agresti, 1990; Jöreskog, 1994)

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

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

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

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

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

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

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

Now, to apply (9) to the second stage estimates we need the asymptotic distribution of $\sqrt{N}\left(\dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2\left(\boldsymbol{\rho}, \hat{\boldsymbol{\tau}}\right)\right)$. In Appendix 1, we show that

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

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

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

where $\mathbf{B}_{22} = \left(\boldsymbol{\Delta}_{22}'\mathbf{D}_2\boldsymbol{\Delta}_{22}\right)^{-1}\boldsymbol{\Delta}_{22}'\mathbf{D}_2$, $\mathbf{D}_2 = \mathrm{Diag}\left(\dot{\boldsymbol{\pi}}_2\right)^{-1}$, and $\boldsymbol{\Delta}_{22} = \dfrac{\partial\dot{\boldsymbol{\pi}}_2}{\partial\boldsymbol{\rho}'}$. 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

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

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

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

where provided $n > 3$, the elements of $\tilde{\boldsymbol{\Gamma}}$ are fourth order marginal probabilities. Thus, we find by (14) and (15) that

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

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

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

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

Now, 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 $\tilde{\boldsymbol{\Delta}} = \dfrac{\partial \boldsymbol{\kappa}}{\partial \boldsymbol{\theta}'}$,

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \overset{a}{=} \mathbf{H}\sqrt{N}\left(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}\right) \tag{18}$$

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

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) simplify to

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

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 our presentation we assume a multivariate standard normal density that 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. For completeness, these are provided in Appendix 4 following Muthén (1978, 1993). Finally, in Appendix 5 we discuss the estimation of a discretized multivariate normal density with some mean and covariance structure following 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 are only meaningful if the distributional restrictions hold (i.e., if the data arises 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 we propose here a test of the joint distributional restrictions $\dot{\boldsymbol{\pi}}_2(\boldsymbol{\kappa})$. 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 $\mathbf{e}_s = \hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}\left(\hat{\boldsymbol{\theta}}\right)$. Using standard results for weighted least squares estimators

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

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

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

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

In (23) the $\chi_1^2\,'s$ are independent chi-square variables with one degree of freedom and the $\alpha_i\,'s$ are the non-null eigenvalues of

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

When $\hat{\mathbf{W}} = \hat{\boldsymbol{\Xi}}^{-1}$, (23) simplify to $T_s \xrightarrow{d} \chi^2_{r_s}$. On the other hand, when $\hat{\mathbf{W}} = \left(\mathrm{Diag}\left(\hat{\boldsymbol{\Xi}}\right)\right)^{-1}$ 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)

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

where $\overline{T}_s$ and $\overline{\overline{T}}_s$, 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 $d_s = \dfrac{\mathrm{Tr}\left(\mathbf{M}_s\right)^2}{\mathrm{Tr}\left(\mathbf{M}_s^2\right)/\,r_s}$ degrees of freedom.

4.2 Goodness of fit testing of the distributional restrictions

Consider now the distributional residuals $\mathbf{e}_d = \dot{\mathbf{p}}_2 - \dot{\boldsymbol{\pi}}_2\left(\hat{\boldsymbol{\kappa}}\right)$ and let $\boldsymbol{\Delta} = \dfrac{\partial \dot{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\kappa}'} = \left(\boldsymbol{\Delta}_{21} \mid \boldsymbol{\Delta}_{22}\right)$. In Appendix 1 we show that

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

From (15) and (26) we immediately have

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

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 non-null eigenvalues of $\mathbf{V}_d$ and the number of degrees of freedom available for testing is $\left(K^2 - 2K\right)\dfrac{n(n-1)}{2}$. 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

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

where $\overline{T}_d$ and $\overline{\overline{T}}_d$, 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 $d_d = \dfrac{\mathrm{Tr}\left(\mathbf{M}_d\right)^2}{\mathrm{Tr}\left(\mathbf{M}_d^2\right)/\,r_d}$ degrees of freedom.

4.3 Goodness of fit testing of the overall restrictions

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

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

From (15) and (30) we immediately have

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

Akin to (28), to test the overall restrictions of the model, $\dot{\boldsymbol{\pi}}_2\left(\boldsymbol{\theta}\right)$, 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 non-null eigenvalues of $\mathbf{V}_o$ and the number of degrees of freedom available for testing is $r_o = n\left(K-1\right) + \left(K-1\right)^2 \dfrac{n\left(n-1\right)}{2} - q$. Goodness of fit tests of the distributional 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

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

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 $d_o = \dfrac{\mathrm{Tr}\left(\mathbf{M}_o\right)^2}{\mathrm{Tr}\left(\mathbf{M}_o^2\right)/r_o}$ 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 $\mathbf{e}_s$

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

This is shown in Appendix 1. In this appendix we also show that

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

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

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

Thus, 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 $\dot{\mathbf{p}}_2$, which we denote by $\dot{\boldsymbol{\Gamma}}$ is of dimension $K^2\,\frac{n(n-1)}{2}$. Clearly, the size of this matrix grows very rapidly for increasing n 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 $\dot{\boldsymbol{\Gamma}}$. 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 additional computation. The approach employed here relies heavily in Jöreskog (1994). The approach taken here is not applicable in general to the computation of the overall tests.

5.1 Asymptotic covariance matrix of sample thresholds and polychoric correlations

Akin to (10) we have

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

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

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

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

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

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

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

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

$$N\,\mathrm{Acov}\left(\hat{\rho}_{ii'}, \hat{\boldsymbol{\tau}}_j\right) = \mathbf{G}_2^{(ii')}\left(\mathbf{C}_{ii'j} - \pi_{ii'}\pi_j\right)\mathbf{B}_{11}^{(j)\prime} \tag{42}$$

where $\mathbf{C}_{ii'j}$ is a $K^2 \times K$ table of trivariate probabilities. Finally, letting $\left(i, i', j, j'\right)$ 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

$$N\,\mathrm{Acov}\left(\hat{\rho}_{ii'}, \hat{\rho}_{jj'}\right) = \mathbf{G}_2^{(ii')}\left(\mathbf{C}_{ii'jj'} - \pi_{ii'}\pi_{jj'}\right)\mathbf{G}_2^{(jj')\prime} \tag{43}$$

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 $-\pi_{ii'}\pi_{jj'}$ can be dropped from (43).

To compute $\hat{\boldsymbol{\Xi}}$ we store into memory all $(K - 1) \times K$ $\mathbf{B}_{11}^{(i)}$ 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

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

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

Now, to obtain $\overline{T}_d$ and $\overline{\overline{T}}_d$, we need $\mathrm{Tr}\left(\mathbf{V}_d\right)$ and $\mathrm{Tr}\left(\mathbf{V}_d^2\right)$ 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

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

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

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

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 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

To illustrate the small sample behavior of the sequential estimation procedure under consideration and of the proposed distributional and overall tests we performed a small simulation study. We considered a three factor correlation structure model, $\boldsymbol{P}_{z^*} = \mathrm{Off}\left(\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'\right)$, with unrestricted thresholds for $n = 12$ variables, where each variable consists of $K = 3$ categories with thresholds $\boldsymbol{\tau}_i = (-0.5, 0.5)'$ and

$$\boldsymbol{\Lambda} = \begin{pmatrix} 0.7 & 0.6 & 0.5 & 0.4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.7 & 0.6 & 0.5 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0.6 & 0.5 & 0.4 \end{pmatrix}' \quad \boldsymbol{\Phi} = \begin{pmatrix} 1 & 0.3 & 0.4 \\ 0.3 & 1 & 0.5 \\ 0.4 & 0.5 & 1 \end{pmatrix}.$$

Two sample sizes were considered: $N = 200$ and $N = 1000$. ULS was employed in the third estimation stage. It is known that the asymptotically optimal WLS has a poorer small sample behavior than ULS due to the instability of the four-way proportions in small samples (Muthén, 1993). Furthermore, when ULS is employed no weight matrix needs to be inverted. Thus, larger models can be handled by ULS than by WLS. Alternatively, DWLS could have been used in the third stage. In the binary case, Maydeu-Olivares (2001) has shown that the small sample behavior of ULS and DWLS is very similar.

A summary of the parameter estimates and asymptotic standard errors are shown in Table 1. As can be seen in this table, a sample size of 200 observations suffices to obtain

----

Insert Tables 1 and 2 about here

----

accurate parameter estimates as there is no consistent bias in the parameter estimates. Also, 200 observations suffice to obtain accurate standard errors because although we observe a consistent downward bias, the relative bias does not exceed 7%. Of course when $N = 1000$ we obtain more accurate parameter estimates and standard errors. In this case there is no consistent bias neither in the parameter estimates nor in the standard errors, and the relative bias of the standard errors does not exceed 5%.

A summary of the goodness of fit results is shown in Table 2. As can be seen in this table, when $N = 1000$ both the scaled and the adjusted statistics match very well their reference distributions when testing the structural restrictions. When testing the distributional and overall restrictions, the mean and variance adjusted statistics also match well their reference distributions, particularly in the critical region $0.01 \le \alpha \le 0.20$. The mean scaled statistic, on the other hand, is too optimistic in this region. When $N = 200$, the mean and variance adjusted statistic matches adequately its reference distribution when testing the structural restrictions until $\alpha = 0.40$ whereas the mean statistic is too optimistic. Above this point, the mean scaled statistic behaves very well, better than the mean and variance adjusted statistic which is too optimistic. Finally, when testing the distributional and overall restrictions, the mean and variance adjusted statistics are too conservative within the critical region $0.01 \le \alpha \le 0.20$, whereas the mean scaled statistic is too liberal within this region. This led us to investigate the use of a heuristic procedure consisting in averaging the p-values obtained from these two statistics. Thus, in this table we have included an additional column, H, obtained by this heuristic procedure. As can be seen in this table, our heuristic procedure enable us to draw meaningful inferences about the distributional and overall fit of the model within the critical region $0.01 \le \alpha \le 0.20$ when $N = 200$. This is remarkable, as the number of degrees of freedom for testing the distributional and overall restrictions are 198 and 249, respectively.

In closing this section, we note that for the model under consideration and $N = 1000$ the correlation between the p-values obtained when testing the structural and distributional restrictions was 0.95, between the structural and overall p-values 0.32, and between the distributional and overall p-values 0.04. When $N = 200$, these correlations were 0.94, 0.33, and 0.03, respectively.

7. Applications

Three applications are considered. In the first two a covariance structure is assumed. Since these covariance structures are scale invariant and no restrictions are imposed on the thresholds, the parameters of the covariance structure can be estimated in the third stage by minimizing a discrepancy function of the polychoric correlations alone (see Appendices 4 and 5). In the first example, the variables consist of five categories, in the second example the data is binary. Finally, the third application involves a mean and covariance structure model in which the covariance structure is not scale invariant.

7.1 Life Orientation test

The Life Orientation Test (LOT: Scheier & Carver, 1985), is a eight item questionnaire designed to measure optimism and pessimism where each item consists of 5 categories. Chang, D'Zurilla and Maydeu-Olivares (1994) fitted the following covariance structure model to this questionnaire: $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a diagonal matrix, $\boldsymbol{\Lambda}' = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{41} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \lambda_{52} & \cdots & \lambda_{82} \end{pmatrix}$, and $\boldsymbol{\Psi} = \begin{pmatrix} 1 & \psi_{21} \\ \psi_{21} & 1 \end{pmatrix}$. 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, $\boldsymbol{\theta}' = (\lambda_{11}, \cdots, \lambda_{82}, \psi_{21})$ can be estimated in the third stage by minimizing a discrepancy function of the polychoric correlations only where for identification purposes $\boldsymbol{\Theta} = \mathbf{I} - \mathrm{Diag}(\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}')$.

Chang et al. (1994) used WLS and found that this model reproduced well the polychoric matrix. We shall re-analyze their data here which consists of 389 observations. Using ULS in the third stage we find that the model reproduces well the polychoric matrix $\overline{T}_s = 25.4$ on 14 df, $p = 0.15$ and $\overline{\overline{T}}_s = 15.4$ on 11.5 df, $p = 0.19$. But this test is only meaningful if the distributional restrictions hold. 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 $\alpha = 0.01$. Thus, it is not clear what to conclude. Our tests of the distributional assumptions, however, reveal that the hypothesis of joint categorized bivariate normality is to be rejected: $\overline{T}_d = 1070.9$ on 420 df, $p < 0.01$, and $\overline{\overline{T}}_d = 252.1$ on 98.9 df, $p < 0.01$. Not surprisingly, overall, the model fails to fit the bivariate tables: $\overline{T}_o =$ 1112.1 on 439 df, $p < 0.01$, and $\overline{\overline{T}}_o = 253.8$ on 100.2 df, $p < 0.01$. Thus, although we are able to reproduce well the matrix of polychoric correlations, the model does not really fit the bivariate tables because the distributional restrictions do not hold.

7.2 LSAT 6 data

These data, consisting of 1000 observations on 5 binary variables was originally reported in Bock and Lieberman (1970). The data have been re-analyzed 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, $p = 0.44$, and we computed Pearson's statistic using their parameter estimates obtaining $X^2 = 18.03$, $p = 0.65$.

We fitted a one factor model to these data using ULS in the third stage. The structural tests yielded $\overline{T}_s = 4.67$ on 5 df, $p = 0.46$ and $\overline{\overline{T}}_s = 4.31$ on 4.6 df, $p = 0.45$, so the model fits well the tetrachoric correlations. 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 $\overline{T}_o = 3.95$ on 5 df, $p = 0.56$, and $\overline{\overline{T}}_o = 3.55$ on 4.50 df, $p = 0.56$, which is similar to Bock and Lieberman's results.

7.3 Agresti's Soft drink data

This data set (Agresti, 1992) consists of 61 subjects comparing the taste of Coke, Classic Coke and Pepsi using a five point preference scale in a paired comparison design {Coke vs. Classic Coke, Coke vs. Pepsi, Classic Coke vs. Pepsi}. Böckenholt and Dillon (1997) collapsed the two extreme categories to reduce data sparseness, reducing the categories to {"Preference for i", "Indiference", ""Preference for i'"}. They fitted the following mean and covariance structure to the resulting 3^3 contingency table

$$\boldsymbol{\mu} = \mathbf{A}\boldsymbol{\nu} \qquad \boldsymbol{\Sigma} = \sigma^2\mathbf{A}\mathbf{A}' + 2\left(1-\sigma^2\right)\mathbf{I} \tag{48}$$

where $\boldsymbol{\nu}' = (\nu_1, \nu_2, 0)$ and

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}.$$

In addition, letting $\boldsymbol{\alpha}_1' = \left(\alpha_{1_1}, \alpha_{2_1}, \alpha_{3_1}\right)$ and $\boldsymbol{\alpha}_2' = \left(\alpha_{1_2}, \alpha_{2_2}, \alpha_{3_2}\right)$, the following restrictions are assumed on the thresholds: $\boldsymbol{\alpha}_1 = \gamma\mathbf{1}$, and $\boldsymbol{\alpha}_2 = -\gamma\mathbf{1}$. It is easy to show that this model is equivalent to a Thurstone's Case V model for graded paired comparisons data under the assumption of no order effects in the comparison of the stimuli. For an overview of Thurstonian models for graded paired comparisons see Tsai and Böckenholt (in press) and Maydeu-Olivares (in press).

Now, by (73) in Appendix 5 we find that

$$\begin{aligned} \boldsymbol{\tau}_1 &= \mathbf{D}(\gamma\mathbf{1} - \mathbf{A}\boldsymbol{\nu}) \\ \boldsymbol{\tau}_2 &= \mathbf{D}(-\gamma\mathbf{1} - \mathbf{A}\boldsymbol{\nu}) \end{aligned} \qquad \boldsymbol{P}_{z^*} = \mathbf{D}\left(\sigma^2\mathbf{A}\mathbf{A}' + 2\left(1-\sigma^2\right)\mathbf{I}\right)\mathbf{D} \tag{49}$$

where now $\boldsymbol{\tau}_1' = \left(\tau_{1_1}, \tau_{2_1}, \tau_{3_1}\right)$, $\boldsymbol{\tau}_2' = \left(\tau_{1_2}, \tau_{2_2}, \tau_{3_2}\right)$, and $\mathbf{D} = \mathrm{Diag}\left(\sigma^2\mathbf{A}\mathbf{A}' + 2\left(1-\sigma^2\right)\mathbf{I}\right)^{-\frac{1}{2}}$.

We estimated this model using ULS in the third stage. Böckenholt and Dillon (1997) estimated this model using full information maximum likelihood. In Table 3 we provide our parameter estimates and standard errors along with Böckenholt and Dillon's. As it can be

----

Insert Table 3 about here

----

seen in this Table, they are very similar. Böckenholt and Dillon (1997) reported a full information likelihood ratio statistic $G^2 = 23.65$ on 22 df, $p = 0.37$. We computed Pearson's statistic using their parameter estimates obtaining $X^2 = 20.35$, $p = 0.56$. Our tests of the distributional restrictions yield $\bar{T}_d = 7.57$ on 9 df, $p = 0.58$, and $\bar{\bar{T}}_d = 6.43$ on 7.64 df, $p = 0.56$. Hence, the assumption of joint categorized bivariate normality is not rejected. Also, our tests of the structural restrictions yield $\bar{T}_s = 1.34$ on 5 df, $p = 0.93$ and $\bar{\bar{T}}_s = 1.03$ on 3.8 df, $p = 0.89$. Hence the model reproduces very well the sample thresholds and polychorics. Finally, our tests of the overall restrictions yield $\bar{T}_o = 7.63$ on 14 df, $p = 0.91$, and $\bar{\bar{T}}_o = 5.50$ on 10.09 df, $p = 0.86$. Thus, the model reproduces very well the bivariate tables.

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. As an alternative to these statistics, one could consider the use of a weighted quadratic form using a generalized inverse of a consistent estimate of the asymptotic covariance matrix of the distributional and overall residuals as weight matrix. These generalized Wald tests (Moore, 1977) would be asymptotically chi-squared distributed. However, these tests would only be computationally feasible for small models as the asymptotic covariance matrix of the distributional and overall residuals is of dimension $K^2 \frac{n(n-1)}{2}$. In addition, the generalized inverses required by these asymptotically chi-square tests are computationally demanding except for small models and they may be unstable in small samples as the matrix to be inverted depends on four-way proportions. On the other hand, we have shown that the proposed distributional tests can be computed very efficiently for very large models. It does not seem possible in general to compute the proposed overall tests without storing the large asymptotic covariance matrix of the bivariate proportions in memory. So, in general, there is a limitation in the size of the models that can be tested using our proposed overall tests.

We have investigated the small sample performance of the sequential estimator and of the proposed tests. We have shown that for a covariance structure model for 12 variables that has been tricotomized, one can obtain accurate parameter estimates and standard errors with as few as 200 observations. Furthermore, one can draw meaningful inferences about the structural, distributional and overall misfit of the model with this small sample size.

Clearly, as the number of categories and variables increases, the number of degrees of freedom available for testing the distributional and overall restrictions grows very rapidly. Thus, in applications the distributional and overall null hypotheses are very likely to be rejected when the model under consideration is large. Further work is needed to develop a test of close fit to these null hypotheses along the lines of Browne and Cudeck (1993). Also, further work is needed to investigate the robustness of the sequential estimation procedure under mispecification of the distributional assumptions. Finally, a test of the joint distributional assumptions when all the observed variables are dichotomous is needed.

We have not considered in this paper structured multivariate normal models in which some but not all the variables are categorized. Neither we have 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 & Satorra, 1995; Muthén, du Toit & Spisic, 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.

References

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

Agresti, A. (1992). Analysis of ordinal paired comparison data. Applied Statistics, 41, 287-297.

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., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. Psychometrika, 35, 179-197.

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

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, 16, 769-771.

Böckenholt, U. & Dillon, W.R. (1997). Modeling within-subject dependencies in ordinal paired comparison data, Psychometrika, 62, 411-434.

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

Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long [Eds.] Testing structural equation models (pp. 136-162). Newbury Park, CA: Sage.

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. Dept. 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 x 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. (in press). Limited information estimation and testing of Thurstonian models for preference data. Mathematical Social Sciences.

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., 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.

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.

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 Eye 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.

Tsai, R.C. & Böckenholt, U. (in press). Two-level linear paired comparison models: Estimation and identifiability issues. Mathematical Social Sciences.

TABLE 1

Simulation results for a three factor model: Parameter estimates and standard errors

par	*true*	$N = 200$			$N = 1000$		
		$\overline{x}$ *est*	$\overline{x}$ *SE*	*std SE*	$\overline{x}$ *est*	$\overline{x}$ *SE*	*std SE*
$\lambda_{1,1}$	0.7	0.70	0.09	0.10	0.70	0.04	0.04
$\lambda_{2,1}$	0.6	0.60	0.09	0.10	0.60	0.04	0.04
$\lambda_{3,1}$	0.5	0.50	0.09	0.09	0.50	0.04	0.04
$\lambda_{4,1}$	0.4	0.40	0.10	0.10	0.40	0.04	0.04
$\lambda_{5,2}$	0.7	0.69	0.09	0.09	0.70	0.04	0.04
$\lambda_{6,2}$	0.6	0.60	0.09	0.09	0.60	0.04	0.04
$\lambda_{7,2}$	0.5	0.50	0.09	0.09	0.50	0.04	0.04
$\lambda_{8,2}$	0.4	0.40	0.09	0.10	0.40	0.04	0.04
$\lambda_{9,3}$	0.7	0.69	0.09	0.09	0.70	0.04	0.04
$\lambda_{10,3}$	0.6	0.60	0.09	0.09	0.60	0.04	0.04
$\lambda_{11,3}$	0.5	0.49	0.09	0.09	0.50	0.04	0.04
$\lambda_{12,3}$	0.4	0.40	0.09	0.10	0.40	0.04	0.04
$\phi_{2,1}$	0.3	0.31	0.11	0.12	0.30	0.05	0.05
$\phi_{3,1}$	0.4	0.41	0.11	0.12	0.40	0.05	0.05
$\phi_{3,2}$	0.5	0.50	0.11	0.11	0.50	0.05	0.05

Notes: ULS was employed in the third stage; all variables had 3 categories.

TABLE 2

Simulation results for a three factor model: Goodness of fit tests

		Structural restrictions						*Distributional restrictions*						*Overall restrictions*					
		$N = 200$			$N = 1000$			$N = 200$			$N = 1000$			$N = 200$			$N = 1000$		
		$\bar{T}_s$	$\bar{\bar{T}}_s$	H	$\bar{T}_s$	$\bar{\bar{T}}_s$	H	$\bar{T}_d$	$\bar{\bar{T}}_d$	H	$\bar{T}_d$	$\bar{\bar{T}}_d$	H	$\bar{T}_o$	$\bar{\bar{T}}_o$	H	$\bar{T}_o$	$\bar{\bar{T}}_o$	H
Mean		52.1	36.6	--	51.0	42.9	--	201.1	93.7	--	199.7	148.9	--	252.9	103.6	--	250.8	176.0	--
Var.		129.6	57.4	--	106.5	74.3	--	489.2	100.0	--	469.6	250.2	--	643.7	95.3	--	614.7	288.3	--
rejection rates	1%	2.5	0.8	1.1	1.0	0.4	0.5	2.5	0.2	0.5	1.7	0.7	1.0	2.7	<0.1	0.6	1.8	0.5	0.5
	5%	8.3	4.9	6.1	5.2	4.1	4.8	9.4	2.4	4.5	8.0	4.8	6.5	9.9	1.7	4.1	8.1	4.5	6.1
	10%	13.8	9.6	11.9	10.0	8.4	9.2	16.1	6.7	10.5	13.8	10.3	12.7	16.4	4.9	10.2	14.5	9.6	12.3
	20%	24.0	21.2	22.9	20.1	18.6	19.5	25.6	17.7	22.0	24.9	21.7	23.6	26.1	16.4	22.1	25.3	21.1	23.0
	30%	34.3	32.3	33.0	31.5	30.3	30.8	34.6	28.9	32.2	34.0	32.0	32.9	36.1	28.0	33.0	34.4	31.4	33.0
	40%	44.2	43.3	43.7	41.7	41.4	41.6	44.3	41.8	42.7	42.8	42.1	42.3	46.3	42.6	44.3	43.2	41.8	42.8
	50%	52.9	53.9	53.2	51.5	51.6	51.5	55.0	57.6	55.7	53.1	43.1	53.0	56.0	57.4	56.4	52.4	52.8	52.7
	60%	62.5	65.2	63.5	60.7	61.9	61.6	64.2	70.9	66.8	63.6	63.6	62.6	65.8	72.2	68.3	61.7	63.6	62.4
	70%	71.3	75.4	72.8	69.1	70.6	69.9	73.9	81.0	77.9	70.1	73.9	71.8	74.4	82.8	78.5	70.3	73.7	72.2
	80%	79.9	85.1	82.3	79.0	81.2	80.1	81.2	90.2	85.5	78.8	83.1	80.7	82.5	90.9	86.6	79.8	83.1	81.4
	90%	89.4	94.4	92.9	89.0	91.1	90.0	90.5	96.6	92.2	89.4	93.2	91.3	89.9	97.2	94.1	88.8	92.3	91.0

Notes: ULS was employed in the third stage; $\bar{T}$ and $\bar{\bar{T}}$ denote the mean scaled and mean and variance adjusted statistics, respectively; H is a heuristic p-value obtained by averaging the p-values obtained from $\bar{T}$ and $\bar{\bar{T}}$; $df_s = 51$, $df_d = 178$, and $df_o = 249$.

TABLE 3

Parameter estimates and standard errors for Agresti's soft drink data

par.	*Böckenholt & Dillon*	*sequential estimator (ULS)*
γ	-0.37 (0.07)	-0.37 (0.06)
ν_1	-0.48 (0.20)	-0.47 (0.19)
ν_2	-0.24 (0.19)	-0.23 (0.18)
σ^2	0.40 (0.18)	0.40 (0.15)

Notes: standard errors in parentheses

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 $\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho},\hat{\boldsymbol{\tau}})\overset{a}{=}\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho},\boldsymbol{\tau})+\boldsymbol{\Delta}_{21}(\hat{\boldsymbol{\tau}}-\boldsymbol{\tau})$, where $\boldsymbol{\Delta}_{21}=\dfrac{\partial\dot{\boldsymbol{\pi}}_2}{\partial\boldsymbol{\tau}'}$. Thus, $\sqrt{N}\left(\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho},\hat{\boldsymbol{\tau}})-\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho},\boldsymbol{\tau})\right)\overset{a}{=}\boldsymbol{\Delta}_{21}\sqrt{N}(\hat{\boldsymbol{\tau}}-\boldsymbol{\tau})$. Now, by (11), $\sqrt{N}\left(\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho},\hat{\boldsymbol{\tau}})-\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho},\boldsymbol{\tau})\right)\overset{a}{=}\boldsymbol{\Delta}_{21}\mathbf{B}_{11}\mathbf{T}\sqrt{N}\left(\dot{\mathbf{p}}_2-\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho},\boldsymbol{\tau})\right)$. Equation (12) follows by noting that $\sqrt{N}\left(\dot{\mathbf{p}}_2-\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho},\hat{\boldsymbol{\tau}})\right)=\sqrt{N}\left(\dot{\mathbf{p}}_2-\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho},\boldsymbol{\tau})\right)-\sqrt{N}\left(\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho},\hat{\boldsymbol{\tau}})-\dot{\boldsymbol{\pi}}_2(\boldsymbol{\rho},\boldsymbol{\tau})\right)$ □

Proof of Equation (26):

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

Proof of Equation (30):

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

Proof of Equation (34):

By (21) and (14)

$$\sqrt{N}\mathbf{e}_s\overset{a}{=}\left(\mathbf{I}-\tilde{\boldsymbol{\Delta}}\mathbf{H}\right)\mathbf{G}\sqrt{N}\left(\dot{\mathbf{p}}_2-\dot{\boldsymbol{\pi}}_2\right). \tag{50}$$

Now from (30), $\mathbf{e}_o\overset{a}{=}\left(\mathbf{I}-\boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G}\right)\left(\dot{\mathbf{p}}_2-\dot{\boldsymbol{\pi}}_2\right)$. Thus, $\mathbf{e}_o\overset{a}{=}\left(\dot{\mathbf{p}}_2-\dot{\boldsymbol{\pi}}_2\right)+\boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G}\left(\dot{\mathbf{p}}_2-\dot{\boldsymbol{\pi}}_2\right)$. Now, adding and subtracting $\boldsymbol{\Delta}\mathbf{G}\left(\dot{\mathbf{p}}_2-\dot{\boldsymbol{\pi}}_2\right)$ to this equation and re-arranging terms, $\mathbf{e}_o\overset{a}{=}\left(\mathbf{I}-\boldsymbol{\Delta}\mathbf{G}\right)\left(\dot{\mathbf{p}}_2-\dot{\boldsymbol{\pi}}_2\right)+\boldsymbol{\Delta}\left(\mathbf{I}-\tilde{\boldsymbol{\Delta}}\mathbf{H}\right)\mathbf{G}\left(\dot{\mathbf{p}}_2-\dot{\boldsymbol{\pi}}_2\right)$, and (34) follows immediately from (26) and (50). □

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), $T_d \overset{a}{=} N\mathbf{e}'\mathbf{A}\mathbf{e}$. Also, from (50) $\mathbf{W}^{\frac{1}{2}}\sqrt{N}\mathbf{e}_s \overset{a}{=} \mathbf{W}^{\frac{1}{2}}\left(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H}\right)\mathbf{G}\sqrt{N}\mathbf{e}$. Then, letting $\mathbf{B} = \mathbf{G}'\left(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H}\right)'\mathbf{W}\left(\mathbf{I} - \tilde{\boldsymbol{\Delta}}\mathbf{H}\right)\mathbf{G}$, $T_s \overset{a}{=} N\mathbf{e}'\mathbf{B}\mathbf{e}$. Finally, letting $\mathbf{C} = \left(\mathbf{I} - \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G}\right)'\left(\mathbf{I} - \boldsymbol{\Delta}\tilde{\boldsymbol{\Delta}}\mathbf{H}\mathbf{G}\right)$, $T_o \overset{a}{=} N\mathbf{e}'\mathbf{C}\mathbf{e}$. Equations (35), (36) and (37) then follow from Theorem 3.2d.4 in Mathai and Provost (1992). □

<u>Appendix 2</u>: Derivatives involved in Δ_{11}, Δ_{21}, and Δ_{22}

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

$$\pi_{i_k} = \Phi_1\left(\tau_{i_k}\right) - \Phi_1\left(\tau_{i_{k-1}}\right) \tag{51}$$

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

$$\Phi_1\left(\tau_{i_0}\right) = 0 \qquad\qquad \Phi_1\left(\tau_{i_K}\right) = 1 \tag{52}$$

Then, the elements of $\boldsymbol{\Delta}_{11}$ can be obtained using (51) and (52) with

$$\frac{\partial \Phi_1\left(\tau_{i_k}\right)}{\partial \tau_{i_k}} = \phi_1\left(\tau_{i_k}\right). \tag{53}$$

We also note that because of (51), (4) has a closed form solution

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

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

$$\pi_{i_k i'_{k'}} = \Phi_2\left(\tau_{i_k}, \tau_{i'_{k'}}, \rho_{ii'}\right) - \Phi_2\left(\tau_{i_{k-1}}, \tau_{i'_{k'}}, \rho_{ii'}\right) - \Phi_2\left(\tau_{i_k}, \tau_{i'_{k'-1}}, \rho_{ii'}\right) + \Phi_2\left(\tau_{i_{k-1}}, \tau_{i'_{k'-1}}, \rho_{ii'}\right) \tag{54}$$

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

$$\begin{array}{lll} \Phi_2\left(\tau_{i_k}, \tau_{i'_K}, \rho_{ii'}\right) = \Phi_1\left(\tau_{i_k}\right) & \Phi_2\left(\tau_{i_K}, \tau_{i'_K}, \rho_{ii'}\right) = 1 & \Phi_2\left(\tau_{i_0}, \tau_{i'_{k'}}, \rho_{ii'}\right) = 0 \\ \Phi_2\left(\tau_{i_K}, \tau_{i'_{k'}}, \rho_{ii'}\right) = \Phi_1\left(\tau_{i'_k}\right) & \Phi_2\left(\tau_{i_0}, \tau_{i'_0}, \rho_{ii'}\right) = 0 & \Phi_2\left(\tau_{i_k}, \tau_{i'_0}, \rho_{ii'}\right) = 0 \end{array} \tag{55}$$

Then, the elements of $\boldsymbol{\Delta}_{21}$ can be obtained using (52) through (55), and

$$\frac{\partial \Phi_2\left(\tau_{i_k}, \tau_{i'_{k'}}, \rho_{ii'}\right)}{\partial \tau_{i_k}} = \phi_1\left(\tau_{i_k}\right)\Phi_1\left(\frac{\tau_{i'_{k'}} - \rho_{ii'}\tau_{i_k}}{\sqrt{1-\rho_{ii'}^2}}\right) \tag{56}$$

(Olsson, 1979: Equation 12). Finally, the elements of $\boldsymbol{\Delta}_{22}$ can also be obtained using (52) through (55), and

$$\frac{\partial \Phi_2\left(\tau_{i_k}, \tau_{i'_{k'}}, \rho_{ii'}\right)}{\partial \rho_{ii'}} = \phi_2\left(\tau_{i_k}, \tau_{i'_{k'}} : \rho_{ii'}\right) \tag{57}$$

(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 two mathematically independent univariate probabilities, say $\ddot{\pi}_i = \Pr(y_i = 1)$ and $\ddot{\pi}_{i'} = \Pr(y_{i'} = 1)$, and one mathematically independent bivariate probability, say $\ddot{\pi}_{ii'} = \Pr[(y_i = 1) \cap \Pr(y_{i'} = 1)]$. Let $\ddot{\boldsymbol{\pi}}_1 = (\ddot{\pi}_1, \cdots, \ddot{\pi}_n)'$, $\ddot{\boldsymbol{\pi}}_2 = (\ddot{\pi}_{21}, \ddot{\pi}_{31}, \cdots, \ddot{\pi}_{nn-1})'$ and $\ddot{\boldsymbol{\pi}} = (\ddot{\boldsymbol{\pi}}_1, \ddot{\boldsymbol{\pi}}_2)'$, with sample counterparts $\ddot{\mathbf{p}} = (\ddot{\mathbf{p}}_1, \ddot{\mathbf{p}}_2)'$. Muthén (1978) estimates each threshold and tetrachoric correlation separately using

$$\hat{\tau}_i = -\Phi_1^{-1}(\ddot{p}_i) \tag{58}$$

$$\hat{\rho}_{ii'} = \Phi_2^{-1}(\ddot{p}_{ii'} | -\hat{\tau}_i, -\hat{\tau}_{i'}) \tag{59}$$

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 (58) and (59) is equivalent to employing (4) and (5) (Hamdan, 1970). Now,

$$\dot{\boldsymbol{\pi}}_2 = \mathbf{c} + \mathbf{C}\ddot{\boldsymbol{\pi}} = \left(\mathbf{C}_1 \;\vdots\; \mathbf{C}_2\right)\begin{pmatrix} \ddot{\boldsymbol{\pi}}_1 \\ \ddot{\boldsymbol{\pi}}_2 \end{pmatrix}, \tag{60}$$

illustrated here for $n = 2$

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

Now, by (14) and (60), $\sqrt{N}(\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}) \overset{a}{=} \mathbf{GC}\sqrt{N}(\ddot{\mathbf{p}} - \ddot{\boldsymbol{\pi}})$. Furthermore, it is easy to verify that

$$\mathbf{GC} = \begin{pmatrix} \left(\dfrac{\partial \ddot{\boldsymbol{\pi}}_1}{\partial \boldsymbol{\tau}'}\right)^{-1} & \mathbf{0} \\ -\left(\dfrac{\partial \ddot{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\rho}'}\right)^{-1} \dfrac{\partial \ddot{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\tau}'} \dfrac{\partial \ddot{\boldsymbol{\pi}}_1}{\partial \boldsymbol{\tau}'} & \left(\dfrac{\partial \ddot{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\rho}'}\right)^{-1} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \ddot{\boldsymbol{\pi}}_1}{\partial \boldsymbol{\tau}'} & \mathbf{0} \\ \dfrac{\partial \ddot{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\tau}'} & \dfrac{\partial \ddot{\boldsymbol{\pi}}_2}{\partial \boldsymbol{\rho}'} \end{pmatrix}^{-1} = \left(\frac{\partial \ddot{\boldsymbol{\pi}}}{\partial \boldsymbol{\kappa}'}\right)^{-1} = \ddot{\mathbf{G}}^{-1}. \tag{61}$$

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

$$\boldsymbol{\Xi} = \ddot{\mathbf{G}}^{-1}\ddot{\boldsymbol{\Gamma}}\ddot{\mathbf{G}}^{-1\prime} \tag{62}$$

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: 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

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

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

Appendix 4: Correlation structure models with unrestricted thresholds

Suppose that a parametric structure is imposed on $\boldsymbol{\rho}$, say $\boldsymbol{\rho}$ ($\boldsymbol{\theta}$), but that $\boldsymbol{\tau}$ is left unconstrained. Now, we partition $\hat{\mathbf{W}} = \begin{pmatrix} \hat{\mathbf{W}}_{11} & \hat{\mathbf{W}}_{21}' \\ \hat{\mathbf{W}}_{21} & \hat{\mathbf{W}}_{22} \end{pmatrix}$ in (6) according to the partitioning of $\boldsymbol{\kappa} = \left(\boldsymbol{\tau}', \boldsymbol{\rho}'\right)'$. Then, we can estimate $\boldsymbol{\theta}$ by minimizing in the third stage the weighted least squares function

$$F_2 = \left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\boldsymbol{\theta})\right)' \hat{\mathbf{W}}_{22} \left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\boldsymbol{\theta})\right) \tag{63}$$

which is computationally more convenient than (6). In this case, $\hat{F} = \hat{F}_2$ and the parameter estimates for $\boldsymbol{\theta}$ and their standard errors estimated by minimizing F_2 will equal those obtained by minimizing F (Muthén, 1978: p. 554; Maydeu-Olivares & Hernández, 2000: Appendix 1). Also, denoting the asymptotic covariance matrix of the sample polychoric correlations by $\boldsymbol{\Xi}_{22}$, we have when $\hat{\mathbf{W}}_{22} = \hat{\boldsymbol{\Xi}}_{22}^{-1}$ (WLS), $\hat{\mathbf{W}}_2 = diag\left(\hat{\boldsymbol{\Xi}}_2\right)^{-1}$ (DWLS), and $\hat{\mathbf{W}}_2 = \mathbf{I}$ (ULS).

Let $\tilde{\boldsymbol{\Delta}}_{22} = \dfrac{\partial \boldsymbol{\rho}}{\partial \boldsymbol{\theta}'}$ and $\mathbf{H}_2 = \left(\tilde{\boldsymbol{\Delta}}_{22}{}'\mathbf{W}_{22}\tilde{\boldsymbol{\Delta}}_{22}\right)^{-1} \tilde{\boldsymbol{\Delta}}_{22}{}'\mathbf{W}_{22}$. When a parametric structure is imposed on $\boldsymbol{\rho}$, say $\boldsymbol{\rho}(\boldsymbol{\theta})$, $\boldsymbol{\tau}$ is left unconstrained and $\boldsymbol{\theta}$ is estimated by minimizing (63) we have

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \overset{a}{=} \mathbf{H}_2 \sqrt{N}\left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\right) \tag{64}$$

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{H}_2 \boldsymbol{\Xi}_{22} \mathbf{H}_2{}'\right) \tag{65}$$

$$\sqrt{N}\mathbf{e}_s^{(2)} := \sqrt{N}\left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\left(\hat{\boldsymbol{\theta}}\right)\right) \overset{a}{=} \left(\mathbf{I} - \tilde{\boldsymbol{\Delta}}_{22}\mathbf{H}_2\right)\sqrt{N}\left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\right) \tag{66}$$

$$\sqrt{N}\mathbf{e}_s^{(2)} \xrightarrow{d} N\left(\mathbf{0}, \mathbf{V}_{22}\right) \qquad \mathbf{V}_s^{(22)} = \left(\mathbf{I} - \tilde{\boldsymbol{\Delta}}_{22}\mathbf{H}_2\right)\boldsymbol{\Xi}_{22}\left(\mathbf{I} - \tilde{\boldsymbol{\Delta}}_{22}\mathbf{H}_2\right)' \tag{67}$$

$$T_s^{(2)} = N\hat{F}_2 \xrightarrow{d} \sum_{i=1}^{r_s} \alpha_i \chi_1^2 \tag{68}$$

where the degrees of freedom available for testing the structural restrictions $\boldsymbol{\rho}(\boldsymbol{\theta})$ are now $r_s = \dfrac{n(n-1)}{2} - q$. In (68) the α_i's are now the non-null eigenvalues of $\mathbf{M}_s^{(22)} = \mathbf{W}_{22}\left(\mathbf{I} - \tilde{\boldsymbol{\Delta}}_{22}\mathbf{H}_2\right)\boldsymbol{\Xi}_{22}$. Also, when $\hat{\mathbf{W}}_{22} = \hat{\boldsymbol{\Xi}}_{22}^{-1}$, (65) and (68) simplify to

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\xrightarrow{d} N\left(\mathbf{0},\left(\tilde{\boldsymbol{\Delta}}_{22}{}'\boldsymbol{\Xi}_{22}^{-1}\tilde{\boldsymbol{\Delta}}_{22}\right)^{-1}\right) \qquad T_s^{(2)}\xrightarrow{d}\chi^2_{r_s}\,. \tag{69}$$

On the other hand, when $\hat{\mathbf{W}}_{22}=\left(\mathrm{Diag}\left(\hat{\boldsymbol{\Xi}}_{22}\right)\right)^{-1}$ or $\hat{\mathbf{W}}_{22}=\mathbf{I}$, a goodness of fit of the model can be obtained using

$$\overline{T}_s^{(2)}=\frac{T_s^{(2)}}{\mathrm{Tr}\left(\mathbf{M}_s^{(22)}\right)/\,r_s} \qquad \overline{\overline{T}}_s^{(2)}=\frac{T_s^{(2)}}{\mathrm{Tr}\left(\mathbf{M}_s^{(22)2}\right)/\,r_s} \tag{70}$$

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 $d_s=\dfrac{\mathrm{Tr}\left(\mathbf{M}_s^{(22)}\right)^2}{\mathrm{Tr}\left(\mathbf{M}_s^{(22)2}\right)/\,r_s}$ degrees of freedom.

Appendix 5: Mean and covariance structure models

Throughout our presentation we have assumed that the model's underlying normal density had a correlation structure and a zero mean vector. Suppose instead that our model is $\mathbf{y}^* \sim N\left(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta})\right)$ where $\boldsymbol{\Sigma}$ denotes a covariance matrix and suppose that each variable y_i^* has been categorized using $y_i = k_i$ if $\alpha_{i_k} < y_i^* < \alpha_{i_{k+1}}$ where $\alpha_{i_0} = -\infty, \alpha_{i_k} = \infty$. According to this model

$$\Pr\left[\bigcap_{i=1}^{n}\left(y_i = k_i\right)\right] = \int \underset{\tilde{\mathbf{R}}}{\cdots} \int \phi_n\left(\mathbf{y}^* : \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta})\right) d\mathbf{y}^* \tag{71}$$

where $\phi_n(\bullet)$ denotes a n-dimensional normal density function, and $\tilde{\mathbf{R}}$ is a n-dimensional area of integration with intervals $R_i = \left(\alpha_{i_k}, \alpha_{i_{k+1}}\right)$.

Now, these pattern probabilities are unchanged when we perform a change of variable of integration in (71) standardizing $\mathbf{y}^*$ using

$$\mathbf{z}^* = \mathbf{D}\left(\mathbf{y}^* - \boldsymbol{\mu}\right) \qquad\qquad \mathbf{D} = \mathrm{Diag}\left(\boldsymbol{\Sigma}\right)^{-\frac{1}{2}} \tag{72}$$

where note that the diagonal matrix $\mathbf{D}$ depends on $\boldsymbol{\theta}$. Then, at $y_i^* = \alpha_{i_k}$, $\tau_{i_k} := \dfrac{\alpha_{i_k} - \mu_i}{\sqrt{\sigma_i^2}}$,

where σ_i^2 denotes a diagonal element of $\boldsymbol{\Sigma}$. Letting $\boldsymbol{\tau}_k = \left(\tau_{1_k}, \cdots, \tau_{n_k}\right)'$ and $\boldsymbol{\alpha}_k = \left(\alpha_{1_k}, \cdots, \alpha_{n_k}\right)'$ the transformation (72) yields (1) where $\boldsymbol{\mu}_{z^*} = \mathbf{0}$ and

$$\boldsymbol{\tau}_k = \mathbf{D}\left(\boldsymbol{\alpha}_k - \boldsymbol{\mu}\right) \qquad\qquad \boldsymbol{P}_{z^*} = \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}. \tag{73}$$

Maydeu-Olivares and Hernández (2000) showed that if and only if $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scale invariant, then it is possible to find a reparameterization of $\boldsymbol{\theta}$ so that $\boldsymbol{P}_{z^*}$ has the same functional form as $\boldsymbol{\Sigma}$. Thus, in this case we can take rid of the diagonal matrix $\mathbf{D}$.

Consider now the situation in which $\boldsymbol{\alpha}$ is unconstrained and $\boldsymbol{\mu} = \mathbf{0}$. In this case, if $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is not scale invariant, then the restrictions imposed on the thresholds $\boldsymbol{\tau}$ and on the polychoric correlations $\boldsymbol{P}$ are $\boldsymbol{\tau}_k = \mathbf{D}\boldsymbol{\alpha}_k$ and $\boldsymbol{P}_{z^*} = \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}$. Therefore, the parameters of a categorized covariance structure model that is not scale invariant may not be estimated in the third stage from the polychoric correlations alone because the thresholds $\boldsymbol{\tau}$ also depend on $\boldsymbol{\theta}$ through the model-based matrix $\mathbf{D}$. For further details, see Maydeu-Olivares and Hernández (2000).