# SOME REMARKS ON ESTIMATING A COVARIANCE STRUCTURE MODEL FROM A SAMPLE CORRELATION MATRIX

Alberto Maydeu-Olivares
and Adolfo Hernández Estrada

00-62

WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 00-62
Statistics and Econometrics Series 27
September 2000

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

# SOME REMARKS ON ESTIMATING A COVARIANCE STRUCTURE MODEL FROM A SAMPLE CORRELATION MATRIX

Alberto Maydeu-Olivares and Adolfo Hernández Estrada*

**Abstract**

A popular model in structural equation modeling involves a multivariate normal density with a structured covariance matrix that has been categorized according to a set of thresholds. In this setup one may estimate the covariance structure parameters from the sample tetrachoric/ polychoric correlations but only if the covariance structure is scale invariant. Doing so when the covariance structure is not scale invariant results in estimating a more restricted covariance structure than the one intended. When the covariance structure is not scale invariant, then the model parameters must be estimated jointly from the sample thresholds and tetrachoric/ polychoric correlations.

In general, when fitting a covariance structure from a sample correlation matrix one should consider the population correlation structure under the null hypothesis. This is obtained by pre and post-multiplying the covariance structure by a diagonal matrix consisting of the inverse of the square root of the diagonal of the covariance structure under consideration.

We provide computer algebra code for assessing whether a covariance structure is scale invariant and for assessing the identification of threshold and correlation structures.

**Keywords:** Lisrel; Mplus; Mathematica; tau-equivalent model; normal ogive model

*Maydeu-Olivares, Dept. of Psychology, University of Barcelona; Hernández, Dept. of Statistics and Econometrics; Universidad Carlos III de Madrid; e-mail: adolfo@est-econ.uc3m.es

## Abstract

Estimating a covariance structure from a sample correlation matrix requires complex non-linear constraints among the model parameters. When the covariance structure is scale invariant, then a simpler approach is possible. We cover the continuous observed variables case, as well as the categorical observed variables case, the latter under discretized multivariate normality assumptions. In the categorical case, we show that the covariance structure parameters can be estimated from the sample polychoric correlations if and only if the covariance structure is scale invariant. Otherwise, they must be estimated from the sample thresholds and polychoric correlations jointly. In the continuous case, we show that one can estimate any covariance structure from a sample correlation matrix by minimizing a normal theory discrepancy function for sample covariances.

Because determining whether a model is scale invariant leads to considerable simplifications, we provide computer algebra code that may be employed to determine whether a covariance structure model is scale invariant. Also, because when estimating a covariance structure from a sample correlation matrix not all covariance structure parameters may be identified we also provide computer algebra code to be used to determine which parameters of the covariance structure can be estimated from sample correlations.

Numerical examples in which scale and non-scale invariant models are estimated from categorical and continuous data are provided.


Keywords: Lisrel, Mplus, Mathematica, tau-equivalent model, normal ogive model

# 1. Introduction

In covariance structure analysis, one wishes to model the variances and covariances of the observed variables. That is, one assumes that the population covariance matrix $\boldsymbol{\Sigma}$ of the observed variables depends on a parameter vector $\boldsymbol{\theta}$, say $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, whereas no structure is imposed on the population mean vector $\boldsymbol{\mu}$. The objective is then to estimate the parameter vector $\boldsymbol{\theta}$ from a sample covariance matrix. In contrast, in correlation structure analysis, one wishes to model the correlations among the observed variables. Thus, in this case it is the population correlation matrix $\boldsymbol{P}$ which is assumed to depend on a parameter vector $\boldsymbol{\theta}$, say $\boldsymbol{P}(\boldsymbol{\theta})$, whereas, as before, no structure is imposed on the population mean vector $\boldsymbol{\mu}$. Correlation structure analysis is often chosen when the observed variables have different and arbitrary scales. In this case, researchers may feel that it is more meaningful to transform the observed variables to standard deviation scales. In contrast, when all observed variables are on the same scale, researchers may feel that it is more appropriate to fit a covariance structure.

It is not the aim of this paper to elaborate on when to perform covariance vs. correlation structure analysis. Rather, this paper aims at discussing the case in which a researcher wishes to estimate a covariance structure but s/he is unable to do so from a sample covariance matrix because only a sample correlation matrix is available for analysis. Estimating a covariance structure from a sample correlation matrix is not a trivial matter. Cudeck (1989) thoroughly reviewed this topic pointing out that doing this may result in (a) fitting a different model that the one intended, (b) incorrect $\chi^2$ and other goodness-of-fit measures, and (c) incorrect standard errors. Given these problems, one should estimate a covariance structure from a sample covariance matrix if at all possible. However, in some cases is not possible. For instance

a) When all observed variables are categorical. Although covariance structure

analysis was originally developed as a technique for continuous variables, over the last fifteen years the most popular software packages for structural equation modeling (LISREL: Jöreskog & Sörbom, 1993; MPLUS: Muthén & Muthén, 1998; EQS: Bentler, 1995) have incorporated routines for performing covariance structure analysis for categorical dependent variables as well by assuming that these arise by discretizing a multivariate normal distribution according to a set of thresholds. Nevertheless, when all the observed variables are categorical, then the parameters of the underlying covariance structure can not be estimated from a sample covariance matrix, as only the correlation matrix of the underlying normal variates (a matrix of tetrachoric/polychoric correlations) may be estimated.

b) When all observed variables are continuous but only a correlation matrix of is available (e.g., when one is interested in estimating a covariance structure from a published correlation matrix). Since in this case only the correlation matrix is available, estimation must proceed under multivariate normal assumptions.

Clearly, the first instance will be encountered more frequently than the second, and correspondingly, it will be the main focus of the present research. The standard procedure to fit a covariance structure to categorical observed variables when no restrictions are imposed on the thresholds consists in estimating each sample threshold and polychoric correlation separately from the first and second order marginals of the observed contingency table. Then, the parameters of the underlying covariance structure are estimated from the sample tetrachoric/polychoric correlations alone using a weighted least squares discrepancy function. By using this approach, one can estimate the covariance model parameters, obtain asymptotically correct goodness-of-fit measures and standard errors for the parameter estimates. But, as we shall show, if and only if the covariance structure being fitted is scale invariant. If this procedure is used to estimate a covariance structure that is not scale invariant, then one ends up fitting a different (and more restricted) covariance structure

than the one intended. We shall also show that to fit covariance structure that is not scale invariant to categorical observed variables one must use in the final stage of the estimation procedure a weighted least squares discrepancy function using both the sample thresholds and tetrachoric/polychoric correlations. To illustrate our discussion, we shall provide a numerical example in which we fit scale invariant and non-scale invariant factor models to the well known LSAT 6 dataset (Bock & Lieberman, 1970).

Next, we shall discuss how to fit a covariance structure model to a sample correlation matrix of continuous variables. Covariance structure models can be estimated from a sample correlation matrix by minimizing a normal theory generalized least squares function of the sample correlations under normality assumption (Jennrich, 1970; Browne & Shapiro, 1990). However, this is actually not needed. One can estimate any covariance structure from a sample correlation matrix by minimizing a normal theory discrepancy function for sample covariances. This is convenient, because to our knowledge discrepancy functions for sample correlations have not been implemented in standard software packages such as LISREL, EQS or MPLUS. Unfortunately, no standard software package can currently estimate a non-scale invariant covariance structure from a sample correlation matrix. To illustrate our discussion of the continuous case we shall use some data originally published by Jöreskog (1978) and also considered by Cudeck (1989).

Because determining whether a model is scale invariant is critical in applications in which a covariance structure is estimated from a sample correlation matrix, we provide in an appendix computer algebra code in Mathematica (Wolfram, 1999) that may be employed to determine whether a covariance structure model is scale invariant using results from Bekker, Merckens and Wansbeek (1994). Also, because when estimating a covariance structure from a sample correlation matrix not all covariance structure parameters may be identified we provide in another appendix Mathematica computer algebra code to be used to investigate

the identification of the model parameters.

## 2. Covariance structure analysis for categorical dependent variables

Let $\mathbf{y}^* \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ and suppose that each variable $y_i^*$ has been categorized using

$$y_i = h \quad \text{if} \quad \alpha_{i_h} < y_i^* < \alpha_{i_{h+1}} \qquad h = 0, \cdots, k-1;\ i = 1, \ldots, n \tag{1}$$

where $\alpha_{i_0} = -\infty, \alpha_{i_k} = \infty$. That is, for notational ease, we assume that all variables $y_i$ have the same number of categories, $k$. Our objective is to estimate the $q$-dimensional parameter vector $\boldsymbol{\theta}$ from the observed categorical variables $\mathbf{y}$.

According to this model, the probability of observing any categorical pattern $\mathbf{y}_c$ is

$$\Pr(\mathbf{y}_c) = \int_{\mathbf{R}} \cdots \int \phi_n\left(\mathbf{y}^* : \boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right) d\mathbf{y}^* \qquad c = 1, \ldots, k^n \tag{2}$$

where $\phi_n(\bullet)$ denotes a $n$-variate normal density and the intervals of the area of integration $\mathbf{R}$ are $R_i = \left(\alpha_{i_h}, \alpha_{i_{h+1}}\right)$ if $y_i = h$.

Because the underlying variables $\mathbf{y}^*$ are normal, the pattern probabilities (2) are unchanged when we standardize each $y_i^*$ by subtracting its mean and dividing it by its standard deviation using

$$\mathbf{z}^* = \mathbf{D}_\theta\left(\mathbf{y}^* - \boldsymbol{\mu}\right) \qquad \mathbf{D}_\theta = \mathrm{Diag}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})\right)^{-\frac{1}{2}} \tag{3}$$

where Diag($\bullet$) denotes a square matrix whose non-diagonal elements have been set to 0. Denoting by $\sigma_{ii}(\boldsymbol{\theta})$ a diagonal element of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, the diagonal elements of $\mathbf{D}_\theta$ are of the type

$$\delta_i = \frac{1}{\sqrt{\sigma_{ii}(\boldsymbol{\theta})}} \tag{4}$$

As a result of (3), $\mathbf{z}^*$ has mean zero and correlation structure

$$\mathbf{P}(\boldsymbol{\theta}) = \mathbf{D}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D}_{\boldsymbol{\theta}} \tag{5}$$

i.e., $\boldsymbol{P}(\boldsymbol{\theta})$ has ones along its diagonal. Furthermore, defining

$$\tau_{i_h} := \delta_i\left(\alpha_{i_h} - \mu_i\right) \tag{6}$$

when we change the variable of integration in (2) using (3) we find that at $y_i^* = \alpha_{i_h}$, $z_i^* = \tau_{i_h}$. Thus, (2) can be equivalently written as

$$\Pr(\mathbf{y}_c) = \int_{\tilde{\mathbf{R}}} \cdots \int \phi_n\left(\mathbf{z}^* : \mathbf{0}, \mathbf{P}(\boldsymbol{\theta})\right) d\mathbf{z}^* \tag{7}$$

with intervals of integration $\tilde{R}_i = \left(\tau_{i_h}, \tau_{i_{h+1}}\right)$ if $y_i = h$, where $\tau_{i_0} = -\infty, \tau_{i_k} = \infty$.

Now, because (2) and (7) are equivalent, we see that only the correlation structure (5) can be identified (estimated) from categorical data. There is an additional identification problem in (6), namely, that the μ's can not be separately estimated from the α's. The easiest way to solve this identification problem is to assume in applications that $\boldsymbol{\mu} = \mathbf{0}$. We shall do so in the remainder of this paper. Note, however, that if we were to generate data according to this model with $\boldsymbol{\mu} \neq \mathbf{0}$, we would be estimating $\alpha_{i_h}^{\circ} = \alpha_{i_h} - \mu_i$ rather than $\alpha_{i_h}$

We shall now introduce some notation. Let $\boldsymbol{\alpha}_h = \left(\alpha_{1_h}, \cdots, \alpha_{n_k}\right)'$, $\boldsymbol{\alpha}' = \left(\boldsymbol{\alpha}_1', \cdots, \boldsymbol{\alpha}_{k-1}'\right)$, $\boldsymbol{\vartheta}' = \left(\boldsymbol{\alpha}', \boldsymbol{\theta}'\right)$ and $\boldsymbol{\tau}_h(\boldsymbol{\vartheta}) = \left(\tau_{1_h}(\boldsymbol{\vartheta}), \cdots, \tau_{n_k}(\boldsymbol{\vartheta})\right)'$, where from (6) and the identification restriction $\boldsymbol{\mu} = \mathbf{0}$,

$$\boldsymbol{\tau}_h(\boldsymbol{\vartheta}) = \mathbf{D}_{\boldsymbol{\theta}}\boldsymbol{\alpha}_h \tag{8}$$

Furthermore, let $\boldsymbol{\tau}'(\boldsymbol{\vartheta}) = \left(\boldsymbol{\tau}_1'(\boldsymbol{\vartheta}), \cdots, \boldsymbol{\tau}_{k-1}'(\boldsymbol{\vartheta})\right)$, and $\boldsymbol{\kappa}'(\boldsymbol{\vartheta}) = \left(\boldsymbol{\tau}'(\boldsymbol{\vartheta}), \boldsymbol{\rho}'(\boldsymbol{\vartheta})\right)$ where $\boldsymbol{\rho}(\boldsymbol{\vartheta})$ is obtained by stacking the lower diagonal elements of $\boldsymbol{P}(\boldsymbol{\vartheta})$ excluding the diagonal onto a column vector. Note that in fact $\boldsymbol{\rho}$ depends only on the covariance structure parameters $\boldsymbol{\theta}$,

see (5).

As pointed out in the introduction, standard software programs such as EQS, LISREL and MPLUS estimate $\theta$ using several stages (see Jöreskog, 1994; Lee, Poon & Bentler, 1995; Muthén, 1978, 1984, 1993; Muthén, du Toit & Spisic, in press; Muthén & Satorra, 1995). First, the sample thresholds $\tau$ are estimated from the first order marginals of the contingency table. Then, the polychoric correlations $\rho$ are estimated from second order marginals of the contingency table given the estimated sample thresholds.

Consider now the estimation of $\vartheta$ from the parameters estimated in the first two stages, $\hat{\kappa}' = (\hat{\tau}', \hat{\rho}')$. Before estimating the model parameters in the last stage using (9), however, we must investigate its identification. Most often, when estimating $\vartheta$ from $\hat{\kappa}$, $\theta$ will not be identified even if the covariance structure model $\Sigma(\theta)$ is identified. Denoting by $\theta^*$ the subset of identified parameters in $\theta$, a general approach to estimate the identified model parameters $\vartheta^{*\prime} = \left(\alpha', \theta^{*\prime}\right)$ from $\hat{\kappa}$ is by minimizing

$$F_1(\vartheta) = \left(\hat{\kappa} - \kappa(\vartheta)\right)' \hat{\mathbf{W}} \left(\hat{\kappa} - \kappa(\vartheta)\right) \tag{9}$$

where $\hat{\mathbf{W}}$ is a matrix converging in probability to $\mathbf{W}$, a non-negative definite matrix, and from (5) and (8)

$$\tau_h\left(\vartheta^*\right) = \mathbf{D}_{\theta^*}\alpha_h \qquad \mathbf{P}\left(\vartheta^*\right) = \mathbf{D}_{\theta^*}\Sigma\left(\theta^*\right)\mathbf{D}_{\theta^*} \tag{10}$$

To use this general approach we need to be able to model simultaneously the thresholds and tetrachoric/polychoric correlations. In addition, we need to be able to enforce the complex non-linear constraints (4). MPLUS (Muthén & Muthén, 1998) can be used to do the former, but not the latter. LISREL (Jöreskog & Sörbom, 1993) and EQS (Bentler, 1995) only have capabilities for modeling tetrachoric/polychoric correlations.

Letting $\hat{\Xi}$ be a consistent estimate of the asymptotic covariance matrix of $\hat{\kappa}$, then,

obvious choices of $\hat{\mathbf{W}}$ in (9) are $\hat{\mathbf{W}} = \hat{\boldsymbol{\Xi}}^{-1}$ (WLS: Muthén, 1978), $\hat{\mathbf{W}} = diag\left(\hat{\boldsymbol{\Xi}}\right)^{-1}$ (DWLS: Muthén, du Toit & Spisic, in press), and $\hat{\mathbf{W}} = \mathbf{I}$ (ULS: Muthén, 1993). WLS estimation has asymptotically optimal properties (i.e., minimum variance) among the class of estimators (9). However, it has been found repeatedly in simulation studies (e.g., Muthén & Kaplan, 1992; Muthén, 1993; Reboussin & Liang, 1998) that unless the model is very small and the sample size very large WLS has an unacceptable small sample behavior. Furthermore, ULS and DWLS behave well in small samples (Muthén, 1993; Muthén et al., in press), the difference between the two being negligible (Maydeu-Olivares, in press).

Suppose now that the covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scale invariant. A covariance structure is scale invariant (e.g., Browne & Shapiro, 1991) if for any parameter vector $\boldsymbol{\theta}$ belonging to the parameter space $\boldsymbol{\Theta}$ and a diagonal matrix $\mathbf{D}_\delta$ with non-zero and distinct elements $\delta_{ii}$ one can find a parameter vector $\tilde{\boldsymbol{\theta}}$ belonging to $\boldsymbol{\Theta}$ such that

$$\boldsymbol{\Sigma}\left(\tilde{\boldsymbol{\theta}}\right) = \mathbf{D}_\delta \boldsymbol{\Sigma}(\boldsymbol{\theta}) \mathbf{D}_\delta \tag{11}$$

Since (8) is a special case of (11), when a covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scale invariant

(a) one can always find a parameter vector $\tilde{\boldsymbol{\theta}}$ satisfying $\mathbf{P}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}\left(\tilde{\boldsymbol{\theta}}\right)$, and

(b) exactly $n$ elements of $\tilde{\boldsymbol{\theta}}$ will not be identified because $\tilde{\boldsymbol{\theta}}$ must satisfy the constraint (Cudeck, 1989: p. 319)

$$\mathrm{Diag}\left(\boldsymbol{\Sigma}\left(\tilde{\boldsymbol{\theta}}\right)\right) = \mathbf{I} \tag{12}$$

Thus, when a covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scale invariant one can always find a subset of identified parameters in $\tilde{\boldsymbol{\theta}}$, say $\tilde{\boldsymbol{\theta}}^*$, such that (12) is satisfied. Then, letting $\tilde{\boldsymbol{\alpha}}_h := \mathbf{D}_\theta \boldsymbol{\alpha}_h$ and $\tilde{\boldsymbol{\vartheta}}^{*\prime} = \left(\tilde{\boldsymbol{\alpha}}', \tilde{\boldsymbol{\theta}}^{*\prime}\right)$,

$$\tau_k\left(\tilde{\vartheta}^*\right) = \tilde{\alpha}_k \qquad\qquad \mathbf{P}\left(\tilde{\vartheta}^*\right) = \boldsymbol{\Sigma}\left(\tilde{\theta}^*\right) \tag{13}$$

is equivalent to (10), where $\boldsymbol{\Sigma}\left(\tilde{\theta}^*\right)$ has ones along its diagonal and has the same functional form as the original covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

Thus, when $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scale invariant it is always possible to reparameterize $\vartheta^*$ as $\tilde{\vartheta}^*$, where the latter is greatly preferable from a computationally point of view. On the one hand, when the thresholds and polychoric correlations are parameterized as a function of $\tilde{\vartheta}^*$ one takes rid of the non-linear constraints (4). On the other hand, as there is a one to one relationship between the parameter vector $\tilde{\alpha}$ and $\boldsymbol{\tau}$, and as $\boldsymbol{\rho}$ depends only on $\tilde{\theta}^*$, one may estimate the (reparameterized) covariance structure parameters $\tilde{\theta}^*$ from the estimated tetrachoric/polychoric correlations $\hat{\rho}$ only by minimizing

$$F_2\left(\tilde{\theta}^*\right) = \left(\hat{\rho} - \rho\left(\tilde{\theta}^*\right)\right)' \hat{\tilde{W}} \left(\hat{\rho} - \rho\left(\tilde{\theta}^*\right)\right) \tag{14}$$

In Appendix 1 we show, following Muthén (1978, p. 554), that when $\tilde{\theta}^*$ is estimated using (14) with $\hat{\tilde{W}} = \left\{\hat{\Xi}_\rho^{-1}, diag\left(\hat{\Xi}_\rho\right)^{-1}, \text{ or } \mathbf{I}\right\}$, and $\hat{\Xi}_\rho$ is a consistent estimate of the asymptotic covariance matrix of $\hat{\rho}$, one obtains the same parameter estimates for $\tilde{\theta}^*$ than when (9) is minimized with respect to $\tilde{\vartheta}^*$ with $\hat{W} = \left\{\hat{\Xi}^{-1}, diag\left(\hat{\Xi}\right)^{-1}, \text{ or } \mathbf{I}\right\}$, respectively. Furthermore, $\hat{F}_1 = \hat{F}_2$. However, the estimates for $\tilde{\alpha}$ will be the same if ULS or DWLS is employed, but not when WLS is employed. LISREL (Jöreskog & Sörbom, 1993), MPLUS (Muthén & Muthén, 1998) and EQS (Bentler, 1995) all have capabilities for estimating a covariance structure from categorical data using a sequential procedure with (14) in the last stage.

In sum, scale invariance of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is a sufficient condition to estimate the parameter vector $\boldsymbol{\theta}$ only from the sample polychoric correlations. It is a necessary condition as well. Often times, we can turn $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ into a correlation structure by enforcing $\text{Diag}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})\right) = \mathbf{I}$. In so doing we are fitting to the data the model on the left hand side of (11). When $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is not

scale invariance, then the models on the left and right hand side of (11) are not equivalent and the model on the left hand side of (11) is a restrictive version of the model on the right hand side of (11). Hence, when $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is not scale invariant and we turn it into a correlation structure by enforcing $\text{Diag}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})\right) = \mathbf{I}$, we are actually fitting to the data at hand a different and more restrictive model than the one intended.

Because when $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scale invariant considerable computational gains are obtained in performing covariance structure analysis when all the observed variables are categorical, it becomes critical in applications to be able to assess whether $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scale invariant. In Appendix 2 we provide computer algebra code in Mathematica (Wolfram, 1999) that will enable researchers to determine whether $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is locally scale invariant.

We shall now apply this general theory to a particular class of covariance structures.

## 3. An application of the general theory: The common factor model

Consider the class of covariance structures implied by the common factor model,

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \tag{15}$$

where $\boldsymbol{\Psi}$ is a diagonal matrix. We shall assume that enough restrictions have been imposed on the model so that the covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is identified and that $\boldsymbol{\mu} = \mathbf{0}$ for identification purposes. The threshold and correlation structure implied by this model are by (8) and (5)

$$\tau_h(\vartheta) = \mathbf{D}_\theta \alpha_h \qquad \mathbf{P}(\vartheta) = \mathbf{D}_\theta \left(\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}\right)\mathbf{D}_\theta \tag{16}$$

where $\mathbf{D}_\theta = \text{diag}\left(\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}\right)^{-\frac{1}{2}}$. We shall now consider how to estimate the parameter vector $\boldsymbol{\theta}$ from the estimated thresholds and polychoric correlations.

One way of estimating any member of this class is to introduce enough restrictions in

$\boldsymbol{\theta}$ so that (16) is identified. The identified parameters, $\boldsymbol{\theta}^*$, are then estimated from $\hat{\boldsymbol{\kappa}}$ simultaneously with $\boldsymbol{\alpha}$ using (9).

Consider now the subset of models of (15) that are scale invariant. For these models

$$\tau_h\left(\tilde{\vartheta}\right) = \tilde{\alpha}_h \qquad \boldsymbol{\Sigma}\left(\tilde{\vartheta}\right) = \tilde{\boldsymbol{\Lambda}}\boldsymbol{\Phi}\tilde{\boldsymbol{\Lambda}}' + \tilde{\boldsymbol{\Psi}} \tag{17}$$

is equivalent to (16) where

$$\tilde{\alpha}_h = \mathbf{D}_\theta \alpha_h \qquad \tilde{\boldsymbol{\Lambda}} = \mathbf{D}_\theta \boldsymbol{\Lambda} \qquad \tilde{\boldsymbol{\Psi}} = \mathbf{D}_\theta \boldsymbol{\Psi} \mathbf{D}_\theta \tag{18}$$

To identify (17) and fulfill (12) we may simply let

$$\tilde{\boldsymbol{\Psi}} = \mathbf{I} - \mathrm{Diag}\left(\tilde{\boldsymbol{\Lambda}}\boldsymbol{\Phi}\tilde{\boldsymbol{\Lambda}}'\right) \tag{19}$$

Substituting (19) in (17), we obtain

$$\tau_h\left(\tilde{\alpha}_h\right) = \tilde{\alpha}_t \qquad \mathbf{P}\left(\tilde{\boldsymbol{\theta}}^*\right) = \tilde{\boldsymbol{\Lambda}}\boldsymbol{\Phi}\tilde{\boldsymbol{\Lambda}}' + \tilde{\boldsymbol{\Psi}} \tag{20}$$

Thus, in this case, one can estimate $\tilde{\boldsymbol{\theta}}^*$ in the last stage of the estimation procedure simply using (14). When the number of categories $k$ and the number of items $n$ are large this is greatly preferable from a computational viewpoint than to estimate all the identified parameters in $\vartheta$ using (9) with (16).

We shall now consider the results of estimating a covariance structure that is not scale invariant by introducing the constraints (19) to identify the model. When we use (20) with (19) we are effectively postulating that $\mathbf{P}_{y\cdot}$, rather than $\boldsymbol{\Sigma}_{y\cdot}$, has the parametric structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. When the model is scale invariant this has no effect as $\mathbf{P}_{y\cdot}$ and $\boldsymbol{\Sigma}_{y\cdot}$ have the same structure. However, when the model is not scale invariant $\mathbf{P}_{y\cdot}$ and $\boldsymbol{\Sigma}_{y\cdot}$ have different structures and thus fitting the model to $\mathbf{P}_{y\cdot}$ rather than to $\boldsymbol{\Sigma}_{y\cdot}$ results in fitting a more restrictive model. Another way to put it is to say that applying (19) with (20) implies

fitting $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ to the standardized variables $\mathbf{z}^*$, rather than to the unstandardized variables $\mathbf{y}^*$. Again, when $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scale invariant, it is irrelevant whether one imposes this structure on $\mathbf{z}^*$ or on $\mathbf{y}^*$. But when it is not scale invariant, however, the covariance structures of $\mathbf{z}^*$ and $\mathbf{y}^*$ have different parametric forms. Thus, when $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is not scale invariant, imposing this structure on $\mathbf{z}^*$ will always results in poorer fit that imposing the same structure on $\mathbf{y}^*$.

To illustrate the present discussion, consider a $n$-variate normal distribution $\mathbf{y}^*$ with mean zero that have been dichotomized via a threshold relationship (1). Note that since we are considering dichotomous variables, there is only one set of thresholds. The following four covariance structures for $\mathbf{y}^*$ will be considered

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Psi} \tag{21}$$

$$\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}\right) = \hat{\lambda}^2\mathbf{1}\mathbf{1}' + \hat{\boldsymbol{\Psi}} \tag{22}$$

$$\boldsymbol{\Sigma}\left(\breve{\boldsymbol{\theta}}\right) = \mathbf{1}\mathbf{1}' + \breve{\boldsymbol{\Psi}} \tag{23}$$

$$\boldsymbol{\Sigma}\left(\dot{\boldsymbol{\theta}}\right) = \dot{\lambda}^2\mathbf{1}\mathbf{1}' + \mathbf{I} \tag{24}$$

where all matrices $\boldsymbol{\Psi}$ are diagonal with elements $\psi_i$. The covariance structures (21) and (22) correspond to the well-known one factor and tau-equivalent models, respectively. Using the computer algebra code provided in Appendix 2, one may easily verify that (21) is scale invariant, whereas (22), (23), and (24) are not scale invariant.

By (16), the threshold and correlation structures corresponding to models (21) to (24) have elements

$$\tau_i(\boldsymbol{\vartheta}) = \frac{\alpha_i}{\sqrt{\lambda_i^2 + \psi_i}} \qquad \rho_{ii'}(\boldsymbol{\vartheta}) = \frac{\lambda_i\lambda_{i'}}{\sqrt{\lambda_i^2 + \psi_i}\sqrt{\lambda_{i'}^2 + \psi_{i'}}} \tag{25}$$

$$\tau_i\left(\hat{\vartheta}\right)=\frac{\hat{\alpha}_i}{\sqrt{\hat{\lambda}^2+\hat{\psi}_i}} \qquad \rho_{ii'}\left(\hat{\vartheta}\right)=\frac{\hat{\lambda}^2}{\sqrt{\hat{\lambda}^2+\hat{\psi}_i}\sqrt{\hat{\lambda}^2+\hat{\psi}_{i'}}} \tag{26}$$

$$\tau_i\left(\breve{\vartheta}\right)=\frac{\breve{\alpha}_i}{\sqrt{1+\breve{\psi}_i}} \qquad \rho_{ii'}\left(\breve{\vartheta}\right)=\frac{1}{\sqrt{1+\breve{\psi}_i}\sqrt{1+\breve{\psi}_{i'}}} \tag{27}$$

$$\tau_i\left(\dot{\vartheta}\right)=\frac{\dot{\alpha}_i}{\sqrt{\dot{\lambda}^2+1}} \qquad \rho_{ii'}\left(\dot{\vartheta}\right)=\frac{\dot{\lambda}^2}{\dot{\lambda}^2+1} \tag{28}$$

After introducing suitable (if any) identification constraints, any of these threshold and correlation structures can be estimated employing (9). In Appendix 3 we provide computer algebra code in Mathematica (Wolfram, 1999) that will enable users to determine whether these threshold and correlation structures are locally identified using results of Bekker, Merckens and Wansbeek (1994).

We shall first consider the one factor model (25). Using the code in Appendix 3 we find that $n$ constraints need to be introduced in this model for this structure to be identified. The constraint $\boldsymbol{\Psi}=\boldsymbol{I}$ identifies the model. One set of identified parameters is therefore $\vartheta^*=\left(\alpha_1,\cdots,\alpha_n,\lambda_1,\cdots,\lambda_n\right)'$ and we may rewrite (25) as

$$\tau_i\left(\vartheta^*\right)=\frac{\alpha_i}{\sqrt{\lambda_i^2+1}} \qquad \rho_{ii'}\left(\vartheta^*\right)=\frac{\lambda_i\lambda_{i'}}{\sqrt{\lambda_i^2+1}\sqrt{\lambda_{i'}^2+1}} \tag{29}$$

Now, because the one factor model is scale invariant using (19) and (20) we can reparameterize it as

$$\tau_i\left(\tilde{\vartheta}^*\right)=\tilde{\alpha}_i \qquad \rho_{ii'}\left(\tilde{\vartheta}^*\right)=\tilde{\lambda}_i\tilde{\lambda}_{i'} \tag{30}$$

with $\tilde{\boldsymbol{\Psi}}=\mathbf{I}-diag\left(\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\Lambda}}'\right)$. The parameterization (30) is considerably more convenient than (29) because the parameters of the covariance structure can be estimated in the third stage as a correlation structure problem using (14) rather than as a threshold and correlation structure problem using (9). Furthermore, the non-linear restrictions in (30) are considerably

simpler than in (29). The relationship between the parameterizations (29) and (30) is given by

$$\tilde{\alpha}_i = \frac{\alpha_i}{\sqrt{\lambda_i^2 + 1}} \qquad \tilde{\lambda}_i = \frac{\lambda_i}{\sqrt{\lambda_i^2 + 1}} \tag{31}$$

Consider now the tau-equivalent model (26). Using computer algebra we find that just one constraint needs to be introduced in this model to identify it. The constraint $\hat{\lambda} = 1$ identifies the model. Alternatively, the constraint $\hat{\psi}_i = 1$ also identifies the model. If we use $\hat{\lambda} = 1$ to identify the model, (26) becomes

$$\tau_i\left(\hat{\vartheta}^*\right) = \frac{\hat{\alpha}_i}{\sqrt{1 + \hat{\psi}_i}} \qquad \rho_{ii'}\left(\hat{\vartheta}^*\right) = \frac{1}{\sqrt{1 + \hat{\psi}_i}\sqrt{1 + \hat{\psi}_{i'}}} \tag{32}$$

which is identical to (27). But if we substitute

$$\tilde{\alpha}_i = \frac{\alpha_i}{\sqrt{1 + \hat{\psi}_i}} \qquad \tilde{\lambda}_i = \frac{1}{\sqrt{1 + \hat{\psi}_i}} \tag{33}$$

into (30), we also see that (27) is equivalent to (30). Thus, the covariance structures (21), (22) and (23) are equivalent when only categorical data is observed. This is a remarkable result.

We shall now consider the results of applying (20) with (19) to a covariance structure that is not scale invariant, such as the tau-equivalent covariance structure (22), in order to estimate it as a correlation structure only via (14). In this case, letting $\tilde{\tilde{\mathbf{\Psi}}} = \mathbf{I} - diag\left(\tilde{\tilde{\lambda}}^2 \mathbf{1}\mathbf{1}'\right)$, we would estimate a threshold and correlation structure with elements

$$\tau_i\left(\tilde{\tilde{\vartheta}}^*\right) = \tilde{\tilde{\alpha}}_i \qquad \rho_{ii'}\left(\tilde{\tilde{\vartheta}}^*\right) = \tilde{\tilde{\lambda}}^2 \tag{34}$$

Clearly, (30) and (34) are not equivalent models. Thus, applying (20) with (19) to estimate a covariance structure that is not scale invariant from a sample correlation matrix results in

estimating a different, more restrictive, model than the one intended. To what covariance structure for $\mathbf{y}^*$ corresponds the threshold and correlation structure (34)? Consider the covariance structure (24). It can be readily verified by substituting

$$\tilde{\dot{\alpha}}_i = \frac{\dot{\alpha}_i}{\sqrt{\dot{\lambda}^2 + 1}} \qquad \tilde{\dot{\lambda}}^2 = \frac{\dot{\lambda}^2}{\dot{\lambda}^2 + 1} \tag{35}$$

into (34) that (34) and (28) are equivalent, and therefore that by fitting (34) we are actually estimating the covariance structure (24).

We shall now provide a numerical example to illustrate our discussion. The covariance structures (21) to (24) will be fitted to a small binary dataset. We chose the well studied LSAT 6 dataset (Bock & Lieberman, 1970) for this example. This dataset consists of 1000 observations on 5 binary variables.

The following table summarizes the covariance structures fitted, the parameterization employed in their threshold and correlation structure, and how they were estimated in the last stage of the sequential procedure employed.

| Model | Covariance structure | Parameterized as | Estimated in the third stage using |
|---|---|---|---|
| A | (21) | (29) | (9) |
| B | (21) | (30) | (14) |
| C | (22) | (32) | (9) |
| D | (24) | (35) | (9) |
| E | (24) | (34) | (14) |

To estimate these models, the elements of $\boldsymbol{\kappa} = (\boldsymbol{\tau}, \boldsymbol{\rho})'$ and their asymptotic covariance matrix $\boldsymbol{\Xi}$ were estimated as in Muthén (1978). Parameter estimates, their asymptotic standard errors and goodness of fit tests for the structural restrictions $\boldsymbol{\kappa}(\hat{\boldsymbol{\theta}})$ were

obtained employing DWLS in the third stage as in Muthén, du Toit and Spisic (in press).

The parameter estimates and standard errors for these models are shown in Table 1.

---

Insert Table 1 about here

---

The so called models {A, B, C} are equivalent (they are just reparameterizatios of each other) and so are models {D, E}. The Satorra-Bentler's scaled statistic for assessing the structural restrictions imposed on the threshold and correlation structures by models {A, B, C} is $T_s = 4.741$, 5 d.f., $p = 0.448$, and for models {D, E} is $T_s = 5.269$, 9 d.f., $p = 0.810$. A nested test (Satorra and Bentler, 1999) reveals that the less restricted models {A, B, C} do not fit significantly better these data than the more restricted ones: $T_{dif} = 0.856$, 4 d.f., $p = 0.931$. Furthermore, one can verify in Table 1 the equivalencies among the models: Parameter estimates for models A and B are related by (31), for models B and C by (33), and for models D and E are by (35).

## 4. Estimating a covariance structure model from a sample correlation matrix of continuous variables

When a covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is to be estimated from a sample correlation matrix one must obtain the population correlation structure associated with the covariance structure. We saw in previous sections that there are two ways to do this: By using scaling constraints

$$\mathbf{P}(\boldsymbol{\theta}) = \mathbf{D}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D}_{\boldsymbol{\theta}} \qquad \mathbf{D}_{\boldsymbol{\theta}} = \mathrm{Diag}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})\right)^{-\frac{1}{2}} \tag{36}$$

or by reparameterization

$$\mathbf{P}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}\left(\tilde{\boldsymbol{\theta}}\right) \tag{37}$$

were one can employ (37) if and only if $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scale invariant, whereas (36) can be used to estimate a covariance structure from a sample correlation regardless of whether $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scale invariant or not. The application of (37) to estimate a covariance structure that is not scale invariant results in estimating a different and more restrictive covariance structure than intended. In any case, not all the parameters in $\boldsymbol{\theta}$ can be estimated, and the same number of identification constraints must be imposed if one uses (36) or (37) to estimate a scale invariant covariance structure. To identify (37) one simply needs to enforce $\text{Diag}\left(\boldsymbol{\Sigma}\left(\tilde{\boldsymbol{\theta}}\right)\right) = \mathbf{I}$, whereas identifying (36) is more complex and we provide computer algebra code in Appendix 3 to do so.

In sum, because in general estimating a covariance structure from a sample correlation matrix (a) requires enforcing complex constraints among the covariance structure parameters, and (b) not all the parameters of the covariance structure can be estimated, one should not estimate a covariance structure from a correlation matrix unless one is forced to do so because only the sample correlation matrix is available. When only the sample correlation matrix among the observed continuous variables is available, then estimation must proceed under multivariate normal assumptions.

One can estimate the identified subset of $\boldsymbol{\theta}$ or of $\tilde{\boldsymbol{\theta}}$ by minimizing a normal theory (NT) generalized least squares (GLS) discrepancy function for sample correlations (Jennrich, 1970; Browne & Shapiro, 1990). To our knowledge this discrepancy function has not been implemented in any standard software package for covariance structure analysis. Fortunately, it is not needed to employ a NT discrepancy function for sample correlations to correctly estimate a covariance structure from sample correlations. One may simply employ a NT discrepancy function for sample covariances provided (a) the degrees of freedom are

correctly computed as $\frac{n(n-1)}{2} - q^*$ where $q^*$ is the number of identified parameters, and (b) one imposes the constraints among the identified parameters $\mathbf{D}_{\theta} = \text{Diag}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})\right)^{-\frac{1}{2}}$ if (36) is employed, or $\text{Diag}\left(\boldsymbol{\Sigma}\left(\tilde{\boldsymbol{\theta}}\right)\right) = \mathbf{I}$ if (37) is employed.

Both LISREL and MPLUS can be used to fit scale invariant covariance structures to a sample correlation matrix by using (37) enforcing $\text{Diag}\left(\boldsymbol{\Sigma}\left(\tilde{\boldsymbol{\theta}}\right)\right) = \mathbf{I}$ and a NT discrepancy function for sample covariances. To our knowledge, the current version of EQS can not enforce constraints $\text{Diag}\left(\boldsymbol{\Sigma}\left(\tilde{\boldsymbol{\theta}}\right)\right) = \mathbf{I}$ and hence, it can not be used to correctly estimate a covariance structure from a sample correlation matrix of continuous variables. Neither LISREL, MPLUS, nor EQS can enforce the complex non-linear constraints implied by (36), and hence, these programs can not be used to estimate a non-scale invariant covariance structure from a sample correlation matrix.

To illustrate our present discussion numerically, we shall use a sample covariance matrix considered by Cudeck (1989) and originally published in Jöreskog (1978). The sample covariance matrix and its corresponding correlation matrix are given in Table 2.

---

Insert Table 2 about here

---

Consider a factor analysis covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$ with the following constraints:

*Model A*

$$\boldsymbol{\Lambda}' = \begin{pmatrix} \lambda_1 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & \lambda_4 \end{pmatrix} \tag{38}$$

*Model B*

$$\boldsymbol{\Lambda}' = \begin{pmatrix} \lambda_1 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & \lambda_2 \end{pmatrix} \tag{39}$$

For both models, $\mathbf{\Phi} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, and $\mathbf{\Psi} = \text{diag}\left(\psi_1, \cdots, \psi_4\right)$. Model A is scale invariant, whereas Model B is not. Both models are identified if estimated from sample covariances.

The following table summarizes the various submodels to be fitted.

| Submodel | Cov. structure | Associated corr. structure obtained by | Estimated from sample |
|---|---|---|---|
| A | A | -- | covariances |
| $A_1$ | A | scaling constraints | correlations |
| $A_2$ | A | reparameterization | correlations |
| B | B | --- | covariances |
| $B_1$ | B | scaling constraints | correlations |

Estimation in all cases was performed by minimizing a maximum likelihood discrepancy function for sample covariances. The resulting parameter estimates, standard errors and goodness of fit tests are shown in Table 3.

---

Insert Table 3 about here

---

Consider first Model A. To fit this model from sample correlations, we obtain its associated correlation structure using (36). Because the covariance structure is scale invariant, when estimating it from a sample correlation matrix exactly $n$ elements in $\boldsymbol{\theta}$ can not be estimated. Using the methods given in Appendix 3, we find that the constraint $\mathbf{\Psi} = \mathbf{I}$ identifies the model. This is submodel $A_1$. Because Model A is scale invariant when estimating it from sample covariances or correlations we obtain the same (a) goodness of fit, (b) parameter estimates and standard errors for scale free parameters (in this case for $\rho$) -see

Cudeck (1989). However, we obtain different parameter estimates for the elements of $\boldsymbol{\Lambda}$ in A and $A_1$ because these parameters are not scale free. In this example, because we have both the sample covariance and correlation matrices we can obtain the same parameter estimates for $\boldsymbol{\Lambda}$ using correlations than covariances if instead of fixing $\boldsymbol{\Psi} = \mathbf{I}$ when estimating the model from correlations, we fix these values at the values estimated using covariances. This is submodel $A_1^*$. Note that the standard errors for non-scale free parameters estimated from correlations are larger. Finally, because Model A is scale invariant, we can alternatively use the reparameterization approach (37) to fit it from sample correlations and estimate the reparameterized matrices of factor loadings (18) and uniquenesses (19).

Consider now Model B. Because it is not scale invariant, in can only be estimated from correlations using scaling constraints. Using the methods given in Appendix 3, we find that two constraints need to be introduced in the parameter vector $\boldsymbol{\theta}$ to estimate it from sample correlations. The constraints $\lambda_1 = 1, \lambda_2 = 1$ identify the model. This is model $B_1$. The goodness of fit of models B and $B_1$ are different because model $B_1$ is a constrained version of model B. In fact, $B_1$ is equivalent to models $A_1$ and $A_2$. That is, although Models A and B are distinct covariance structures, they have equivalent associated correlation structures.

## 6. Conclusions

When fitting a covariance structure from a sample correlation matrix one must consider the population correlation structure associated with it under the null hypothesis. This is obtained by pre and post-multiplying the covariance structure specified by the null hypothesis by a model-based diagonal matrix. That is, this diagonal matrix consists of the inverse of the square root of the diagonal of the covariance structure under consideration. As a result, in general, estimating a covariance structure from a sample correlation matrix requires estimating complicated non-linear functions of the covariance structure parameters.

However, it is well known (see for instance Cudeck, 1989) that if the covariance structure is scale invariant then one can find a reparameterization of this correlation structure that has the same functional form as the covariance structure specified by the null hypothesis. This reparameterization approach to estimate covariance structures is greatly preferable from a computational point of view, but it is only possible with scale invariant models. Furthermore, the goodness of fit indices obtained when estimating a covariance structure from sample correlations and from a sample covariances will be the same only if the covariance structure is scale invariant because not all the parameters of the covariance structure can be estimated from sample correlations. Hence the substantive conclusions a researcher may reach if s/he estimates a covariance structure that is NOT scale invariant from sample covariances or correlations may be different. Hence, assessing whether a covariance structure is scale invariant is critical in estimating it from a sample correlation matrix.

When all the observed variables are categorical these problems can not be avoided, as in this case one can only estimate a matrix of sample tetrachoric/polychoric correlations. Furthermore, we have shown that in this case the common practice of estimating the covariance structure parameters from a matrix of sample tetrachoric/polychoric correlations when no restrictions are imposed on the thresholds is admissible only if the covariance structure specified by the null hypothesis is scale invariant. Otherwise, one estimates a covariance structure that is more restrictive than that specified by the null hypothesis. We have also shown that to correctly estimate a covariance structure that is not scale invariant from categorical observed variables, one has to do so jointly from the sample thresholds and tetrachoric/polychoric correlations.

Appendix 1:

## Proof of the equivalence of (9) and (14) for scale invariant models

When $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scaled invariant, $\boldsymbol{\kappa}\left(\tilde{\boldsymbol{\vartheta}}^{*}\right)' = \left[\boldsymbol{\tau}(\tilde{\boldsymbol{\alpha}})', \boldsymbol{\rho}\left(\tilde{\boldsymbol{\theta}}^{*}\right)'\right]$. Now, letting $\mathbf{e}_1 := \hat{\boldsymbol{\tau}} - \boldsymbol{\tau}(\tilde{\boldsymbol{\alpha}})$, $\mathbf{e}_2 := \hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\left(\tilde{\boldsymbol{\theta}}^{*}\right)$, and partitioning $\mathbf{W}$ according to the partitioning of $\boldsymbol{\kappa}$, (9) may be rewritten as

$$F_1 = \begin{pmatrix} \mathbf{e}_1' & \mathbf{e}_2' \end{pmatrix} \begin{pmatrix} \hat{\mathbf{W}}_{11} & \hat{\mathbf{W}}_{21}' \\ \hat{\mathbf{W}}_{21} & \hat{\mathbf{W}}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix} = \mathbf{e}_1'\hat{\mathbf{W}}_{11}\mathbf{e}_1 + 2\mathbf{e}_1'\hat{\mathbf{W}}_{21}'\mathbf{e}_2 + \mathbf{e}_2'\hat{\mathbf{W}}_{22}\mathbf{e}_2 \tag{40}$$

Now, since there is a one-to-one relationship between $\boldsymbol{\tau}$ and $\tilde{\boldsymbol{\alpha}}$, from the first order condition for minimizing (14)

$$\frac{\partial F_1}{\partial \tilde{\boldsymbol{\alpha}}'} = \frac{\partial F_1}{\partial \boldsymbol{\tau}'} = 2\hat{\mathbf{W}}_{11}\mathbf{e}_1 + 2\hat{\mathbf{W}}_{21}'\mathbf{e}_2 = 0 \quad \Rightarrow \quad \mathbf{e}_1 = -\hat{\mathbf{W}}_{11}^{-1}\hat{\mathbf{W}}_{21}'\mathbf{e}_2 \tag{41}$$

and substituting this into (40), we obtain

$$F_1 = \mathbf{e}_2'\left(\hat{\mathbf{W}}_{22} - \hat{\mathbf{W}}_{21}\hat{\mathbf{W}}_{11}\hat{\mathbf{W}}_{21}'\right)\mathbf{e}_2 = \mathbf{e}_2'\left[\hat{\mathbf{W}}^{-1}\right]_{22}\mathbf{e}_2 := F_2 \tag{42}$$

where the last equality follows from a well-known result for the inverse of a partitioned matrix (e.g., Mardia, Kent & Bibby, p. 459). For instance, when $\hat{\mathbf{W}} = \hat{\boldsymbol{\Xi}}^{-1}$, $\hat{\mathbf{W}}^{-1} = \hat{\boldsymbol{\Xi}}$, and $\left[\hat{\mathbf{W}}^{-1}\right]_{22} = \boldsymbol{\Xi}_{22} := \boldsymbol{\Xi}_{\boldsymbol{\rho}}$.

Hence, since $\hat{F}_1 = \hat{F}_2$, when the covariance structure parameters $\tilde{\boldsymbol{\theta}}^{*}$ are estimated by minimizing $F_2$ the resulting parameter estimates and their standard errors will equal those obtained had these been estimated by minimizing $F_1$. If one is interested in estimating the threshold parameters $\tilde{\boldsymbol{\alpha}}$ after minimizing $F_2$, from (41) one may use

$$\tilde{\boldsymbol{\alpha}} = \hat{\boldsymbol{\tau}} + \hat{\mathbf{W}}_{11}^{-1}\hat{\mathbf{W}}_{21}{}'\left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\left(\tilde{\boldsymbol{\theta}}^{*}\right)\right) \tag{43}$$

This implies that $\hat{\tilde{\boldsymbol{\alpha}}} = \hat{\boldsymbol{\tau}}$ if and only if $\hat{\mathbf{W}}$ is a diagonal matrix (that is for DWLS or ULS). Else, if one takes $\hat{\mathbf{W}} = \hat{\boldsymbol{\Xi}}^{-1}$, the estimates and standard errors of $\tilde{\boldsymbol{\alpha}}$ depend on the parameter estimates $\hat{\tilde{\boldsymbol{\theta}}}^{*}$.

## Appendix 2:

## Assessing local scale invariance using computer algebra

Assessing whether $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is scale invariant amounts to verifying if we can find an alternative parameter vector $\tilde{\boldsymbol{\theta}}$ such that (11) is satisfied, i.e., $\boldsymbol{\Sigma}\left(\tilde{\boldsymbol{\theta}}\right) = \mathbf{D}_{\delta}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D}_{\delta}$, under the additional conditions that (a) $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$ belong to the same parameter space $\boldsymbol{\Theta}$ and (b) the elements of the diagonal matrix $\mathbf{D}_{\delta}$ are non-zero and distinct elements. Most often $\boldsymbol{\Sigma}$ is a non-linear function of $\boldsymbol{\theta}$. In that case, it is very difficult to solve the system of non-linear equations (11) unless the model is small, even with the aid of software systems capable of performing symbolic computations, such as Mathematica (Wolfram, 1999).

Note, however, that $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is nested within $\mathbf{D}_{\delta}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D}_{\delta}$. Thus assessing scale invariance amounts to assessing whether two nested models are equivalent. To do so, we may apply a result due to Bekker et al. (1994: Section 2.8) by which, under appropriate regularity conditions, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\mathbf{D}_{\delta}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D}_{\delta}$ are <u>locally</u> equivalent (and hence $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ will be scale invariant) if and only if

$$\operatorname{rank}\left[\frac{\partial \operatorname{vecs}\left(\mathbf{D}_{\delta}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{D}_{\delta}\right)}{\partial\left(\boldsymbol{\theta}', \boldsymbol{\delta}'\right)}\right] = \operatorname{rank}\left[\frac{\partial \operatorname{vecs}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})\right)}{\partial \boldsymbol{\theta}'}\right] + n \tag{44}$$

where $n$ is the number of observed variables, and vecs(•) denotes a column vector obtained by stacking the lower triangular elements of a matrix, including the diagonal, into a column vector.

Condition (44) can be very easily verified using a software package with symbolic computational capabilities, often for large models. Consider the covariance structure models A and B described in Section 4. We shall now provide some very simple Mathematica code to assess whether these models are scale invariant using (44). The code consists of four parts.

We first need the following function definitions

```
T[matrix_List] := Transpose[matrix]
L[matrix_List] := Length[matrix]
Diag[matrix_List] := Table[If[i == j, matrix[[i, j]], 0], {i, L[matrix]}, {j, L[matrix]}]
VecLow[matrix_List] := Flatten[MapIndexed[Take[#1, First[#2] - 1] &, matrix]]
VecLowDiag[matrix_List] := Flatten[MapIndexed[Take[#1, First[#2]] &, matrix]]
VecDiag[matrix_List] := Table[matrix[[i, i]], {i, Length[matrix]}]
```

(45)

where VecDiag(•), VecLow(•), and VecLowDiag(•) vectorize the diagonal, below the diagonal, and below and diagonal elements of a matrix, respectively. Diag(•) simply sets the off-diagonal elements of a matrix equal to zero.

The second block of the program simply constructs $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. For model A, this would simply be

```
n = 4;
la = {{l1,0},{l2,0},{0,l3},{0,l4}};
phi = {{1,r},{r,1}};
psi = DiagonalMatrix[Table[ToExpression["ps" <> ToString[i]], {i, n}]]
sigma = la . phi . T[la] + psi;
```

(46)

The third block of the program constructs $\text{vecs}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})\right)$ and $\boldsymbol{\theta}$. The latter is accomplished by vectorizing $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, putting them together and dropping constants and repeated parameters.

```
omega=VecLowDiag[sigma];
Print["This is the parameter vector theta"]
theta =Cases[Union[Flatten[la],VecLowDiag[phi],VecDiag[psi]],_Symbol]
```

(47)

Finally, the fourth block constructs $\Im_1 = \dfrac{\partial\, \text{vecs}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})\right)}{\partial \boldsymbol{\theta}'}$, $\Im_2 = \dfrac{\partial\, \text{vecs}\left(\mathbf{D}_\delta \boldsymbol{\Sigma}(\boldsymbol{\theta}) \mathbf{D}_\delta\right)}{\partial\left(\boldsymbol{\theta}', \boldsymbol{\delta}'\right)}$, and informs the user of whether $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is (locally) scale invariant or not by verifying

$$\operatorname{rank}\left(\aleph\left(\Im_2\right)\right) = \operatorname{rank}\left(\aleph\left(\Im_1\right)\right) + n$$

where $\aleph(\Im)$ denotes a basis for the nullspace of the Jacobian matrix $\Im$.

```
j =Outer[D,omega,theta];
lnu1=L[NullSpace[j]];
d=DiagonalMatrix[Table[ToExpression["d"<>ToString[i]],{i,n}]];
j2 =Outer[D,VecLowDiag[d . sigma . d],Join[theta,VecDiag[d]]];
lnu2=L[NullSpace[j2]];
If[lnu1 + n == lnu2,Print["The covariance structure is scale invariant"],Print["The covariance structure
is NOT scale invariant"]]
```

(48)

Using (45), (46), (47) and (48) one may readily verify model A is scale invariant but model B is not.

## Appendix 3:

## Assessing local model identification using computer algebra

Following Bekker et al. (1994) a necessary and sufficient condition (under appropriate regularity conditions) for the local identification of $\boldsymbol{\vartheta}$ in the parametric structure $\boldsymbol{\kappa}(\boldsymbol{\vartheta})$ is that the Jacobian matrix $\Im = \frac{\partial \boldsymbol{\kappa}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}'}$ be of full column rank. This condition may be verified by constructing a basis for the nullspace of $\Im$, say $\aleph$, such that $\aleph\Im' = 0$, and checking that $\aleph$ is an empty set. Whenever the model is not identified, the number of constraints we need to introduce in the parameter vector $\boldsymbol{\vartheta}$ will be given by the rank of $\aleph$. Furthermore, a zero column in $\aleph$ indicates an identified parameter.

We shall now provide some very simple Mathematica code to assess whether a threshold and correlation structure is locally identified using these results. We shall apply it to investigate the identification of the tau-equivalent covariance structure (22) for binary data. The code consists of four blocks.

The first block is simply (45). The second block constructs the threshold and correlation structure of the model of interest. In this case, it would be

```
n = 4;
la = Table[l, {n}, {1}];
psi = DiagonalMatrix[Table[ToExpression["ps" <> ToString[i]], {i, n}]];
sigma = la . T[la]  + psi;
d = Inverse[Sqrt[Diag[sigma]]];
alpha = Table[ToExpression["a" <> ToString[i]], {i, n}];
t = d . alpha;
rho = d . sigma . d;
```

(49)

The third block constructs $\boldsymbol{\kappa}(\boldsymbol{\vartheta})$, prints out the parameter vector $\boldsymbol{\vartheta}$ and the degrees

of freedom of the threshold and correlation structure.

```
omega=Join[t, VecLow[rho]];
Print["This is the parameter vector theta"]                                    (50)
theta =Cases[Union[alpha,Flatten[la],VecDiag[psi]],_Symbol]
Print["The number of degrees of freedom is ", L[omega] -L[theta]]
```

Finally, the fourth block obtains the Jacobian matrix $\Im = \frac{\partial \kappa(\vartheta)}{\partial \vartheta'}$, finds a basis for its nullspace, and reports that the threshold and correlation structure is (locally) identified if $\aleph$ is an empty set. Otherwise, when $\aleph$ is not an empty set, it reports that the model is not identified, it yields a list of the non-identified parameters, and it reports how many restrictions must be enforced among them to identify the model.

```
j =Outer[D,omega,theta];
nu =Simplify[NullSpace[j]];                                                   (51)
If[L[nu]==0, Print["The model is identified"], Print["The model is NOT identified. ", L[nu], " constraint(s)
need to be introduced among these parameters"] &&
    Print[theta[[Complement[Range[L[theta]],Flatten[Position[T[nu],Table[0,{L[nu]}]]]]]]];
```

In this example, the program reports that none of the parameters is identified, and that one constraint must be introduced in the model to identified. At this point, one can check whether the estimated $\aleph$ is actually a basis of the nullspace of $\Im$ verifying that

```
Simplify[nu .T[ j]]
```

yields a zero matrix, or print $\aleph$ using MatrixForm[nu], which in this example yields,

$$\left( \frac{\alpha_1}{2\psi_4} \quad \cdots \quad \frac{\alpha_4}{2\psi_4} \quad \frac{\lambda}{2\psi_4} \quad \frac{\psi_1}{\psi_4} \quad \cdots \quad \frac{\psi_3}{\psi_4} \quad 1 \right)$$

Finally, one can fix one of the non-identified parameters, say $\lambda = 1$, and re-run the program to verify that the model is identified for any number of observed variables $n$. A word of

caution. Because of the non-linear constraints (4), finding a basis for the nullspace in these models requires considerable computer resources unless the model is small.

References

Bekker, P.A., Merckens, A., & Wansbeek, T.J. (1994). Identification, equivalent models and computer algebra. San Diego: Academic Press.

Bentler, P.M. (1995). EQS 5. Encino, CA: Multivariate Software.

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

Browne, M.W. & Shapiro, A. (1991). Invariance of covariance structures under groups of transformations. Metrika, 38, 345-355.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. Psychological Bulletin, 105, 317-327.

Jennrich, R.I. (1970). An asymptotic chi-square test for the equality of two correlation matrices. Journal of the American Statistical Association, 65, 904-912.

Jöreskog, K.G. (1978). Structural analysis of covariance and correlation matrices. Psychometrika, 43, 443-477.

Jöreskog, K.G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. Psychometrika, 59, 381-390.

Jöreskog, K.G. & Sörbom, D. (1993). LISREL 8. User's reference guide. Chicago, IL: Scientific Software.

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.

Mardia, KV., Kent, J.T. & Bibby, J.M. (1979). Multivariate analysis. London: Academic Press.

Maydeu-Olivares, A. (in press). Limited information estimation and testing of Thurstonian models for paired comparison data under multinomial sampling. Psychometrika

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. Psychometrika, 43, 551-560.

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. Newbury Park, CA: Sage.

Muthén, B., du Toit, S.H.C. & Spisic, D. (in press). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Psychometrika.

Muthén, B. & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. British Journal of Mathematical and Statistical Psychology, 45, 19-30.

Muthén, L. & Muthén, B. (1998). 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.

Reboussin, B.A. & Liang, K.Y. (1998). An estimating equations approach for the LISCOMP model. Psychometrika, 63, 165-182.

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.

Satorra, A. & Bentler, P.M. (1999). A scaled difference chi-square test statistic for moment structure analysis. UPF Economic Working Papers Series # 412.

Shapiro, A. & Browne, M.W. (1990). On the treatment of correlation structures as covariance structures. Linear Algebra and its Applications, 127, 567-587.

Wolfram, S. (1999). Mathematica 4. Champaign, IL: Wolfram Research.

Table 1

Parameter estimates and standard errors for some models applied to the LSAT 6 data

| Parameter | Model A | Model B | Model C | Model D | Model E |
|---|---|---|---|---|---|
| $\alpha_1$ | -1.555 | -1.433 | -3.678 | -1.432 | -1.563 |
| | (0.100) | (0.059) | (1.072) | (0.059) | (0.067) |
| $\alpha_2$ | -0.600 | -0.550 | -1.386 | -0.550 | -0.600 |
| | (0.051) | (0.042) | (0.310) | (0.042) | (0.046) |
| $\alpha_3$ | -0.151 | -0.133 | -0.283 | -0.133 | -0.145 |
| | (0.046) | (0.040) | (0.100) | (0.040) | (0.043) |
| $\alpha_4$ | -0.773 | -0.716 | -1.900 | -0.716 | -0.781 |
| | (0.054) | (0.044) | (0.437) | (0.044) | (0.049) |
| $\alpha_5$ | -1.199 | -1.126 | -3.290 | -1.127 | -1.229 |
| | (0.067) | (0.050) | (0.909) | (0.050) | (0.057) |
| $\lambda_1$ | 0.423 | 0.389 | 1[a] | 0.400[c] | 0.436[c] |
| | (0.143) | (0.112) | | (0.031) | (0.041) |
| $\lambda_2$ | 0.433 | 0.397 | 1[a] | 0.400[c] | 0.436[c] |
| | (0.107) | (0.083) | | (0.031) | (0.041) |
| $\lambda_3$ | 0.534 | 0.471 | 1[a] | 0.400[c] | 0.436[c] |
| | (0.128) | (0.088) | | (0.031) | (0.041) |
| $\lambda_4$ | 0.407 | 0.377 | 1[a] | 0.400[c] | 0.436[c] |
| | (0.105) | (0.083) | | (0.031) | (0.041) |
| $\lambda_5$ | 0.364 | 0.342 | 1[a] | 0.400[c] | 0.436[c] |
| | (0.112) | (0.093) | | (0.031) | (0.041) |
| $\psi_1$ | 1[a] | 0.848[b] | 5.593 | 0.840[b] | 1[a] |
| | | (0.087) | (3.781) | (0.025) | |
| $\psi_2$ | 1[a] | 0.842[b] | 5.342 | 0.840[b] | 1[a] |
| | | (0.066) | (2.640) | (0.025) | |
| $\psi_3$ | 1[a] | 0.778[b] | 3.502 | 0.840[b] | 1[a] |
| | | (0.083) | (1.683) | (0.025) | |
| $\psi_4$ | 1[a] | 0.858[b] | 6.039 | 0.840[b] | 1[a] |
| | | (0.063) | (3.102) | (0.025) | |
| $\psi_5$ | 1[a] | 0.883[b] | 7.531 | 0.840[b] | 1[a] |
| | | (0.064) | (4.621) | (0.025) | |

Notes: [a] parameter fixed for identification purposes; [b] parameter constrained to be a function of other parameters for identification purposes; [c] parameter constrained to be equal across variables.

Table 2

Sample covariance and correlation matrices for some vocabulary tests

| | 15 items, untimed | 15 items, timed | 75 items, untimed | 75 items, timed |
|---|---|---|---|---|
| 15 items, untimed | 86.40 | .67 | .62 | .64 |
| 15 items, timed | 57.78 | 86.26 | .65 | .65 |
| 75 items, untimed | 56.87 | 59.32 | 97.29 | .76 |
| 75 items, timed | 58.90 | 59.67 | 73.82 | 97.82 |

Notes: N = 649; covariances below the diagonal, correlations above the diagonal.

From "Structural Analysis of Covariance and Correlation Matrices" by K.G. Jöreskog, 1978, *Psychometrika.*

Table 3

Results of fitting some models to the vocabulary tests data

**Model A**

| | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\psi_4$ | $\rho$ | $\chi^2$ | d.f. | p-val. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 7.50 | 7.70 | 8.51 | 8.68 | 30.13 | 21.93 | 24.89 | 22.56 | 0.90 | 0.70 | 1 | 0.40 |
| | (0.32) | (0.32) | (0.33) | (0.33) | (2.47) | (2.43) | (2.36) | (2.35) | (0.02) | | | |
| $A_1$ | 1.37 | 1.49 | 1.71 | 1.83 | 1[a] | 1[a] | 1[a] | 1[a] | 0.90 | 0.70 | 1 | 0.40 |
| | (0.08) | (0.10) | (0.11) | (0.12) | | | | | (0.02) | | | |
| $A_1^*$ | 7.50 | 7.70 | 8.51 | 8.68 | 30.13[a] | 21.93[a] | 24.89[a] | 22.56[a] | 0.90 | 0.70 | 1 | 0.40 |
| | (0.45) | (0.49) | (0.53) | (0.58) | | | | | (0.02) | | | |
| $A_2$ | 0.81 | 0.83 | 0.86 | 0.88 | 0.35[b] | 0.31[b] | 0.26[b] | 0.23[b] | 0.90 | 0.70 | 1 | 0.40 |
| | (0.02) | (0.02) | (0.01) | (0.01) | (0.03) | (0.03) | (0.02) | (0.02) | (0.02) | | | |

**Model B**

| | $\lambda_1$ | $\lambda_2$ | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\psi_4$ | $\rho$ | $\chi^2$ | d.f. | p-val. |
|---|---|---|---|---|---|---|---|---|---|---|
| B | 7.60 | 8.59 | 29.71 | 27.39 | 24.41 | 23.05 | 0.90 | 1.28 | 3 | 0.74 |
| | (0.27) | (0.28) | (2.35) | (2.26) | (2.15) | (2.1) | (0.02) | | | |
| $B_1$ | 1[a] | 1[a] | 0.54 | 0.45 | 0.34 | 0.30 | 0.90 | 0.70 | 1 | 0.40 |
| | | | (0.07) | (0.06) | (0.04) | (0.04) | (0.02) | | | |

Notes: [a] parameter fixed for identification purposes; [b] parameter constrained to be a function of other parameters for identification purposes.