Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

STRUCTURAL EQUATION MODELING WITH ORDINAL VARIABLES

BY KARL G. JÖRESKOG
Uppsala University

The statistical models used in structural equation modeling are described. The estimation theory for these models is reviewed for the case when all variables are continuous. Estimation theory for the case when all observed variables are ordinal is developed. This involves fitting the structural equation model to a matrix of polychoric correlations by weighted least squares. The weight matrix is a consistent estimate of the inverse of the asymptotic covariance matrix of the polychoric correlations. The asymptotic covariance matrix of the estimated polychoric correlations is derived for the case when the thresholds are estimated from the univariate marginals and the polychoric correlations are estimated from the bivariate marginals for given thresholds. Computational aspects are also discussed.

1. Introduction. Structural equation models have proven useful in solving many substantive research problems in the social and behavioral sciences. Such models have been used in the study of macroeconomic policy formation, intergenerational occupational mobility, racial discrimination in employment, housing and earnings, studies of antecedents and consequences of drug use, scholastic achievement, evaluation of social action programs, voting behavior, studies of genetic and cultural effects, factors in cognitive test performance, consumer behavior, and many other phenomena.

Methodologically, the models have many names, including simultaneous equation systems, linear causal analysis, confirmatory factor analysis, path analysis, structural equation models, recursive and non-recursive models for cross-sectional and longitudinal data, and covariance structure models.

The basic ideas and methods of structural equation models are explained in Bollen (1989). Bibliographies on the theory and applications of structural equation models are found in Jöreskog and Sörbom (1989) and Austin and

---

AMS 1980 Subject Classifications: Primary 62H12, 62P25.

Key words and phrases: Ordinal variables, polychoric correlations, maximum likelihood, asymptotic covariance matrix.

Wolfle (1991). A number of different but almost equivalent model formulations for structural equation models have been considered by Jöreskog (1973, 1977, 1978, 1981), McDonald (1978), Bentler and Weeks (1980) and McArdle and McDonald (1984). The most commonly used is the LISREL model of Jöreskog (1977) and we shall follow this formulation here.

In its most general form, the LISREL model consists of a set of linear structural equations. Variables in the equation system may be either directly observed variables or unmeasured latent (theoretical) variables that are not observed but relate to observed variables. The model assumes that there is a "causal" structure among a set of latent variables, and that the observed variables are indicators or symptoms of the latent variables. Sometimes the latent variables appear as linear composites of observed variables, other times as intervening variables in a "causal chain". The LISREL methodology is particularly designed to accommodate models that include latent variables, measurement errors, and reciprocal causation.

Section 2 describes the statistical models used in structural equation modeling. The estimation of the model is reviewed in Section 3 for the case when all variables are continuous. Estimation theory for the case of ordinal observed variables is developed in Section 4.

2. Models. As there is seldom any interest in means of latent variables and intercept terms in the equations, it is assumed here that all variables, observed as well as latent, are measured in deviations from their means. The LISREL model may then be defined as follows.

Consider random vectors $\boldsymbol{\eta}' = (\eta_1, \eta_2, \cdots, \eta_m)$ and $\boldsymbol{\xi}' = (\xi_1, \xi_2, \cdots, \xi_n)$ of latent dependent and independent variables, respectively, and the following system of linear structural relations

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{1}$$

where $\boldsymbol{B}(m \times m)$ and $\boldsymbol{\Gamma}(m \times n)$ are coefficient matrices and $\boldsymbol{\zeta}' = (\zeta_1, \zeta_2, \cdots, \zeta_m)$ is a random vector of residuals (errors in equations, random disturbance terms). The element β_{ij} of $\boldsymbol{B}$ represents the direct effect of η_j on η_i and the element γ_{ij} of $\boldsymbol{\Gamma}$ represents the direct effect of ξ_j on η_i. It is assumed that $\boldsymbol{\zeta}$ is uncorrelated with $\boldsymbol{\xi}$ and that $\boldsymbol{I} - \boldsymbol{B}$ is non-singular.

Vectors $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are not observed, but instead vectors $\boldsymbol{y}' = (y_1, y_2, \cdots, y_p)$ and $\boldsymbol{x}' = (x_1, x_2, \cdots, x_q)$ are observed, such that

$$\boldsymbol{y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{2}$$

and

$$\boldsymbol{x} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}, \tag{3}$$

where $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are vectors of error terms (errors of measurement or measure-specific components). These equations represent the multivariate regressions of $\boldsymbol{y}$ on $\boldsymbol{\eta}$ and of $\boldsymbol{x}$ on $\boldsymbol{\xi}$, respectively. It is convenient to refer to $\boldsymbol{y}$ and $\boldsymbol{x}$ as the observed variables and $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ as the latent variables.

In summary, the full LISREL model is defined by the three equations,

Structural Equation Model :	$\boldsymbol{\eta} = \boldsymbol{B\eta} + \boldsymbol{\Gamma\xi} + \boldsymbol{\zeta}$
Measurement Model for $\boldsymbol{y}$:	$\boldsymbol{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon}$
Measurement Model for $\boldsymbol{x}$:	$\boldsymbol{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}$

with the assumptions,

1. $\boldsymbol{\zeta}$ is uncorrelated with $\boldsymbol{\xi}$
2. $\boldsymbol{\varepsilon}$ is uncorrelated with $\boldsymbol{\eta}$
3. $\boldsymbol{\delta}$ is uncorrelated with $\boldsymbol{\xi}$
4. $\boldsymbol{\zeta}$ is uncorrelated with $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$
5. $\boldsymbol{I}$ - $\boldsymbol{B}$ is non-singular.

Let $\boldsymbol{\Phi}(n \times n)$ and $\boldsymbol{\Psi}(m \times m)$ be the covariance matrices of $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$, respectively, let $\boldsymbol{\Theta}_\varepsilon(p \times p)$ and $\boldsymbol{\Theta}_\delta(q \times q)$ be the covariance matrices of $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$, respectively and let $\boldsymbol{\Theta}_{\delta\varepsilon}(q \times p)$ be the covariance matrix between $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$. Then it follows, from the above assumptions, that the covariance matrix $\boldsymbol{\Sigma}[(p+q) \times (p+q)]$ of $(\boldsymbol{y}', \boldsymbol{x}')'$ is

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Lambda}_y\boldsymbol{A}(\boldsymbol{\Gamma\Phi\Gamma}' + \boldsymbol{\Psi})\boldsymbol{A}'\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\varepsilon & \boldsymbol{\Lambda}_y\boldsymbol{A\Gamma\Phi\Lambda}_x' + \boldsymbol{\Theta}_{\delta\varepsilon}' \\ \boldsymbol{\Lambda}_x\boldsymbol{\Phi\Gamma}'\boldsymbol{A}'\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_{\delta\varepsilon} & \boldsymbol{\Lambda}_x\boldsymbol{\Phi\Lambda}_x' + \boldsymbol{\Theta}_\delta \end{pmatrix}, \tag{4}$$

where $\boldsymbol{A} = (\boldsymbol{I} - \boldsymbol{B})^{-1}$.

The elements of $\boldsymbol{\Sigma}$ are functions of the elements of the parameter matrices $\boldsymbol{\Lambda}_y, \boldsymbol{\Lambda}_x, \boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Theta}_\delta, \boldsymbol{\Theta}_\varepsilon$, and $\boldsymbol{\Theta}_{\delta\varepsilon}$. In applications, some of the elements of the parameter matrices are fixed equal to assigned values, others are equal but unknown and still others are free unknown parameters.

3. Estimation: Continuous Variables. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_t)$ be a vector of all independent parameters in the model. Then $\boldsymbol{\Sigma}$ in (4) is regarded as a function $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$. The data is assumed to be a random sample of cases (individuals) on which the observable variables have been actually observed or measured. From this data a sample covariance matrix $\boldsymbol{S}$ is computed and it is this matrix which is used to fit the model to the data and to test the model.

Following Browne (1984), the model is estimated by minimizing a fit function of the form

$$F(\boldsymbol{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{W}) = (1/2)(\boldsymbol{s} - \boldsymbol{\sigma})'\boldsymbol{W}^{-1}(\boldsymbol{s} - \boldsymbol{\sigma}), \tag{5}$$

where

$$\boldsymbol{s}' = (s_{11}, s_{21}, s_{22}, s_{31}, \cdots, s_{kk}),$$

is a vector of the elements in the lower half, including the diagonal, of the covariance matrix $\boldsymbol{S}$ of order $k \times k$ ($k = p + q$ is the number of observed variables in the model),

$$\boldsymbol{\sigma}' = (\sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{31}, \cdots, \sigma_{kk}), \tag{6}$$

is the vector of corresponding elements of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ reproduced from the model parameters $\boldsymbol{\theta}$, and $\boldsymbol{W}$ is any symmetric positive definite matrix. To estimate the model parameters $\boldsymbol{\theta}$, the fit function is minimized with respect to $\boldsymbol{\theta}$.

Let $\boldsymbol{\Delta}(s \times t) = \partial\boldsymbol{\sigma}/\partial\boldsymbol{\theta}'$, where $s = k(k+1)/2$. The gradient vector of F is

$$\partial F/\partial\boldsymbol{\theta} = -\boldsymbol{\Delta}'\boldsymbol{W}^{-1}(\boldsymbol{s} - \boldsymbol{\sigma}),$$

and the information matrix is

$$\boldsymbol{E} = E(\partial^2 F/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}') = \boldsymbol{\Delta}'\boldsymbol{W}^{-1}\boldsymbol{\Delta}. \tag{7}$$

These quantities can be evaluated at any admissible point $\boldsymbol{\theta}$ of the parameter space and can therefore be used in an iterative procedure to minimize F.

The family of fit functions (5) includes most of the fit functions that are used in practice, i.e., ULS, GLS, ML, DWLS, and WLS, see, e.g., Jöreskog and Sörbom (1989). Fit functions based on elliptic distributions have been developed by Bentler (1983), Browne (1984), and Shapiro and Browne (1987).

To obtain consistent estimates, any positive definite matrix $\boldsymbol{W}$ may be used. Under very general assumptions, if the model holds in the population and if $\boldsymbol{S}$ converges in probability to $\boldsymbol{\Sigma}$ as the sample size increases, any fit function of the form (5) with a positive definite $\boldsymbol{W}$ will give a consistent estimator of $\boldsymbol{\theta}$. In practice, numerical results obtained by one fit function are often close enough to the results that would be obtained by another fit function to give the same substantive interpretations of the results.

Further assumptions must be made, however, if one needs an asymptotically correct chi-square measure of goodness-of-fit and asymptotically correct standard errors of parameter estimates.

To clarify this a bit further, assume that $\boldsymbol{S}$ converges in probability to $\boldsymbol{\Sigma}_0$ as the sample size increases and let $\boldsymbol{\theta}_0$ be the value of $\boldsymbol{\theta}$ that minimizes $F(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{W})$, that is, $\boldsymbol{\theta}_0$ is the minimizing value when $\boldsymbol{S}$ equals the true population covariance matrix $\boldsymbol{\Sigma}_0$. We say that the model holds if $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. Furthermore, let $\hat{\theta}$ be the value of $\boldsymbol{\theta}$ that minimizes F for the given sample covariance matrix $\boldsymbol{S}$, and let $\boldsymbol{\Omega} = nA\mathrm{Cov}\,(\boldsymbol{s})$, where n is the sample size minus one. Then the asymptotic covariance matrix of $\hat{\theta}$ is given by (Browne, 1984, eq. 2.12a)

$$nA\mathrm{Cov}\,(\hat{\theta}) = \boldsymbol{E}_0^{-1}(\boldsymbol{\Delta}_0'\boldsymbol{W}^{-1}\boldsymbol{\Omega}\boldsymbol{W}^{-1}\boldsymbol{\Delta}_0)\boldsymbol{E}_0^{-1}, \tag{8}$$

where $\boldsymbol{\Delta}_0$ and $\boldsymbol{E}_0$ are $\boldsymbol{\Delta}$ and $\boldsymbol{E}$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, and $\hat{\boldsymbol{\sigma}}$ is obtained from the non-duplicated elements of $\boldsymbol{\Sigma}(\hat{\theta})$.

Furthermore, let $\widehat{\boldsymbol{\Delta}}$ and $\widehat{\boldsymbol{E}}$ be $\boldsymbol{\Delta}$ and $\boldsymbol{E}$ evaluated at $\boldsymbol{\theta} = \hat{\theta}$ and let $\widehat{\boldsymbol{\Omega}}$ be a consistent estimate of $\boldsymbol{\Omega}$. Then a consistent estimate of $n A\mathrm{Cov}\,(\hat{\theta})$ may be obtained by substituting $\widehat{\boldsymbol{\Delta}}$, $\widehat{\boldsymbol{\Omega}}$, and $\widehat{\boldsymbol{E}}$ for $\boldsymbol{\Delta}, \boldsymbol{\Omega}$, and $\boldsymbol{E}$ in (8).

To test the model, one can use

$$c = n(\boldsymbol{s} - \hat{\boldsymbol{\sigma}})'[\widehat{\boldsymbol{\Omega}}^{-1} - \widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\Delta}}(\widehat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\Delta}})^{-1})\widehat{\boldsymbol{\Delta}}'\widehat{\boldsymbol{\Omega}}^{-1}](\boldsymbol{s} - \hat{\boldsymbol{\sigma}}). \tag{9}$$

as a test statistic (Browne, 1984, eq. 2.20b). If the model holds and is identified, this is approximately distributed as χ^2 with $d = s - t$ degrees of freedom.

If the model does not hold, denote by $\widehat{\boldsymbol{\Sigma}}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, which is not equal to $\boldsymbol{\Sigma}_0$ and let $\hat{\boldsymbol{\sigma}}_0$ be the corresponding vector formed from the non-duplicated elements of $\widehat{\boldsymbol{\Sigma}}_0$. Then c is distributed as non-central χ^2 with $s - t$ degrees of freedom and non-centrality parameter (Browne, 1984, eq. 2.21b)

$$\lambda = n(\boldsymbol{\sigma}_0 - \hat{\boldsymbol{\sigma}}_0)'[\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0'\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}_0)^{-1})\boldsymbol{\Delta}_0'\boldsymbol{\Omega}^{-1}](\boldsymbol{\sigma}_0 - \hat{\boldsymbol{\sigma}}_0). \tag{10}$$

In practice, $\boldsymbol{\Omega}$ is not known and a $\boldsymbol{W}$ must be chosen. The usual way of choosing $\boldsymbol{W}$ in weighted least squares is to let $\boldsymbol{W}$ be a consistent estimate of the asymptotic covariance matrix $\boldsymbol{\Omega}$ of $\boldsymbol{s}$, i.e., $\boldsymbol{W} = \widehat{\boldsymbol{\Omega}}$. This yields efficient parameter estimates. In this case, we say that $\boldsymbol{W}^{-1}$ is a *correct weight matrix* and we may substitute $\boldsymbol{W}$ for $\widehat{\boldsymbol{\Omega}}$ in (9). Equation (9) then simplifies to

$$c = n(\boldsymbol{s} - \hat{\boldsymbol{\sigma}})'\boldsymbol{W}^{-1}(\boldsymbol{s} - \hat{\boldsymbol{\sigma}}), \tag{11}$$

which is $2n$ times the minimum value of the fit function. Also note that (8) becomes

$$n A\mathrm{Cov}\,(\hat{\theta})] = \boldsymbol{E}_0^{-1}, \tag{12}$$

In the case of a correct weight matrix, the asymptotic covariance matrix of the residuals $\boldsymbol{s} - \hat{\boldsymbol{\sigma}}$ is given by (cf. Bentler and Dijkstra, 1985, eqs. 1.7.4–1.7.5)

$$n A\mathrm{Cov}\,(\boldsymbol{s} - \hat{\boldsymbol{\sigma}}) = \boldsymbol{\Omega} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0'\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0'. \tag{13}$$

For the same case, Browne and Cudeck (1993) develop an unbiased estimate and a confidence interval for the non-centrality parameter λ and demonstrates how these can be used in model evaluation.

Once the validity of a model has been established, tests of structural hypotheses about the parameters $\boldsymbol{\theta}$ in the model can be developed. One can test hypotheses of the forms

- that certain θ's have particular values
- that certain θ's are equal
- that certain θ's are specified linear or nonlinear functions of other parameters.

Each of these types of hypotheses leads to a model with fewer parameters $\boldsymbol{\nu}$, where $\boldsymbol{\nu}$ $(u \times 1)$ is a subset of the parameters in $\boldsymbol{\theta}$, $u < t$. In conventional statistical terminology, the model with parameters $\boldsymbol{\nu}$ is called the *null hypothesis* H_0 and the model with parameters $\boldsymbol{\theta}$ is called the *alternative hypothesis* H_1. Let c_0 and c_1 be the value of c for models H_0 and H_1, respectively. The test statistic for testing H_0 against H_1 is then

$$D^2 = c_0 - c_1$$

which is used as χ^2 with $d = t - u$ degrees of freedom. The degrees of freedom can also be computed as the difference between the degrees of freedom associated with c_0 and c_1.

4. Estimation: Ordinal Variables. The theory developed above holds in the case that the observed variables are continuous and a sample covariance matrix $\boldsymbol{S}$ is used to estimate a covariance structure $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. In practice, correlation matrices are often analyzed rather than covariance matrices. This is particularly so when ordinal variables are used since these do not have any natural units of measurement.

Theory and applications of structural equation models when some or all of the observed variables are ordinal have been considered by several authors, for example, Muthén (1984), Lee, Poon, and Bentler (1990), Jöreskog (1990), and Aish and Jöreskog (1990). Typically the estimation of the model is done in two steps. The first step involves estimating polychoric, polyserial and other correlations for the observed variables. The second step estimates the parameters of the model by weighted least squares using a weight matrix which must be a consistent estimate of the asymptotic covariance matrix of the correlations estimated in the first step. Different formulas for the weight matrix have been given by Muthén (1984) and by Lee, Poon, and Bentler (1990). None of these is really feasible when the number of ordinal variables is large. For the case when all variables are ordinal, Jöreskog (1993) gives a procedure for estimating the asymptotic covariance matrix of polychoric correlations which is feasible and relatively straightforward even on computers with limited memory.

Equations (12) and (13) do not apply directly to a correlation matrix $\boldsymbol{R}$ instead of $\boldsymbol{S}$. Firstly, a correlation matrix has fixed ones in the diagonal, so that rows and columns of $\boldsymbol{W}$ corresponding to diagonal elements would be zero, which would render $\boldsymbol{W}$ singular and the fit function (5) indeterminate. Secondly, a procedure for estimating the asymptotic covariance matrix of sample variances and covariances does not apply to the correlations in $\boldsymbol{R}$. To resolve this problem it is suggested that the fit function (5) be replaced with

$$F(\boldsymbol{\theta}) = (\boldsymbol{r} - \boldsymbol{\rho})' \boldsymbol{W}_r^{-1} (\boldsymbol{r} - \boldsymbol{\rho}), \tag{14}$$

where

$$\boldsymbol{r} = (r_{21}, r_{31}, r_{32}, r_{41}, r_{42}, r_{43}, \cdots, r_{k,k-1}),$$

and

$$\rho' = (\rho_{21}(\boldsymbol{\theta}), \rho_{31}(\boldsymbol{\theta}), \rho_{32}(\boldsymbol{\theta}), \rho_{41}(\boldsymbol{\theta}), \cdots, \rho_{k,k-1}(\boldsymbol{\theta})),$$

where the $\rho_{ij}(\boldsymbol{\theta})$ are population correlations implied by the model as functions of $\boldsymbol{\theta}$.

The asymptotic covariance matrix $\boldsymbol{W}_r$ in (14) should be a consistent estimate of the asymptotic covariance matrix of $\boldsymbol{r}$. The diagonal elements of the correlation matrix *are not included* in this vector.

A small problem arises here because the fit function (14) is not a function of the diagonal elements and, as a consequence, parameters such as the diagonal elements of Θ_ε and Θ_δ in LISREL (Jöreskog and Sörbom, 1989) cannot be estimated directly. However, the error variances of a standardized observed variable can be estimated as one minus the estimated variance contribution of all the factors that influences the observed variable. Standard errors of these estimated error variances can be obtained by the delta method.

Observations on an ordinal variable are assumed to represent responses to a set of ordered categories, such as a five-category Likert scale. Here it is only assumed that a person who responds in one category has more of a characteristic than a person who responds in a lower category. Ordinal variables are not continuous variables and should not be treated as if they are. Ordinal variables do not have origins or units of measurements. Means, variances, and covariances of ordinal variables have no meaning. To use ordinal variables in structural equation models requires other techniques than those of the previous section.

For each ordinal variable z (z may be a y- or x-variable in the LISREL model), it is assumed that there is an underlying continuous variable z^* which is normally distributed with mean μ_{z^*} and variance $\sigma^2_{z^*}$. The assumption of normality is based on the belief that, for most attitudinal variables, most people are in the middle and fewer people are at the ends. The normality assumption can not be falsified at the univariate level, but for every pair of variables, Jöreskog and Sörbom (1989) provide a test of underlying bivariate normality. Quiroga (1992) studied the robustness of the procedures described here against departures from underlying normality.

We write $z = i$ to mean that z belongs to the ordered category i. The actual score values in the data may be arbitrary and are irrelevant as long as the ordinal information is retained. That is, low scores correspond to low-order categories of z which are associated with smaller values of z^* and high scores correspond to high-order categories which are associated with larger values of z^*.

The connection between z and z^* is

$$z = i \quad \Longleftrightarrow \quad \tau_{i-1} < z^* \leq \tau_i, \qquad i = 1, 2, \cdots, m,$$

where

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

are parameters called threshold values. With m categories, there are $m-1$ threshold parameters.

Since only ordinal information is available about z^*, the mean μ_{z^*} and variance $\sigma^2_{z^*}$ of z^* are usually not identified and are therefore set to zero and one, respectively. However, when the same ordinal variable is measured one or more times, as in longitudinal or panel studies and in multigroup studies, it is possible to estimate the means and variances of the underlying variables (relative to a fixed origin and scale) by specifying the thresholds to be the same for the same variable over time and/or groups. In the following we assume that $\mu_{z^*}=0$ and that $\sigma^2_{z^*}=1$. Otherwise, replace $\tau_i^{(z)}$ by $(\tau_i^{(z)}-\mu_{z^*})/\sigma_{z^*}$ in what follows, where $\tau_i^{(z)}$ is the ith threshold for variable z.

The parameters are estimated from the univariate and bivariate log-likelihoods. These log-likelihoods have the following general form

$$\log L=\sum_a n_a \ln \pi_a(\boldsymbol{\theta}),$$

where a runs over all cells of the marginal distribution, n_a is the frequency (count) in cell a, and $\pi_a(\boldsymbol{\theta})$ is the probability of cell a as a function of a parameter vector $\boldsymbol{\theta}$[1]. The information matrix for $\boldsymbol{\theta}$ is

$$\boldsymbol{E}=\sum_a (1/\pi_a)(\partial\pi_a/\partial\boldsymbol{\theta})(\partial\pi_a/\partial\boldsymbol{\theta})',$$

Maximizing $\log L$ with respect to $\boldsymbol{\theta}$ gives the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. The inverse of the information matrix $\boldsymbol{E}$ evaluated at the maximum of $\log L$ gives an estimate of the covariance matrix of $\hat{\boldsymbol{\theta}}$. The general theory is given in Jöreskog (1993).

Consider k ordinal variables $z_1, z_2, \ldots, z_k$ with $m_1, m_2, \cdots, m_k$ categories, respectively. Altogether there are $\sum_{i=1}^k m_i - k + k(k-1)/2$ parameters to be estimated, namely the thresholds $(\tau_1^{(g)}, \tau_2^{(g)}, \cdots, \tau_{m_g-1}^{(g)})$, $g=1,2,\cdots,k$, and the polychoric correlations ρ_{gh}, $h<g$. The parameters are usually estimated from the univariate and bivariate marginal likelihoods, that is, the thresholds are estimated from the univariate marginal distribution and the polychoric correlations from the bivariate marginal distributions for given thresholds; see Olsson (1979), Muthén (1984), and Jöreskog and Sörbom (1988). The univariate and bivariate marginal likelihoods all have the general form given in the previous paragraph.

Olsson (1979) considered the case $k=2$ and studied two methods for estimating the parameters:

(i) Estimate the thresholds and the polychoric correlation jointly from the bivariate marginal distribution.

---

[1] This is a different parameter vector than the $\boldsymbol{\theta}$ in $\Sigma(\theta)$.

(ii) Estimate the thresholds from the univariate marginal distribution and then the polychoric correlation from the bivariate marginal distribution for given thresholds.

In both methods, an iterative procedure must be used to estimate the parameters. Practical experience suggests that the two methods give almost identical estimates. Method (ii) is computationally simple and is used most often in practice. This is the method considered here. Method (i) would have the disadvantage that different estimates of thresholds for one variable may be obtained from different pairs of variables where this variable is included.

The model for the univariate marginal of variable g is

$$\pi_a^{(g)}(\boldsymbol{\theta}) = \int_{\tau_{a-1}^{(g)}}^{\tau_a^{(g)}} \phi(u)du, \tag{15}$$

where $\phi(u)$ is the standard normal density function. The parameter vector is

$$\boldsymbol{\theta} = \boldsymbol{\tau}_g = (\tau_1^{(g)}, \tau_2^{(g)}, \cdots, \tau_{m_g-1}^{(g)}).$$

Application of the general theory gives the maximum likelihood estimator $\hat{\boldsymbol{\tau}}_g$ of $\boldsymbol{\tau}_g$ with information matrix $\boldsymbol{E}_g$. The maximum likelihood estimator is given explicitly as

$$\hat{\tau}_a^{(g)} = \Phi^{-1}(p_1 + p_2 + \cdots + p_a), \qquad a = 1, \cdots, m_g - 1,$$

where Φ^{-1} is the inverse of the standard normal distribution function.

$\hat{\boldsymbol{\tau}}_g$ is asymptotically linear in the proportions $\boldsymbol{p}_g$ of the univariate marginal distribution of variable g, $\hat{\boldsymbol{\tau}}_g = \boldsymbol{B}_g' \boldsymbol{p}_g$, say, where $\boldsymbol{B}_g$ is a matrix of order $m_g \times m_g - 1$ (Jöreskog, 1993).

Let $\boldsymbol{\Psi}_{gh}$ be N times the asymptotic covariance matrix of $\hat{\boldsymbol{\tau}}_g$ and $\hat{\boldsymbol{\tau}}_h$. Then:

$$\boldsymbol{\Psi}_{gh} = \boldsymbol{B}_g' \boldsymbol{\pi}_{gh} \boldsymbol{B}_h, \tag{16}$$

since $\mathrm{Cov}\,(\boldsymbol{p}_g, \boldsymbol{p}_h) = \boldsymbol{\pi}_{gh} - \boldsymbol{\pi}_g \boldsymbol{\pi}_h'$, and $\boldsymbol{B}_g' \boldsymbol{\pi}_g = \mathbf{0}$ (Jöreskog , 1993), where $\boldsymbol{\pi}_{gh}$ is a matrix of population probabilities of the bivariate marginal distribution of variables g and h and $\boldsymbol{\pi}_g$ and $\boldsymbol{\pi}_h$ are vectors of population probabilities of the univariate marginal distributions of variables g and h. Equation (16) holds for $g \neq h$. It holds for $g = h$ as well, if $\boldsymbol{\pi}_{gh}$ in (16) is interpreted as the diagonal matrix with elements $\pi_1^{(g)}, \pi_2^{(g)}, \cdots, \pi_{m_g}^{(g)}$. Note that $\boldsymbol{\Psi}_{gg} = \boldsymbol{E}_g^{-1}$.

The model for the bivariate marginal of variables g and h is

$$\pi_{ab}^{(gh)}(\boldsymbol{\theta}) = \int_{\tau_{a-1}^{(g)}}^{\tau_a^{(g)}} \int_{\tau_{b-1}^{(h)}}^{\tau_b^{(h)}} \phi_2(u, v; \rho_{gh})\, du\, dv, \tag{17}$$

where $\phi_2(u, v; \rho)$ is the density function of the standardized bivariate normal distribution with correlation ρ. The parameter vector is

$$\boldsymbol{\theta} = \boldsymbol{\theta}_{gh} = (\tau_1^{(g)}, \tau_2^{(g)}, \cdots, \tau_{m_g-1}^{(g)}, \tau_1^{(h)}, \tau_2^{(h)}, \cdots, \tau_{m_h-1}^{(h)}, \rho_{gh}) = (\boldsymbol{\tau}_g, \boldsymbol{\tau}_h, \rho_{gh}), \tag{18}$$

consisting of the thresholds for the two variables and the polychoric correlation ρ_{gh}.

To estimate the polychoric correlation the bivariate log-likelihood

$$\log L(\rho, \hat{\boldsymbol{\tau}}_g, \hat{\boldsymbol{\tau}}_h) = N \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} p_{ab}^{(gh)} \ln \pi_{ab}^{(gh)}), \tag{19}$$

is maximized with respect to ρ *for given* $\hat{\boldsymbol{\tau}}_g$ and $\hat{\boldsymbol{\tau}}_h$. Here $p_{ab}^{(gh)}$ are the sample proportions in the bivariate marginal distribution for variables g and h. The value of ρ that maximizes $\log L$ is the estimate $\hat{\rho}_{gh}$ of the polychoric correlation ρ_{gh}. This estimate satisfies the equation

$$f(\hat{\rho}, \hat{\boldsymbol{\tau}}_g, \hat{\boldsymbol{\tau}}_h) = 0, \tag{20}$$

where $f = \partial \log L / \partial \rho$.

Olsson (1979, eq. 23) gives a complicated expression for the asymptotic covariance matrix of

$$\hat{\theta}_{gh} = (\hat{\boldsymbol{\tau}}_g, \hat{\boldsymbol{\tau}}_h, \hat{\rho}_{gh}). \tag{21}$$

From this one can obtain the regression of $\hat{\rho}_{gh}$ on $\hat{\boldsymbol{\tau}}_g$ and $\hat{\boldsymbol{\tau}}_h$. However, this regression can be obtained more directly as follows. Expanding f to linear terms, gives

$$a(\hat{\rho} - \rho) + \boldsymbol{b}'(\hat{\boldsymbol{\tau}}_g - \boldsymbol{\tau}_g) + \boldsymbol{c}'(\hat{\boldsymbol{\tau}}_h - \boldsymbol{\tau}_h) = 0, \tag{22}$$

where a =plim $\partial f / \partial \rho$, $\boldsymbol{b}$ =plim $\partial f / \partial \boldsymbol{\tau}_g$, and $\boldsymbol{c}$ =plim $\partial f / \partial \boldsymbol{\tau}_h$. The required "plims" are given by Olsson (1979). Equation (23) can be interpreted to give the asymptotic conditional mean of $\hat{\rho}_{gh}$ for given $\hat{\boldsymbol{\tau}}_g$ and $\hat{\boldsymbol{\tau}}_h$:

$$E(\hat{\rho}_{gh} \mid \hat{\boldsymbol{\tau}}_g, \hat{\boldsymbol{\tau}}_h) = \alpha_{gh} + \boldsymbol{\beta}'^{(gh)}_g \hat{\boldsymbol{\tau}}_g + \boldsymbol{\beta}'^{(gh)}_h \hat{\boldsymbol{\tau}}_h, \tag{23}$$

where α_{gh} is a regression intercept term and $\boldsymbol{\beta}_g^{(gh)} = -a^{-1}\boldsymbol{b}$ and $\boldsymbol{\beta}_h^{(gh)} = -a^{-1}\boldsymbol{c}$ are vectors of regression coefficients.

The estimated thresholds and polychoric correlations are all asymptotically linear in the sample proportions of the univariate and bivariate marginal distributions (Jöreskog 1993). Since these proportions are linear in all the sample proportions of the k-way contingency table, it follows that the joint distribution of all the estimated parameters is asymptotically normal.

Let

$$\hat{\boldsymbol{\tau}} = (\hat{\boldsymbol{\tau}}_1, \hat{\boldsymbol{\tau}}_2, \cdots \hat{\boldsymbol{\tau}}_k),$$

be the vector of all estimated thresholds, and let

$$\hat{\boldsymbol{\rho}} = (\hat{\rho}_{21}, \hat{\rho}_{31}, \hat{\rho}_{32}, \hat{\rho}_{41}, \hat{\rho}_{42}, \hat{\rho}_{43}, \cdots, \hat{\rho}_{k,k-1}),$$

be the vector of estimated polychoric correlations. The asymptotic covariance matrix of $\hat{\boldsymbol{\rho}}$ is obtained as (see Rao, 1965, eq.2b.3.6)

$$A\mathrm{Cov}\,(\hat{\boldsymbol{\rho}}) = A\mathrm{Cov}\,[E(\hat{\boldsymbol{\rho}} \mid \hat{\boldsymbol{\tau}})] + A\mathrm{Cov}\,(\hat{\boldsymbol{\rho}} \mid \hat{\boldsymbol{\tau}}). \tag{24}$$

The first term is obtained from the regressions of each $\hat{\rho}$ on the thresholds. Specifically, using (23), the covariance between $\hat{\rho}_{gh}$ and $\hat{\rho}_{ij}$ *due to variation in* $\hat{\boldsymbol{\tau}}$ *alone* is

$$\begin{aligned} NA\mathrm{Cov}\,(\hat{\rho}_{gh},\ \hat{\rho}_{ij}) =& \boldsymbol{\beta}_g^{'(gh)} \boldsymbol{\Psi}_{gi} \boldsymbol{\beta}_i^{(ij)} + \boldsymbol{\beta}_g^{'(gh)} \boldsymbol{\Psi}_{gj} \boldsymbol{\beta}_j^{(ij)} \\ &+ \boldsymbol{\beta}_h^{'(gh)} \boldsymbol{\Psi}_{hi} \boldsymbol{\beta}_i^{(ij)} + \boldsymbol{\beta}_h^{'(gh)} \boldsymbol{\Psi}_{hj} \boldsymbol{\beta}_j^{(ij)}. \end{aligned} \tag{25}$$

Conditional on $\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}$, it follows from the general theory in Jöreskog (1993) that asymptotically

$$(\hat{\rho}_{gh} - \rho_{gh}) \sim \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} u_{ab}^{(gh)} (p_{ab}^{(gh)} - \pi_{ab}^{(gh)}), \tag{26}$$

where

$$u_{ab}^{(gh)} = \frac{(1/\pi_{ab}^{(gh)}) \partial \pi_{ab}^{(gh)} / \partial \rho_{gh}}{\sum_{a=1}^{m_g} \sum_{b=1}^{m_h} (1/\pi_{ab}^{(gh)}) (\partial \pi_{ab}^{(gh)} / \partial \rho_{gh})^2}. \tag{27}$$

Similar expressions hold for variables i and j. Hence,

$$NA\mathrm{Cov}\,(\hat{\rho}_{gh}, \hat{\rho}_{ij}) = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \sum_{c=1}^{m_i} \sum_{d=1}^{m_j} u_{ab}^{(gh)} N\mathrm{Cov}\,(p_{ab}^{(gh)}, p_{cd}^{(ij)}) u_{cd}^{(ij)}, \tag{28}$$

where

$$N\mathrm{Cov}\,(p_{ab}^{(gh)}, p_{cd}^{(ij)}) = \pi_{abcd}^{(ghij)} - \pi_{ab}^{(gh)} \pi_{cd}^{(ij)}. \tag{29}$$

Here $\pi_{abcd}^{(ghij)}$ are the probabilities for the four-way contingency table for variables g, h, i, j, which can be estimated consistently by the corresponding sample proportions $p_{abcd}^{(ghij)}$. It follows from (27) that

$$\sum_{a=1}^{m_g} \sum_{b=1}^{m_h} u_{ab}^{(gh)} \pi_{ab}^{(gh)} = 0, \tag{30}$$

so that equation (28) simplifies to

$$NA\mathrm{Cov}\,(\hat{\rho}_{gh}, \hat{\rho}_{ij}) = \sum_{a=1}^{m_g} \sum_{b=1}^{m_h} \sum_{c=1}^{m_i} \sum_{d=1}^{m_j} u_{ab}^{(gh)} \pi_{abcd}^{(ghij)} u_{cd}^{(ij)}, \tag{31}$$

which may be estimated as

$$\text{Est}\,[NA\text{Cov}\,(\hat{\rho}_{gh}, \hat{\rho}_{ij}) = \sum_{a=1}^{m_g}\sum_{b=1}^{m_h}\sum_{c=1}^{m_i}\sum_{d=1}^{m_j} \hat{u}_{ab}^{(gh)} p_{abcd}^{(ghij)} \hat{u}_{cd}^{(ij)}, \tag{32}$$

where $\hat{u}_{ab}^{(gh)}$ is (27) evaluated at $\hat{\boldsymbol{\theta}}_{gh}$. Equations (25), (31), and (32) hold for every pair of variables gh and ij with $g \neq h$ and $i \neq j$. Evaluating (25) at sample estimates and adding (32) gives the required estimate of $NA\text{Cov}\,(\hat{\rho}_{gh}, \hat{\rho}_{ij})$.

Equation (32) involves the four-way contingency table for variables g, h, i, j. However, these four-way contingency tables need not be computed. Let $\kappa_{\alpha ghij} = 1/N$, if $z_{\alpha g} = a, z_{\alpha h} = b, z_{\alpha i} = c, z_{\alpha j} = d$ and $\kappa_{\alpha ghij} = 0$, otherwise, where $z_{\alpha i}$ is the αth observation on variable i. Then (32) is

$$\text{Est}\,[NA\text{Cov}\,(\hat{\rho}_{gh}, \hat{\rho}_{ij}) = \sum_{\alpha=1}^{N} \kappa_{\alpha ghij} \hat{u}_{ab}^{(gh)} \hat{u}_{cd}^{(ij)}.$$

Hence, this estimated asymptotic covariance is obtained by reading the raw data and for each case multiplying $\hat{u}_{ab}^{(gh)}$ and $\hat{u}_{cd}^{(ij)}$ and cumulating over all cases in the data.

Acknowledgments. The research reported in this paper has been supported by the Swedish Council for Research in the Humanities and Social Sciences (HSFR) under the program *Multivariate Statistical Analysis.*

REFERENCES

AISH, A.M. and JÖRESKOG, K.G. (1990). A panel model for political efficacy and responsiveness: An application of LISREL 7 with weighted least squares. *Quality and Quantity*, **24**, 405–426.

AUSTIN, J.T. and WOLFLE, L.M. (1991). Annotated bibliography of structural equation modelling: Technical work. *British Journal of Mathematical and Statistical Psychology* **44**, 93–152.

BENTLER, P.M. (1983). Some contributions to efficient statistics for structural models: Specification and estimation of moment structures. *Psychometrika* **48**, 493–517.

BENTLER, P.M. and DIJKSTRA, T. (1985). Efficient estimation via linearization in structural models. Pp 9-24 in P.R. Krishnaiah (Ed.): *Multivariate Analysis VI*, Amsterdam: North-Holland.

BENTLER, P.M. and WEEKS, D.G. (1980). Linear structural equations with latent variables. *Psychometrika* **45**, 289–308.

BOLLEN, K. A. (1989). *Structural Equations with Latent Variables.* New York: Wiley.

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. and CUDECK, R. (1993). Alternative ways of assessing model fit. Forthcoming in *Testing Structural Equation Models.* K. A. Bollen and J. S. Long (Editors). Sage Publications.

JÖRESKOG, K.G. (1973). A general method for estimating a linear structural equation system. In A.S. Goldberger and O.D. Duncan (Eds.): *Structural Equation Models in the Social Sciences.* New York: Seminar Press, 85–112.

JÖRESKOG, K.G. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In P.R. Krishnaiah (Ed.): *Applications of Statistics.* Amsterdam: North-Holland Publishing Co., 265–287.

JÖRESKOG, K.G. (1978). Structural analysis of covariance and correlation matrices, *Psychometrika* **43**, 443–477.

JÖRESKOG, K.G. (1981). Analysis of covariance structures, *Scandinavian Journal of Statistics* **8**, 65–92.

JÖRESKOG, K.G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares, *Quality and Quantity* **24**, 387–404.

JÖRESKOG, K.G. (1993). On the estimation of polychoric correlations and their asymptotic covariance matrix. Accepted for publication in *Psychometrika.*

JÖRESKOG, K.G. and SÖRBOM, D. (1988). *PRELIS - A Program for Multivariate Data Screening and Data Summarization. A preprocessor for LISREL.* Second Edition. Chicago, Illinois: Scientific Software, Inc.

JÖRESKOG, K.G. and SÖRBOM, D. (1989). *LISREL 7 - A Guide to the Program and Applications.* Second Edition. Chicago: SPSS Publications.

LEE, S-Y., POON, W-Y. and BENTLER, P.M. (1990). A three-stage estimation procedure for structural equation models with polytomous variables, *Psychometrika* **55**, 45–51.

MCARDLE, J.J. and MCDONALD, R.P. (1984). Some algebraic properties of the reticular action model for moment structures, *British Journal of Mathematical and Statistical Psychology* **37**, 234–251.

MCDONALD, R.P. (1978). A simple comprehensive model of the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology* **31**, 59–72.

MUTHÉN, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychome-*

trika **49**, 115–132.

OLSSON, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient, *Psychometrika* **44**, 443–460.

QUIROGA, A.M. (1992). *Studies of the Polychoric Correlation and Other Correlation Measures for Ordinal Variables.* PhD Dissertation, Uppsala University, Department of Statistics.

RAO, C.R. (1965). *Linear Statistical Inference and Its Applications.* New York: Wiley.

SHAPIRO, A. and BROWNE, M.W. (1987). Analysis of covariance structures under elliptical distributions, *Journal of the American Statistical Association* **82**, 1092–1097.

DEPARTMENT OF STATISTICS
UPPSALA UNIVERSITY
P.O. BOX 513
S-75120 UPPSALA
SWEDEN