PSYCHOMETRIKA—VOL. 60, NO. 4, 489–503
DECEMBER 1995

TECHNICAL ASPECTS OF MUTHÉN'S LISCOMP APPROACH TO ESTIMATION OF LATENT VARIABLE RELATIONS WITH A COMPREHENSIVE MEASUREMENT MODEL

BENGT O. MUTHÉN*

GRADUATE SCHOOL OF EDUCATION & INFORMATION STUDIES, UCLA

ALBERT SATORRA

DEPARTMENT OF ECONOMICS, UNIVERSITAT POMPEU FABRA

Muthén (1984) formulated a general model and estimation procedure for structural equation modeling with a mixture of dichotomous, ordered categorical, and continuous measures of latent variables. A general three-stage procedure was developed to obtain estimates, standard errors, and a chi-square measure of fit for a given structural model. While the last step uses generalized least-squares estimation to fit a structural model, the first two steps involve the computation of the statistics used in this model fitting. A key component in the procedure was the development of a GLS weight matrix corresponding to the asymptotic covariance matrix of the sample statistics computed in the first two stages. This paper extends the description of the asymptotics involved and shows how the Muthén formulas can be derived. The emphasis is placed on showing the asymptotic normality of the estimates obtained in the first and second stage and the validity of the weight matrix used in the GLS estimation of the third stage.

Key words: structural equation modeling, dichotomous measures, generalized least-squares estimation, asymptotic covariance matrix.

1. Introduction

Drawing on work in Muthén (1978, 1979, 1983) and Muthén and Christoffersson (1981), Muthén (1984) formulated a general model and estimation procedure for structural equation modeling with a mixture of dichotomous, ordered categorical, and continuous measures of latent variables. The model and its estimation also included multiple-group analysis with mean, intercept, and threshold structures. This approach was implemented in the computer program LISCOMP (Muthén, 1987). LISCOMP also included extensions for nonnormal continuous and censored variables subsequently published as Muthén (1989a, 1989b, 1990). For an overview with applications, see Muthén (1989c). An important contribution of Muthén (1984) was the development of a general three-stage procedure to obtain estimates, standard errors, and a chi-square measure of fit for a given structural model. While the last step uses generalized least-squares estimation to fit a structural model, the first two steps involve the computation of the statistics used in this model fitting. A key component in the procedure was the development of a weight matrix corresponding to the asymptotic covariance matrix of the statistics computed in the first two stages.

Since the publication of Muthén (1984), several authors have written papers using closely related modeling and estimation procedures: the conceptual and computational

The research of the first author was supported by grant AA 08651-01 from NIAAA for the project "Psychometric Advances for Alcohol and Depression Studies" and grant 40859 from the National Institute for Mental Health. The research of the second author was partially supported by the Spanish DGICYT grants PB91-0814 and PB93-0403.

Requests for reprints should be sent to Bengt O. Muthén, Graduate School of Education, University of California, Los Angeles, CA 90024-1521.

0033-3123/95/1200-92054$00.75/0

developments by Arminger and his students and colleagues, Küsters (1987, 1990), Arminger and Küsters (1988, 1989), Schepers (1991), Schepers, Arminger, and Küsters (1992), Sobel and Arminger (1992); the alternative two-stage and weight matrix estimators for special cases by Jöreskog (1985, 1991), Jöreskog and Sörbom (1988, 1989), Bermann (1993); and the variations on the general theme by Lee and colleagues, Lee (1985), Lee and Poon (1986, 1987), Lee, Poon, and Bentler (1989, 1990a, 1990b, 1992), Poon and Lee (1987, 1988, 1992). The claims made in Muthén (1984) about the asymptotic behavior of the estimator have been questioned by Lee (February 4, 1988, personal communication) and by Lee, Poon, and Bentler (September 20, 1991, personal communication), see also Lee, Poon, and Bentler (1992, p. 91, p. 102). The technical aspects of the estimation procedure were only briefly described in Muthén (1984). The aim of the present paper is to extend this description so that it is clear that the claims were correct and can be proven. The emphasis will be placed on showing the asymptotic normality of the statistics vector and the asymptotic validity of the Muthén (1984) form of the weight matrix. In order to give concreteness to the formulas, they will be explicated in terms of a specific model that is easy to understand, but still shows the general issues. For simplicity, the notation will be kept in line with that of Muthén (1984), referred to as M hereafter.

2. The General Model

For background material to M's approach, the interested reader is referred not only to the 1984 article but also to the overview given in Muthén (1983) and references therein. Briefly stated, the general model of M specifies the measurement structure for a p-dimensional latent response variable vector y^*,

$$y^* = \nu + \Lambda\eta + \varepsilon, \tag{1}$$

where ν is a p-dimensional parameter vector of intercepts, Λ is a $p \times m$ matrix of loading parameters, η is an m-dimensional vector of latent variables, and ε is a p-dimensional vector of measurement errors. A set of linear structural relations are specified for the m-dimensional latent variable vector η regressed on a q-dimensional observed variable vector x,

$$\eta = \alpha + B\eta + \Gamma x + \zeta, \tag{2}$$

where α is an m-dimensional vector of intercept parameters, B is an $m \times m$ matrix of regression slope parameters, Γ is an $m \times q$ matrix of regression slopes, and ζ is an m-dimensional vector of residuals. With conventional assumptions

$$E(y^*|x) = \pi_1 + \Pi_2 x \tag{3}$$

$$V(y^*|x) = \Omega, \tag{4}$$

where

$$\pi_1 = \nu + \Lambda(I - B)^{-1}\alpha \tag{5}$$

$$\Pi_2 = \Lambda(I - B)^{-1}\Gamma \tag{6}$$

$$\Omega = \Lambda(I - B)^{-1}\Psi(I - B')^{-1}\Lambda' + \Theta. \tag{7}$$

This presentation of the model used in M and in LISCOMP ignores the generality of using multiple groups, including mean, intercept and threshold structures, and using the scaling matrix Δ, representing the standard deviations of the y^*'s; for an exhaustive presentation, see M and Muthén (1987).

The y^* variables of (1) are assumed to be normally distributed conditional on the x's and can be measured as dichotomous, ordered categorical or continuous variables. In work published after M, censored variables were also added to this framework (see Muthén, 1987; 1989a, 1989b; and references therein).

3. An Example: A Structural Probit Model

Here, we will consider the case of binary measurements of y^* for a specific structural model. Consider the structural model for the single latent variable η regressed on the q-dimensional x variable

$$\eta = \gamma' x + \zeta, \tag{8}$$

with the measurement specification for the p latent response variables y^*,

$$y^* = \lambda \eta + \varepsilon, \tag{9}$$

where λ is a p-dimensional vector of loadings. Here, each y_i^* variable is measured by a binary indicator y_i, introducing the threshold parameter τ_i,

$$y_i = \begin{cases} 1, & \text{if } y_i^* \geq \tau_i \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

In this example, we may standardize to $\nu = 0$ and $\alpha = 0$ so that in (5) through (7), π_1 vanishes,

$$\Pi_2 = \lambda \gamma' \tag{11}$$

$$\Omega = \lambda \psi \lambda' + \Theta. \tag{12}$$

Here, ψ is the variance of the residual ζ and Θ is a covariance matrix for the measurement errors of ε, assumed to be diagonal. Since y^* is a latent response variable, the variances on the diagonal of Ω are standardized to unity and Θ does therefore not introduce any parameters into the structural model. The structural parameters are: τ (a p-dimensional vector of thresholds), λ (a p-dimensional vector), ψ, and γ (a q-dimensional vector).

Given the normality assumption, this formulation gives the structural probit regression model introduced by Muthén (1979), so that for respondent r's observation on the i-th y variable and on x, the univariate conditional probability of $y_{ri} = 1$ is

$$\text{Prob } (y_{ri} = 1 | x_r) = \int_{\tau_i}^{\infty} \phi(y_{ri}^*; [\Pi_2 x_r]_i, [\Omega]_{ii}) \, dy_{ri}^*, \tag{13}$$

where ϕ is a univariate normal density and, in line with probit regression, the residual variance $[\Omega]_{ii}$ is standardized to unity. This may be rewritten as

$$\text{Prob } (y_{ri} = 1 | x_r) = \Phi_1\{-\tau_i + [\Pi_2 x_r]_i\}, \tag{14}$$

where Φ_1 is the standardized univariate normal distribution function. Similarly, we may consider the bivariate conditional probability for variables y_i and y_j

$$\text{Prob } (y_{ri} = 1, y_{rj} = 1 | x_r) = \Phi_2\{(-\tau_i + [\Pi_2 x_r]_i), (-\tau_j + [\Pi_2 x_r]_j), [\Omega]_{ij}\}, \tag{15}$$

where Φ_2 is the standardized bivariate normal distribution function and $[\Omega]_{ij}$ is a residual correlation.

This example may be used to define the quantities involved in the estimation of the model. With τ denoting the p-dimensional vector of threshold parameters, define the p-dimensional population vector

$$\sigma_1 = \tau, \tag{16}$$

the $(pq \times 1)$-dimensional vector

$$\sigma_2 = \text{vec}\ \{\Pi_2\}, \tag{17}$$

and the $p(p - 1)/2$-dimensional vector

$$\sigma_3 = \text{vec}^*\ \{\Omega\}, \tag{18}$$

where vec* selects the lower-triangular elements of Ω. We define $\sigma = (\sigma_1', \sigma_1', \sigma_1')'$. Let s_1, s_2, s_3 be vectors of sample statistics corresponding to estimates of these σ's and let $s = (s_1', s_2', s_3')'$ be the estimate $\hat{\sigma}$ of σ.

4. First and Second Stage Estimation in M: Asymptotic Distribution of the Sample Statistics

Section 3.2 of M discussed the asymptotic covariance matrix of the estimate $\hat{\sigma}$ of σ obtained from the first and second stages of estimation; that is, the quantities used as the sample statistics in the third stage. This provided the weight matrix for M's generalized least-squares (GLS) estimation of structural parameters in the third and final estimation step. The discussion in M will be expanded in this section, clarifying the asymptotics underlying the results. Although the probit example will be continued throughout, giving concreteness to the developments, the results are generally applicable to the dichotomous, ordered categorical, censored, and unlimited continuous y variables considered in the LISCOMP framework.

For M's GLS estimator to be correct, the key points are that the asymptotic distribution of $\hat{\sigma}$ is multivariate normal with mean σ and that the weight matrix of M is a consistent estimate of the asymptotic covariance matrix of $\hat{\sigma}$. Küsters (1987, 1990) has given an elaborate proof of the consistency and asymptotic normality of $\hat{\sigma}$ and gives an asymptotic covariance matrix closely related to M's. Here, a somewhat different and simpler approach will be taken to show consistency and asymptotic normality and to arrive at the covariance matrix in M.

In M's first stage estimation, the elements of σ_1 (thresholds or intercepts) and σ_2 (regression coefficients) are estimated as s_1 and s_2 respectively via separate maximum-likelihood estimation of each of the p univariate ordered multinomial probit regressions of the y's on x. For each variable y_i, $i = 1, 2, \ldots, p$, the corresponding ML estimates $s_{1,i}$ and $s_{2,i}$ of $\sigma_{1,i}$ and $\sigma_{2,i}$, respectively, are obtained. The estimates $s_{1,i}$ and $s_{2,i}$ are assembled into vectors s_1 and s_2 respectively. The values y_{ir}'s are taken to be independent for a given sequence of the values x_r's, hence the conditional (log) likelihood function $\ell_i = \ell_i(\sigma_{1,i}, \sigma_{2,i}|x_1, \ldots, x_n)$ of the multinomial ordinal probit regression with response variable y_i decomposes as the sum $\ell_i = \sum_{r=1}^{n} \ell_i^r$ of the individual univariate conditional likelihood functions. See for example Maddala (1983) where the expressions of the above likelihoods, and first and second derivatives, are given.

For example, in binary probit regression we get

$$\ell_i^r \equiv y_{ir} \log \text{Prob}\,(y_{ir} = 1|x_r) + (1 - y_{ir}) \log \text{Prob}\,(y_{ir} = 0|x_r), \tag{19}$$

where the probabilities are obtained by integrating univariate normal density functions.

In M's second estage estimation, the elements of σ_3 (correlations or covariances) are estimated as s_3 via separate pseudo maximum likelihood estimation of bivariate ordered multinomial probit regressions. Specifically, for each combination of variables y_i, y_j, $i \neq j$, the conditional likelihood function $\ell_{ij} = \ell_{ij}(\sigma_{1,i}, \sigma_{2,i}, \sigma_{1,j}, \sigma_{2,j}, \sigma_{3,ij}|x_1, \ldots, x_n)$ is maximized with the values of $\sigma_{1,i}$, $\sigma_{2,i}$, $\sigma_{1,j}$ and $\sigma_{2,j}$ held fixed at the values $s_{1,i}$, $s_{2,i}$, $s_{1,j}$ and $s_{2,j}$, respectively, obtained in the first stage estimation. The sequence of values of (y_{ir}, y_{jr}) are taken to be independent for a given sequence of the values of x_r's, and thus $\ell_{ij} = \sum_{r=1}^n \ell_{ij}^r$, where ℓ_{ij}^r are the individual conditional likelihood functions.

In the example of the binary probit regression, we have that

$$\begin{aligned} \ell_{ij}^r \equiv{} & y_{ir}y_{jr} \log \text{Prob}\,((y_{ir} = 1, y_{jr} = 1)|x_r) + \;.\;. \\ & y_{ir}(1 - y_{jr}) \log \text{Prob}\,((y_{ir} = 1, y_{jr} = 0)|x_r) + \\ & (1 - y_{ir})y_{jr} \log \text{Prob}\,((y_{ir} = 0, y_{jr} = 1)|x_r) + \\ & (1 - y_{ir})(1 - y_{jr}) \log \text{Prob}\,((y_{ir} = 0, y_{jr} = 0)|x_r), \end{aligned} \tag{20}$$

where the probabilities are computed integrating the corresponding bivariate normal density functions as in (15). Maximization of $\ell_{ij} = \ell_{ij}(s_{1,i}, s_{2,i}, s_{1,j}, s_{2,j}, \sigma_{3,ij})$ with respect to $\sigma_{3,ij}$ yields the pseudo maximum likelihood estimator $s_{3,ij}$ of $\sigma_{3,ij}$. The $s_{3,ij}$'s will be assembled into the vector of estimates s_3.

Now, as in (14) of M, we define the vector of first derivatives corresponding to first and second stage estimation

$$g = \sum_{r=1}^{n} g^r, \tag{21}$$

where

$$\begin{aligned} g^r = \Bigg(& \frac{\partial \ell_1^r}{\partial \sigma_{1,1}}, \frac{\partial \ell_1^r}{\partial \sigma_{2,1}}, \frac{\partial \ell_2^r}{\partial \sigma_{1,2}}, \frac{\partial \ell_2^r}{\partial \sigma_{2,2}}, \ldots, \frac{\partial \ell_p^r}{\partial \sigma_{1,p}}, \frac{\partial \ell_p^r}{\partial \sigma_{2,p}}, \frac{\partial \ell_{21}^r}{\partial \sigma_{3,21}}, \\ & \frac{\partial \ell_{31}^r}{\partial \sigma_{3,31}}, \ldots, \frac{\partial \ell_{pp-1}^r}{\partial \sigma_{3,pp-1}} \Bigg). \end{aligned} \tag{22}$$

Note that $g = g(\sigma)$ where $\sigma \equiv (\sigma_1', \sigma_2', \sigma_3')'$. We partition σ as $\sigma = (\sigma_\star', \sigma_3')'$ with $\sigma_\star = (\sigma_1', \sigma_2')'$. Let $\hat\sigma \equiv (s_1', s_2', s_3')'$ and consider $\hat\sigma = (\hat\sigma_\star', \hat\sigma_3')'$, where $\hat\sigma_\star \equiv (s_1', s_2')'$ and $\hat\sigma_3 \equiv s_3$.

Under usual regularity assumptions it can be proved that the vector of first and second stage estimates $\hat\sigma$ is a consistent estimate of σ, that is, plim $\hat\sigma = \bar\sigma$, where $\bar\sigma$ denotes the "true value" of σ and "plim" stands for probability limit (convergence in probability). Furthermore, by the way the first and second stage estimates are obtained, it is verified that

$$g(\hat\sigma) = 0. \tag{23}$$

The consistency of $\hat\sigma$ is proven in the Appendix under a set of general conditions to be satisfied in the estimation setting specified in M and described above. The conditions listed may not be the set of most stringent assumptions under which the results of M hold, but they are a sufficient set of conditions that allows the description of the most fundamental points without cluttering the exposition with specific technical details and without incurring substantial loss of generality. The conditions are certainly verified in the case of iid sampling of the x_r's, but should also hold in a more general setting as for example in the case of non-stochastic x_r's including dummy or classification variables. The consistency proof has been confined to an appendix in order to not distract the reader from the major object of the paper which is to show asymptotic normality and the expression of the covariance matrix of estimates.

Let us now concentrate on the asymptotic normality and the expression of the covariance matrix of parameter estimates. The following regularity assumptions are made. The assumptions are discussed following the presentation of the desired result.

A1. The function $g = g(\sigma)$ is continuously differentiable in an open, convex neighborhood of the true parameter value $\bar{\sigma}$

A2.

$$\frac{n^{-1}\partial g(\sigma)}{\partial \sigma} = n^{-1}\sum_{r=1}^{n} \frac{g^r(\sigma)}{\partial \sigma} \xrightarrow{P} A(\sigma) \equiv \lim n^{-1} E\left(\frac{\partial g(\sigma)}{\partial \sigma}\right)$$

$$= \underline{\lim n^{-1}} \sum_{r=1}^{n} E\left(\frac{\partial g^r(\sigma)}{\partial \sigma}\right), \tag{24}$$

uniformly in σ in a neighborhood of the true parameter value $\bar{\sigma}$, with $A(\sigma)$ being continuous at $\bar{\sigma}$. In addition, the matrix $\underline{A} \equiv A(\bar{\sigma})$ is nonsingular.

A3.

$$n^{-1/2} g(\bar{\sigma}) = n^{-1/2} \sum_{r=1}^{n} g^r(\bar{\sigma}) \xrightarrow{d} N(0, V), \tag{25}$$

where

$$\underline{V} \equiv \lim n^{-1} \sum_{r=1}^{n} E g^r(\bar{\sigma})(g^r(\bar{\sigma}))', \tag{26}$$

a finite matrix.

A4. The terms $\partial g^r(\bar{\sigma})/\partial \sigma$, $r = 1, \ldots, n$, are stochastically independent.

Expanding $g = g(\cdot)$ around $\bar{\sigma}$ by using the Mean Value Theorem gives

$$0 = g(\hat{\sigma}) = g(\bar{\sigma}) + \left(\frac{\partial g(\sigma^\star)}{\partial \sigma}\right)(\hat{\sigma} - \bar{\sigma}), \tag{27}$$

where $\sigma^\star$ is some point between $\hat{\sigma}$ and $\bar{\sigma}$. After rescaling by dividing by $n^{-1/2}$, we obtain

$$n^{1/2}(\hat{\sigma} - \bar{\sigma}) = \left(\frac{-n^{-1}\partial g(\sigma^\star)}{\partial \sigma}\right)^{-1} n^{-1/2} g(\bar{\sigma}). \tag{28}$$

From Assumption A2 and the fact that plim $\sigma^{\star} = \bar{\sigma}$, it follows (apply, e.g., Theorem 4.1.5 of Amemiya, 1985)

$$\text{plim}\left(\frac{n^{-1}\partial g(\sigma^{\star})}{\partial\sigma}\right) = \text{plim}\left(\frac{n^{-1}\partial g(\bar{\sigma})}{\partial\sigma}\right) = A, \tag{29}$$

and then, using (25) and (28), we deduce the desired result of asymptotic normality of $\hat{\sigma}$,

$$n^{1/2}(\hat{\sigma} - \bar{\sigma}) \xrightarrow{d} N(0, A^{-1}VA'^{-1}).$$

Some comments on the assumptions used to attain this asymptotic normality result are in order. The Assumption A1 is in fact a condition for the individual conditional likelihoods involved in the first and second stage estimation, for example, the functions ℓ_i^r and ℓ_{ij}^r of (19) and (20) in the probit regression example. The assumptions A2 and A3 typically result from applying to the sums $\partial g(\bar{\sigma})/\partial\sigma = \sum_{r=1}^{n} \partial g^r(\bar{\sigma})/\partial\sigma$ and $g(\bar{\sigma}) = \sum_{r=1}^{n} g^r(\bar{\sigma})$ some version of the law of large numbers and the central limit theorem, respectively. What is fundamental in order to apply such limit theorems is that the above functions are evaluated at a population value $\bar{\sigma}$, since that guarantees the independence of the terms of the sums under usual regularity conditions. Such independence will be attained for example when the conditioning variables x_r's are fixed, or when they are random but the x_r's are independent. A somewhat restrictive case is the one in which the x_r's are independent and identically distributed (iid). In that case the sums in (24) and (25) are sums of iid terms, and then the standard versions of the law of large numbers and the central limit theorems can be applied. Assumptions A2 and A3, however, can be claimed under weaker conditions than the iid sampling scheme. When the iid assumption is relaxed, however, some general condition on the behavior of the x_r's will need to be imposed. This is true, for example, in ordinary regression when the regressors are assumed to be random, where the limit of the second order moments of the x's need to be finite. It is beyond the scope of the paper to explicate specific, less stringent, assumptions that guarantee the validity of M's result.

Lee et al. (personal communication) raised the issue that standard limit theorems could not be applied when deriving the asymptotic normality, because the terms of the sums over all cases involved in g and $\partial g/\partial\sigma$ are not independent when evaluated at the first-stage estimates. To clarify this point we note again that the functions g and $\partial g/\partial\sigma$ as they appear on the right hand side of (28) and (29) respectively, are evaluated at the population value $\bar{\sigma}$; hence, the corresponding sums will be a sum of independent terms under usual regularity assumptions, allowing the application of the limit theorems. Facing this issue of lack of independence when terms are evaluated at the first stage estimates, Amemiya (1978) was to our knowledge the first to develop a double Taylor series expansion to obtain a sum of independent terms, to which the standard version of the central limit theorem could be applied. In fact, considering the Taylor expansion in (27) restricted only to the components of g of the second stage estimation, we obtain a compact expression of the double Taylor expansion introduced by Amemiya (1978, p. 17, Formulas (15)–(16)) in the context of a two-stage estimator for multivariate logit models. In his developments, Amemiya was interested only in the distribution of the parameters at the second stage of estimation, in our case the parameter σ_3, while we consider the asymptotic normality of the whole parameter estimate $\hat{\sigma}$.

The fact that the parameter $\sigma_{\star} = (\sigma_1', \sigma_2')'$ is substituted by an estimator obtained in the first stage, affects not only the argument of proof as noted above, but also affects the result of the asymptotic variance matrix of the estimator $\hat{\sigma}_3$. The estimator $\hat{\sigma}_3$ will not be an ML estimator and may therefore be less efficient than the true ML estimator. This is reflected by the fact that the asymptotic variance matrix of the two step estimator $\hat{\sigma}_3$ is not V^{-1} but $A^{-1}VA'^{-1}$. Note also that there may also be a loss of efficiency due to the fact that limited information (marginal univariate and bivariate likelihoods) is used instead of full information. See also Parke (1986) and Gong and Samaniego (1981) where the asymptotic variance of an PML estimator like $\hat{\sigma}_3$ is derived and compared with the variance of a true ML estimate.

Now we need to establish the form of the matrix A. Partitioning the rows of A according to a partitioning of g into two subvectors, the first involving the derivatives of the univariate marginal conditional likelihoods, $\partial \ell_i/\partial \sigma_{\bigstar,i}$, and the second involving the derivatives of bivariate marginal conditional likelihoods, $\partial \ell_{ij}/\partial \sigma_{3,ij}$, and partitioning the columns of A according to the partitioning of $\sigma = (\sigma'_{\bigstar}, \sigma'_3)'$, we obtain

$$A = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix},$$

since the likelihood functions ℓ_i do not depend on σ_3 and thus the $\partial \ell_i/\partial \sigma_3$'s are zero.

Using the standard regularity maximum likelihood result of

$$\frac{E\partial^2 \log L}{\partial \sigma \partial \sigma'} = -E\left(\frac{\partial \log L}{\partial \sigma}\right)\left(\frac{\partial \log L}{\partial \sigma}\right)', \tag{30}$$

where L is a regular maximum likelihood function, we deduce that the structural nonzero elements of A will be expressed as

1. Nonzero elements of A_{11}

$$\lim n^{-1}E \frac{\partial^2 \ell_i}{\partial(\sigma_{\bigstar,i})_s \partial(\sigma_{\bigstar,i})_t} = \lim n^{-1} \sum_{r=1}^{n} - E\left(\frac{\partial \ell_i^r}{\partial(\sigma_{\bigstar,i})_s}\right)(\partial \ell_i^r/\partial(\sigma_{\bigstar,i})_t.$$

2. Nonzero elements of A_{21}

$$\lim n^{-1}E \frac{\partial^2 \ell_{ij}}{\partial \sigma_{3,ij} \partial(\sigma_{\bigstar,i})_t} = \lim n^{-1} \sum_{r=1}^{n} - E\left(\frac{\partial \ell_{ij}^r}{\partial \sigma_{3,ij}}\right)\left(\frac{\partial \ell_{ij}^r}{\partial(\sigma_{\bigstar,i})_t}\right).$$

3. Nonzero elements of A_{22}

$$\lim n^{-1}E \frac{\partial^2 \ell_{ij}}{\partial \sigma_{3,ij} \partial \sigma_{3,ij}} = \lim n^{-1} \sum_{r=1}^{n} - E\left(\frac{\partial \ell_{ij}^r}{\partial \sigma_{3,ij}}\right)\left(\frac{\partial \ell_{ij}^r}{\partial \sigma_{3,ij}}\right),$$

where the $(\ )_s$ or $(\ )_t$ denote the s-th or t-th component of the vector enclosed and where all the derivatives are evaluated at the population value $\bar{\sigma}$.

Sample means will be used to obtain the appropriate estimates of expectation involved in the right-hand side of the above equalities; for example, $n^{-1} \sum_{r=1}^{n} (\partial \ell_i^r/\partial(\sigma_{\bigstar,i})_s)(\partial \ell_i^r/\partial(\sigma_{\bigstar,i})_t)$ evaluated at the estimated value $\hat{\sigma}$ will be used as estimate of $\lim n^{-1} \sum_{r=1}^{n} E(\partial \ell_i^r/\partial(\sigma_{\bigstar,i})_s)(\partial \ell_i^r/\partial(\sigma_{\bigstar,i})_t)$, and similarly for the other terms of A. The matrix V of (26) will be estimated as the mean of cross-products $n^{-1} \sum_{r=1}^{n} g^r(\hat{\sigma})(g^r(\hat{\sigma}))'$. Then a consistent estimate of the variance matrix $A^{-1}VA'^{-1}$ will be formed by substituting the population matrices for consistent estimates. This approach to the large-sample approximation of the variance matrix of first and second stage estimates $\hat{\sigma}$ is the one given in (19) and (20) of M. Assembling the above results leads to the consistent estimator of the asymptotic covariance matrix given in (22) of M and shows that the conditions for correct GLS estimation by (23) in M are fulfilled.

Note that the estimates of variance derived above exploit the assumption of correct likelihood specification required for the validity of the information matrix equality of (30). Alternatively, estimates of variance that do not exploit this equality could be considered. For example, the elements of A_{11} could be estimated as $n^{-1} \sum_{r=1}^{n} \partial^2 \ell_i^r / \partial(\sigma_{\star,i})_s \partial(\sigma_{\star,i})_t$, which involves the second order derivatives of the ℓ_i^r's, evaluated at $\hat{\sigma}$. The terms in A_{21} and A_{22} could be estimated analogously involving the second order derivatives. Then, the corresponding consistent estimate of $A^{-1}VA'^{-1}$ could be constructed. Such alternative estimates of variance would provide standard errors asymptotically robust against deviations from the normality assumption, an assumption required for the validity of the information matrix equality of (30). It should be warned, however, that violation of the normality assumption could in fact introduce inconsistency of the parameter estimate $\hat{\sigma}$.

The above discussion concludes our treatment of the general case. The probit example of section 3 is, however, of interest to further clarify the variance matrix components discussed above. Consider

$$A = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}.$$

The corresponding second-order derivatives are given in (18) of M, except that M's matrix is not divided by n. Here, A_{11} has p blocks corresponding to each univariate log likelihood and the approximation is applied to each block separately (compare with (19) in M). The rows of A_{21} and A_{22} draw on $p(p-1)/2$ bivariate log likelihoods and the approximation is applied separately to each such likelihood (compare with (20) and (21) in M). Küsters (1987, 1990) used the approximation for A_{11} and A_{22} but not for A_{21}. We will therefore focus on A_{21} and use the structural probit example for clarification.

Assume for simplicity that the structural probit model has three response variables y_i ($i = 1, 2, 3$, i.e., $p = 3$) and a single x variable ($q = 1$). Regressing on the single x variable, the element $\sigma_{1,i}$ corresponds to the threshold for y_i, $\sigma_{2,i}$ corresponds to the slope from the y_i regression, and $\sigma_{3,ij}$ corresponds to the residual correlation from the y_i, y_j regression. The rows of A_{21} correspond to the three bivariate log-likelihood functions ℓ_{21}, ℓ_{31}, ℓ_{32} (see (20)), while the columns correspond to the threshold and slope for y_1, the threshold and slope for y_2, and the threshold and slope for y_3,

$$A_{21} = \lim n^{-1} E \begin{bmatrix} \frac{\partial^2 \ell_{21}}{\partial \sigma_{3,21} \partial \sigma_{1,1}} & \frac{\partial^2 \ell_{21}}{\partial \sigma_{3,21} \partial \sigma_{2,1}} & \frac{\partial^2 \ell_{21}}{\partial \sigma_{3,21} \partial \sigma_{1,2}} & \frac{\partial^2 \ell_{21}}{\partial \sigma_{3,21} \partial \sigma_{2,2}} & 0 & 0 \\ \frac{\partial^2 \ell_{31}}{\partial \sigma_{3,31} \partial \sigma_{1,1}} & \frac{\partial^2 \ell_{31}}{\partial \sigma_{3,31} \partial \sigma_{2,1}} & 0 & 0 & \frac{\partial^2 \ell_{31}}{\partial \sigma_{3,31} \partial \sigma_{1,3}} & \frac{\partial^2 \ell_{31}}{\partial \sigma_{3,31} \partial \sigma_{2,3}} \\ 0 & 0 & \frac{\partial^2 \ell_{32}}{\partial \sigma_{3,32} \partial \sigma_{1,2}} & \frac{\partial^2 \ell_{32}}{\partial \sigma_{3,32} \partial \sigma_{2,2}} & \frac{\partial^2 \ell_{32}}{\partial \sigma_{3,32} \partial \sigma_{1,3}} & \frac{\partial^2 \ell_{32}}{\partial \sigma_{3,32} \partial \sigma_{2,3}} \end{bmatrix}. \tag{31}$$

Consider as an example the bivariate log likelihood function ℓ_{21} in the first row of A_{21}. For this likelihood, the information matrix is

$$-E\begin{bmatrix} \frac{\partial^2 \ell_{21}}{\partial\sigma_{1,1}\partial\sigma_{1,1}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{1,1}\partial\sigma_{2,1}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{1,1}\partial\sigma_{1,2}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{1,1}\partial\sigma_{2,2}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{1,1}\partial\sigma_{3,21}} \\ \frac{\partial^2 \ell_{21}}{\partial\sigma_{2,1}\partial\sigma_{1,1}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{2,1}\partial\sigma_{2,1}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{2,1}\partial\sigma_{1,2}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{2,1}\partial\sigma_{2,2}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{2,1}\partial\sigma_{3,21}} \\ \frac{\partial^2 \ell_{21}}{\partial\sigma_{1,2}\partial\sigma_{1,1}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{1,2}\partial\sigma_{2,1}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{1,2}\partial\sigma_{1,2}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{1,2}\partial\sigma_{2,2}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{1,2}\partial\sigma_{3,21}} \\ \frac{\partial^2 \ell_{21}}{\partial\sigma_{2,2}\partial\sigma_{1,1}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{2,2}\partial\sigma_{2,1}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{2,2}\partial\sigma_{1,2}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{2,2}\partial\sigma_{2,2}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{2,2}\partial\sigma_{3,21}} \\ \frac{\partial^2 \ell_{21}}{\partial\sigma_{3,21}\partial\sigma_{1,1}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{3,21}\partial\sigma_{2,1}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{3,21}\partial\sigma_{1,2}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{3,21}\partial\sigma_{2,2}} & \frac{\partial^2 \ell_{21}}{\partial\sigma_{3,21}\partial\sigma_{3,21}} \end{bmatrix}, \tag{32}$$

which by regular maximum-likelihood theory can be written as

$$\sum_{r=1}^{n} E\begin{pmatrix} \partial\ell_{21}^r/\partial\sigma_{1,1} \\ \partial\ell_{21}^r/\partial\sigma_{2,1} \\ \partial\ell_{21}^r/\partial\sigma_{1,2} \\ \partial\ell_{21}^r/\partial\sigma_{2,2} \\ \partial\ell_{21}^r/\partial\sigma_{3,21} \end{pmatrix}\begin{pmatrix} \partial\ell_{21}^r/\partial\sigma_{1,1} \\ \partial\ell_{21}^r/\partial\sigma_{2,1} \\ \partial\ell_{21}^r/\partial\sigma_{1,2} \\ \partial\ell_{21}^r/\partial\sigma_{2,2} \\ \partial\ell_{21}^r/\partial\sigma_{3,21} \end{pmatrix}. \tag{33}$$

The non-zero elements of the first row of nA_{21} are found in the last row of (32), that is, in the last row of (33). As an example, the 1,1 element of nA_{21} is found in the 5,1 element in (32), that is, by (32) and (33),

$$E\left(\frac{\partial^2 \ell_{21}}{\partial\sigma_{3,21}\partial\sigma_{1,1}}\right) = -\sum_{r=1}^{n} E\left(\frac{\partial\ell_{21}^r}{\partial\sigma_{3,21}}\right)\left(\frac{\partial\ell_{21}^r}{\partial\sigma_{1,1}}\right). \tag{34}$$

As large-sample approximations of the right-hand side we use the sums of products of the first-order derivatives evaluated at $\hat{\sigma}$.

Having thus obtained consistent estimates $\hat{A}$ and $\hat{V}$ of the matrices A and V, the consistent estimate of the variance matrix of $\hat{\sigma}$ is

$$1/n\hat{A}^{-1}\hat{V}\hat{A}'^{-1} = (n\hat{A})^{-1}(n\hat{V})(n\hat{A}')^{-1}; \tag{35}$$

that is, the asymptotic variance matrix (22) of M.

5. Discussion

The strength of M's approach is its simplicity given its generality. Despite the multivariate nature of the problem, only univariate and bivariate likelihoods are considered. Only a single parameter is estimated in each bivariate likelihood. Furthermore, only first-order derivatives of the univariate and bivariate likelihoods are required. These simplifications are important since it is clear from the above special case that the formulas are complex and become even more complex when covering all the different cases corresponding to combinations of the y variable types: dichotomous, ordered categorical, continuous, and censored.

Muthén (1983, 1984) distinguishes between Case A and Case B models, that is models that exclude or include x variables in the structural relations of (2). The special structural probit example that was given above corresponds to Case B in that x variables are available and conditioned on in the analysis in line with the tradition of simultaneous equation systems (see, e.g., L.-F. Lee, 1982). The estimates from the first two stages are regression-based. In contrast, Case A analysis does not involve such x variables. The conditional normality assumption then turns into a stronger, marginal normality assumption for the y^* variables. The estimates from the first two stages are correlation- or variance-covariance-based. With ordered categorical y's, for example, Case A leads to the analysis of polychoric correlations.

M's estimation approach, the asymptotics of which has been discussed in the present paper, is general in that it covers both Case A and Case B models. Most articles on this topic have covered only Case A models and have not provided a general estimation approach that can be used for Case B models. In the articles on Case A models, the asymptotics have been derived in an alternative way. For example, with ordered categorical y's, asymptotic normality can be deduced since the asymptotic distributions derive from bivariate data corresponding to sample proportions. This was utilized in Christoffersson (1975), Muthén (1978), Christoffersson and Gunsjö (1983), in Lee, Poon, and Bentler (1990b), and in Jöreskog and Sörbom (1988, 1989). This approach, however, is not general enough for use with Case B models. In contrast to drawing on sample proportions (grouped data), the asymptotics of M's general estimation approach draws on likelihood theory and asymptotics using sums of individual observations.

A cautionary note is warranted regarding the rate of convergence to the asymptotic distributions discussed above. Monte Carlo experience has shown that the GLS estimator gives good inference results when the models are small and the sample sizes are large, but can give very poor results for large models. Similar findings have been made for the ADF estimator with non-normal continuous variables (Muthén & Kaplan, 1992) as well as for the Case A approach for binary variables drawing on sample proportions (Muthén, 1993). The asymptotic results discussed above may, however, be useful also for large models using more robust inference techniques. More research is needed in this area, for example along the lines of Muthén (1993) drawing on work of Satorra (see, e.g., Satorra, 1992).

6. Appendix: Proof of Consistency of First and Second Stage Estimates

In this Appendix we will prove consistency of M's estimates arising from first and second stage estimation.

Consider the first and second stage of M's estimates described in sections above. For $i = 1, \ldots, p$ and $i \neq j$, $i, j = 1, \ldots, p$, consider $\ell_i = \ell_i(\vartheta_i | x_1, \ldots, x_n)$ and $\ell_{ij} = \ell_{ij}(\vartheta_{ij} | x_1, \ldots, x_n)$, where $\vartheta_i \equiv \sigma_{\star,i} = (\sigma'_{1,i}, \sigma'_{2,i})'$ and $\vartheta_{ij} \equiv (\sigma'_{\star,i}, \sigma'_{\star,j}, \sigma_{3,ij})'$. The parameter vectors $\overline{\vartheta_i \text{ and}}$ ϑ_{ij} have parameter spaces Ξ_i and Ξ_{ij} respectively. The first and second stage M's estimates are $\hat{\vartheta}_i = \hat{\sigma}_{\star,i} \equiv (s'_{1,i}, s'_{2,i})'$ and $\hat{\sigma}_{3,ij} \equiv s_{3,ij}$, respectively. The subscripts "i" and "ij" will be suppressed when clear from the context.

The following set of assumptions will be used to derive the consistency results. The set of assumptions considered do not strive for generality, rather we want to list typical conditions that can be checked in the estimation context of M and that are sufficient to guarantee the consistency results. All that is said for subscripts i and ij, should be understood to be said for $i = 1, \ldots, p$ and $i, j = 1, \ldots, p$, $i \neq j$.

B1. The parameter spaces Ξ_i (idem Ξ_{ij}) is a compact subset of an Euclidean space, with the "true parameter" value $\bar{\vartheta}_i(\bar{\vartheta}_{ij})$ contained in the interior of $\Xi_i(\Xi_{ij})$.

B2. The function $\ell_i(\ell_{ij})$ is continuously differentiable with respect to $\vartheta_i \in \Xi_i(\vartheta_{ij} \in \Xi_{ij})$ for all y_{ir}'s, and is a measurable function of the y_{ir}'s for all $\vartheta_i \in \Xi_i(\vartheta_{ij} \in \Xi_{ij})$.

B3. The stochastic function $n^{-1}\ell_i(n^{-1}\ell_{ij})$ converge to a nonstochastic function $\bar{\ell}_i(\bar{\ell}_{ij})$ in probability uniformly in $\vartheta_i \in \Xi_i(\vartheta_{ij} \in \Xi_{ij})$, as $n \rightarrow \infty$.

B4. The functions $\bar{\ell}_i$ and $\bar{\ell}_{ij}$ attains a unique global maximum at $\bar{\vartheta}_i$ and $\bar{\vartheta}_{ij}$, respectively.

In the second stage of estimation the parameter vector ϑ_{ij} is partitioned as $\vartheta_{ij} = (\vartheta'_{1,ij}, \vartheta_{2,ij})'$, with $\vartheta_{1,ij} \equiv (\sigma'_{\star,i}, \sigma'_{\star,j})'$ and $\vartheta_{2,ij} \equiv \sigma_{3,ij}$, and the likelihood function will be maximized with respect to $\vartheta_{2,ij}$, with $\vartheta_{1,ij}$ fixed at the first stage estimate $\hat{\vartheta}_{1,ij} \equiv (\hat{\sigma}'_{\star,i}, \hat{\sigma}'_{\star,j})'$. The corresponding maximum will be $\hat{\vartheta}_{2,ij} \equiv s_{3,ij}$. The following additional conditions are assumed (to simplify notation, we omit the subscripts "ij" from ϑ and Ξ until needed):

B5. The two sets of parameters ϑ_1 and ϑ_2 vary independently, that is, they vary in a product space, say $\Xi = \Xi_1 \times \Xi_2$.

B6. The estimate $\hat{\vartheta}_1$ (of the first stage estimation) converge in probability to the population value $\bar{\vartheta}_1$

B7. The function $\bar{\bar{\ell}}_{ij}(\vartheta_2) = \bar{\ell}_{ij}(\bar{\vartheta}_1, \vartheta_2)$ has $\bar{\vartheta}_2$ as a unique maximizer in Ξ_2.

Note that assumptions B1 through B4 are regular conditions ensuring the consistency of the extremum estimators (see, e.g., Theorem 4.1.1 or Theorem 4.1.2 of Amemiya, 1985). We will also see that under assumptions B1 through B7, the pseudo ML estimator $\hat{\vartheta}_2$ converges also in probability to $\bar{\vartheta}_2$. The estimator $\hat{\vartheta}_2$ is a specific case of the "quasi-generalized M-estimator" considered in Gourieroux and Monfort (1989).

Before getting into the proof of the consistency of $\hat{\vartheta}_2$, we consider a technical Lemma.

Lemma. (See Gourieroux and Monfort, 1989, Lemme 24.31.)

Consider the partition $\vartheta = (\vartheta'_1, \vartheta'_2)'$ and assume (B1) through (B7) holds. Define $\hat{\ell}_{ij}(\vartheta_2) \equiv \ell_{ij}(\hat{\vartheta}_1, \vartheta_2)$ and $\bar{\bar{\ell}}_{ij}(\vartheta_2) \equiv \bar{\ell}_{ij}(\bar{\vartheta}_1, \vartheta_2)$. Then $n^{-1}\hat{\ell}_{ij}(\vartheta_2)$ converges to $\bar{\bar{\ell}}(\vartheta_2)$ in probability uniformly in $\vartheta_2 \in \Xi_2$.

Proof. (See Gourieroux and Monfort, 1989.) Since

$$|n^{-1}\ell_{ij}(\hat{\vartheta}_1, \vartheta_2) - \bar{\ell}_{ij}(\bar{\vartheta}_1, \vartheta_2)| \leq |n^{-1}\ell_{ij}(\hat{\vartheta}_1, \vartheta_2) - \bar{\ell}_{ij}(\hat{\vartheta}_1, \vartheta_2)| + |\bar{\ell}_{ij}(\hat{\vartheta}_1, \vartheta_2) - \bar{\ell}_{ij}(\bar{\vartheta}_1, \vartheta_2)|,$$

we have

$$\sup_{\vartheta_2} |n^{-1}\ell_{ij}(\hat{\vartheta}_1, \vartheta_2) - \bar{\ell}_{ij}(\bar{\vartheta}_1, \vartheta_2)| \leq \sup_{\vartheta_2} |n^{-1}\ell_{ij}(\hat{\vartheta}_1, \vartheta_2) - \bar{\ell}_{ij}(\hat{\vartheta}_1, \vartheta_2)| + \sup_{\vartheta_2} |\bar{\ell}_{ij}(\hat{\vartheta}_1, \vartheta_2) - \bar{\ell}_{ij}(\bar{\vartheta}_1, \vartheta_2)|.$$

The first term of the last maximizing expression tends to zero in probability (when $n \rightarrow \infty$) due to the Assumption B3 of uniform convergence, while the second term also tends to zero due to the uniform continuity of the nonstochastic limit function $\bar{\ell}_{ij}$ (Note that this uniform continuity is implied by the smoothness Assumption B2 and the assumption that Ξ is compact.) □

From the result of the above lemma, we can now prove the consistency of $\hat{\vartheta}_2$. In fact, the proof follows the same reasoning as in Amemiya (1985, Proof of Theorem 4.1.1) with $\hat{\ell}_{ij}$ taking the role of the function to be maximized.

Given $\varepsilon > 0$, consider the open ball N_ε of R^q centered at $\bar{\vartheta}_2$ of radius ε (here q denotes the dimension of ϑ_2). Define

$$\varepsilon^\star = \bar{\bar{\ell}}_{ij}(\bar{\vartheta}_2) - \max_{\vartheta \in \Xi_2 \cap \bar{N}_\varepsilon} \bar{\bar{\ell}}_{ij}(\vartheta_2),$$

where $\bar{N}_\varepsilon$ is the complementary in R^q of N_ε. Let A_n be the event

$$|n^{-1}\hat{\ell}_{ij}(\vartheta_2) - \bar{\bar{\ell}}_{ij}(\vartheta_2)| < \frac{\varepsilon^\star}{2}, \quad \forall\, \vartheta_2 \in \Xi_2. \tag{36}$$

Hence, in A_n, it holds

$$\bar{\bar{\ell}}_{ij}(\hat{\vartheta}_2) > n^{-1}\hat{\ell}_{ij}(\hat{\vartheta}_2) - \frac{\varepsilon^\star}{2} \geq n^{-1}\hat{\ell}_{ij}(\bar{\vartheta}_2) - \frac{\varepsilon^\star}{2},$$

and

$$n^{-1}\hat{\ell}_{ij}(\bar{\vartheta}_2) > \bar{\bar{\ell}}_{ij}(\bar{\vartheta}_2) - \frac{\varepsilon^\star}{2}.$$

Consequently,

$$\bar{\bar{\ell}}_{ij}(\hat{\vartheta}_2) > \bar{\bar{\ell}}_{ij}(\bar{\vartheta}_2) - \varepsilon^\star.$$

Thus,

$$P(A_n) \leq P(\hat{\vartheta}_2 \in N_\varepsilon).$$

Using the result of Lemma 1, $P(A_n) \to 1$ when $n \to \infty$, hence $P(\hat{\vartheta} \in N_\varepsilon) \to 1$ when $n \to \infty$. Since this holds for every $\varepsilon > 0$, we have proved that $\hat{\vartheta}_2$ tends in probability to $\bar{\vartheta}_2$.

The consistency of $\hat{\vartheta}_i$ would follow from similar arguments, but in a simpler context, where ℓ_i and $\bar{\ell}_i$ would replace $\hat{\ell}_{ij}$ and $\bar{\bar{\ell}}_{ij}$, respectively. We can also take this proof for granted, under the stated assumptions, since B1 through B4 imply the conditions for consistency of, for example, Theorem 4.1.1 or Theorem 4.1.2 of Amemiya (1985).

Note that the consistency of the $\hat{\vartheta}_i$ and of the $\hat{\vartheta}_{2,ij}$, for $i < j$, $i, j = 1, \ldots, p$, imply the consistency of the whole vector $\hat{\sigma}$ of first and second stage estimates of M.

References

Amemiya, T. (1978). On a two-step estimation of a multivariate logit model. *The Journal of Econometrics*, *8*, 13–21.

Amemiya, T. (1985). *Advanced econometrics*. Cambridge, MA: Harvard University Press.

Arminger, G., & Küsters, U. (1988). Latent trait models with indicators of mixed measurement level. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 51–73). New York: Plenum.

Arminger, G., & Küsters, U. (1989). Construction principles for latent trait models. In C. Clogg (Ed.), *Sociological Methodology* (pp. 369-393). Oxford: Basil Blackwell.

Bermann, G. (1993). *Estimation and inference in bivariate and multivariate ordinal probit models*. Stocholm, Sweden: Almqvist & Wiksell. (Doctoral dissertation, University of Uppsala, Sweden).

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, *40*, 5–32.

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

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

Gourieroux, C., & Montfort, A. (1989). *Statistique et modeles econometriques*, *Vol. I*, *II*. Paris: Economica.

Jöreskog, K. G. (1985, July). *Estimation of the polyserial correlation from summary statistics*. Paper presented at the Fourth European Meeting of the Psychometric Society in Cambridge, England.

Jöreskog, K. G. (1991, July). *Latent variable modeling with ordinal variables*. Paper presented at the International Workshop on Statistical Modeling and Latent Variables in Trento, Italy.

Jöreskog, K. G., & Sörbom, D. (1988). *PRELIS—A program for multivariate data screening and data summarization. A preprocessor for LISREL* (2nd ed.). Chicago, IL: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7—A guide to the program and applications* (2nd ed.). Chicago: SPSS Publications.

Küsters, U. (1987). Hierarchische Mittlewert- und Kovarianzstrukturmodelle mit Nichtmetrischen Endogenen Variablen [Hierarchical mean—and covariance structure models with nonmetric endogenous variables]. Heidelberg: Physica-Verlag.

Küsters, U. (1990). A note on sequential ML estimates and their asymptotic covariances. *Statistical Papers*, *31*, 131–145.

Lee, L.-F. (1982). Health and wage: A simultaneous equation model with multiple discrete indicators. *International Economic Review*, *23*, 199–221.

Lee, S.-Y. (1985). Maximum-likelihood estimation of polychoric correlations in $r \times s \times t$ contingency tables. *Journal of Statistical Computation and Simulations*, *23*, 53–67.

Lee, S.-Y., & Poon, W.-Y. (1986). Maximum likelihood estimation of polyserial correlations. *Psychometrika*, *51*, 113–121.

Lee, S.-Y., & Poon, W.-Y. (1987). Two-step estimation of multivariate polychoric correlations. *Communications in Statistics: Theory and Methods*, *16*, 307–320.

Lee, S.-Y., Poon, W.-Y., & Bentler, P. M. (1989). Simultaneous analysis of multivariate polytomous variates in several groups. *Psychometrika*, *54*, 63–73.

Lee, S.-Y., Poon, W.-Y., & Bentler, P. M. (1990a). Full maximum likelihood analysis of structural equation models with polytomous variables. *Statistics and Probability Letters*, *9*, 91–97.

Lee, S.-Y., Poon, W.-Y., & Bentler, P. M. (1990b). A 3-stage estimation procedure for structural equation models with polytomous variables. *Psychometrika*, *55*, 45–57.

Lee, S.-Y., Poon, W.-Y., & Bentler, P. M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika*, *57*, 89–105.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.

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

Muthén, B. (1979). A structural probit model with latent variables. *Journal of the American Statistical Association*, *74*, 807–811.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*, 48–65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115–132.

Muthén, B. (1987). *LISCOMP. Analysis of linear structural equations with a comprehensive measurement model. Theoretical integration and user's guide*. Mooresville, IN: Scientific Software.

Muthén, B. (1989a). Multiple-group structural modeling with non-normal continuous variables. *British Journal of Mathematical and Statistical Psychology*, *42*, 55–62.

Muthén, B. (1989b). Tobit factor analysis. *British Journal of Mathematical and Statistical Psychology*, *42*, 241–250.

Muthén, B. (1989c). Latent variable modeling in heterogenous populations. *Psychometrika*, *54*, 557–585. Presidential address to the Psychometric Society, July, 1989.

Muthén, B. (1990). Moments of the censored and truncated bivariate normal distribution. *British Journal of Mathematical and Statistical Psychology*, *43*, 131–143.

Muthén, B. (1993). Goodness of fit with categorical and other non-normal variables. In K. Bollen & S. Long (Eds.), Testing structural equation models (pp. 205–243). Newbury Park: Sage Publications.

Muthén, B., & Christofferson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika*, *46*, 407–419.

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.

Parke, W. (1986). Pseudo maximum likelihood estimation: The asymptotic distribution. *The Annals of Statistics*, *14*, 355–357.

Poon, W.-Y., & Lee, S.-Y. (1987). Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika*, *52*, 409–430.

Poon, W.-Y., & Lee, S.-Y. (1988). Analyses of structural equation models with censored or truncated data (Technical report). Hong Kong: The Chinese University of Hong Kong, Department of Statistics.

Poon, W.-Y., & Lee, S.-Y. (1992). Statistical analysis of continuous and polytomous variables in several populations. *British Journal of Mathematical and Statistical Psychology*, *45*, 139–149.

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. In P. V. Marsden (Ed.), *Sociological Methodology 1992* (pp. 249–278). New York: Plenum.

Schepers, A. (1991). Numerische Verfahren und Implementation der Schatzung von Mittlewert und Kovarianz-Strukturmodellen mit Nichmetrische Variablen. Ahaus: Verlag Frank Hartman.

Schepers, A., Arminger, G., & Küsters, U. (1992). The analysis of non-metric endogenous variables in latent variable models: The MECOSA approach. In P. Gruber (Ed.), Econometric decision models: New methods of modeling and applications. Lecture notes in mathematical systems (pp. 459–472). Heidelberg: Springer Verlag.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

Sobel, M. E., & Arminger, G. (1992). Modeling household fertility decision: A non-linear simultaneous probit model. *Journal of the American Statistical Association*, *87*, 38–47.

Manuscript received 8/26/92
Final version received 7/28/94