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

COVARIANCE AND CORRELATION STRUCTURE ANALYSES WITH CONTINUOUS AND POLYTOMOUS VARIABLES

By Sik-Yum Lee, Wai-Yin Poon and P.M. Bentler

Chinese University of Hong Kong and University of California, Los Angeles

The main purpose of this paper is to develop two-stage methods for covariance and correlation structure analyses with continuous and polytomous variables. A full maximum likelihood approach and a partition maximum likelihood approach are used to attain the first stage estimates of the thresholds and the underlying covariance or correlation matrix. Then based on the joint asymptotic distributions of the first stage estimators and appropriate weight matrices, the generalized least squares approach is employed to estimate the structures parameters in the covariance or the correlation structures. Asymptotic properties of the estimates are derived. A simulation study is reported to give some ideas about the accuracy and the asymptotic behaviors of the method.

1. Introduction. Covariance structure analysis (Bentler 11983; Jöreskog, 1978) is a major trend in Psychometrics in the past ten to twenty years. It has been shown by Jöreskog (1970) that this general multivariate method covers multivariate analysis of variance, regression,, principal components, factor analysis and simplex models. Traditionally, it has been carried out in practice under the assumption that the observed variables are continuous with a multivariate normal distribution. Recently, asymptotically distribution-free methods have been developed (Bentler, 1983; Browne, 1984), and robustness of the normal theory methods underviolation of the distributional assumption has been studied (Anderson, 1988; Browne, 1987). Although the works cited above mainly concentrated on continuous variables, some attention has been focused on polytomous variables because most variables in practice are only observable in dichotomous or polytomous form. Examples of these variables are attitude items, performance ratings, etc. Bock and Liberman (1970) considered the maximum likelihood method for a dichotomous factors analysis with only one factor, and which was extended

---

AMS 1980 Subject Classifications: 62H12.

Key words and phrases: Structural equation models, full maximum likelihood, partition maximum likelihood, generalized least squares.

to a multiple factors model by Christoffersson (1975). Recently, Lee, Poon, and Bentler (1990a, b) developed likelihood estimation methods to analyze general structural equation models for polytomous variables. The statistical properties, such as the asymptotic distribution of the estimates and the goodness-of-fit statistics, were also provided.

More general situations that involve both continuous and polytomous variables are more complicated. Olsson, Drasgow, and Dorans (1982) studied the estimation of a bivariate polyserial correlation by means of the maximum likelihood approach and a two-step approach. Poon and Lee (1987) extended the bivariate model to a multivariate polyserial and polychoric model, where a vector of continuous variables and a vector of polytomous variables was involved. Optimal maximum likelihood estimates of the thresholds, polychoric and polyserial correlations were obtained. The main purpose of this article is to discuss two 2-stage procedures to analyze covariance structures for models with continuous and polytomous variables. At the first stage of the procedures, the thresholds and the polychoric and polyserial covariances/correlations are estimated without imposing any structure on the covariance/correlation model. Based on the statistical properties of the first stage estimates, the structural parameters in the model will be estimated at the second stage via the generalized least squares approach. The asymptotic distributions of the estimators and a goodness-of-fit test statistic will be presented. A simulation study is reported to give some ideas about the accuracy and the asymptotic behavior of the developed methods.

2. The Basic Model and Its Likelihood Function.

Suppose $\boldsymbol{X}(r \times 1)$ and $\boldsymbol{Y}(s \times 1)$ are continuous random vectors whose joint distribution is $N[\mathbf{0}, \Sigma_0]$, where elements of $\Sigma_0 = \Sigma(\boldsymbol{\theta}_0)$ are differentiable functions of a q by 1 true population parameter vector $\boldsymbol{\theta}_0$. It is assumed that the model is identifiable, that is, $\Sigma(\boldsymbol{\theta}_1) = \Sigma(\boldsymbol{\theta}_2)$ implies $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. Suppose that exact measurements of $\boldsymbol{Y}$ are not available and that the information in $\boldsymbol{Y}$ is given by an observable polytomous random vector $\boldsymbol{Z} = (Z_1, \cdots, Z_s)'$, such that the ath element of $\boldsymbol{Z}$

$$Z_a = k(a) \qquad \text{if} \quad \alpha_{a,k(a)} \le Y_a < \alpha_{a,k(a)+1}, \tag{1}$$

for $a = 1, \cdots, s$, $k(a) = 1, \cdots, m(a)$. Here $m(a)$ is the number of categories with respect to the ath variable, and $\alpha_{a,k(a)}$ is the unknown threshold parameter expect for $\alpha_{a,1}$ and $a_{a,m(a)+1}$ which are equal to $-\infty$ and $+\infty$ respectively. Let $\boldsymbol{k}$ be the multiple index $(k(1), \cdots, k(s))$, and let $\boldsymbol{Z_k}$ denotes a random observation $\boldsymbol{Z}$ with $Z_a = k(a)$, $a = 1, \cdots, s, k(a) = 1, \cdots, m(a)$. Consider a random sample from $(\boldsymbol{x}'\ \boldsymbol{Z}')'$ of size T, in which observed vectors are of the form $(\boldsymbol{x}_{\boldsymbol{k},i(\boldsymbol{k})}, \boldsymbol{k}')'$, where $i(k)$ denotes the index of the particular polytomous observation with $\boldsymbol{Z} = \boldsymbol{k}$, and $\boldsymbol{x}_{\boldsymbol{k}.i(\boldsymbol{k})}$ is the $i(\boldsymbol{k})$th observed vector of $\boldsymbol{X}$ with $\boldsymbol{Z} = \boldsymbol{z_k}$, and the ath component of $\boldsymbol{Z}, Z_a$, takes a value from $1, 2, \cdots, m(a)$.

Thus, $i(k)$ takes a value from the sequence $1, 2, \cdots, n_{\boldsymbol{k}}$ where $n_{\boldsymbol{k}}$ is the total number of observations with $\boldsymbol{Z} = (k(1), \cdots, k(s))$, and the sum of $n_{\boldsymbol{k}}$ over all $k(a)$ from 1 to $m(a)$ is equal to T. In the following, if the context is clear, we will just write i instead of $i(k)$.

Let Σ_{xx}, Σ_{yy} and Σ_{yx} be matrix function of $\boldsymbol{\theta}$ corresponding to correlation matrices of $\boldsymbol{X}, \boldsymbol{Y}$ and $(\boldsymbol{Y}, \boldsymbol{X})$ respectively, and let $Pr(\boldsymbol{k})$ be the probability of the cell $\boldsymbol{k}$. From (1), it can be shown that

$$Pr(\boldsymbol{k}) = (-1)^s \sum_{j(1)=0}^{1} \cdots \sum_{j(s)=0}^{1} (-1)^{\sum_{u=1}^{s} j(u)} \Phi_s(\alpha_{1,v(1)}, \cdots, \alpha_{s,v(s)}; \Sigma_{yy}) \tag{2}$$

where $v(a) = k(a) + i(a)$, and $\Phi_s(\alpha_1, \cdots, \alpha_s, \Sigma_{yy})$ is equal to

$$\int_{-\infty}^{\alpha_1} \cdots \int_{-\infty}^{\alpha_s} (2\pi)^{-s/2} |\Sigma_{yy}|^{-1/2} \exp\big(- \boldsymbol{y}'\Sigma_{yy}^{-1}\boldsymbol{y}/2\big)\, dy_s \cdots dy_1. \tag{3}$$

The maximum likelihood (ML) estimate of $\boldsymbol{\theta}$ can be obtained by minimizing the following negative logarithm of the likelihood function of $\boldsymbol{\theta}$:

$$- \sum_{k(1)=1}^{m(1)} \cdots \sum_{k(s)=1}^{m(s)} \sum_{i=1}^{n_k} \big\{ \log\, [p_1(\boldsymbol{x}_{\boldsymbol{k},1})] + \log\, [p_2(\boldsymbol{k} \mid \boldsymbol{x}_{\boldsymbol{k},1})] \big\}, \tag{4}$$

in which $p_1(\boldsymbol{x}_{\boldsymbol{k},1})$ is the r-dimensional multivariate normal density function, and $p_2(\boldsymbol{k} \mid \boldsymbol{x}_{\boldsymbol{k},1})$ is the conditional density function of $\boldsymbol{Z}$ given $\boldsymbol{x}_{\boldsymbol{k},1}$. Let σ'_a be the ath row of Σ_{yx}, σ_a^2 be the variance of Y_a, and $\boldsymbol{\sigma}_x$ be the column vector obtained from the $r(r+1)/2$ nonduplicated triangular elements of Σ_{xx} sequentially row by row. It can be shown that

$$p_1(\boldsymbol{x}_{\boldsymbol{k},1}) = (2\pi)^{-r/2} |\Sigma_{xx}|^{-1/2} \exp\big\{ - \boldsymbol{x}'_{\boldsymbol{k},1}\Sigma_{xx}^{-1}\boldsymbol{x}_{\boldsymbol{k},1}/2\big\}, \tag{5}$$

and $p_2(\boldsymbol{k} \mid \boldsymbol{x}_{\boldsymbol{k},1})$ is equal to

$$(-1)^s \sum_{i(1)=0}^{1} \cdots \sum_{i(s)=0}^{1} (-1)^{\sum_{u=1}^{s} i(u)} \Phi_s(\alpha_1^*, \cdots, \alpha_s^*, \boldsymbol{R}^*), \tag{6}$$

where $\boldsymbol{R}^*$ is the correlation matrix of $\boldsymbol{Y} \mid \boldsymbol{X}$, and

$$\alpha_a^* = \big(\alpha_{a,v(a)} - \boldsymbol{\sigma}'_a \Sigma_{xx}^{-1} \boldsymbol{x}_{\boldsymbol{k},1}\big)\big(\sigma_a^2 - \boldsymbol{\sigma}'_a \Sigma_{xx}^{-1} \boldsymbol{\sigma}_a\big)^{-1/2}. \tag{7}$$

It is clear from (4), (5) and (6) that direct minimization of the likelihood function to obtain the ML estimation is very difficult and tedious. Two 2-stage procedures that can significantly simplify the computation are discussed in the following sections.

3. The FML-GLS 2-stage Procedure. In this 2-stage procedure, the thresholds, and the polyserial and polychoric covariance/corrections are estimated at the first stage by the full likelihood without any constraints on Σ and the structural parameter $\boldsymbol{\theta}$ is estimated by the generalized least squares approach at the second stage.

Let $\boldsymbol{\eta}_a' = \boldsymbol{\sigma}_a'/\sigma_a$, and consider the following one-one onto transformation:

$$c_{a,v(a)} = \alpha_{a,v(a)}\left(1 - \boldsymbol{\eta}_a'\Sigma_{xx}^{-1}\boldsymbol{\eta}_a\right)^{-1/2}; \tag{8}$$

$$\boldsymbol{b}_a = -\Sigma_{xx}^{-1}\boldsymbol{\eta}_a\left(1 - \boldsymbol{\eta}_a'\Sigma_{xx}^{-1}\boldsymbol{\eta}_a\right)^{-1/2}; \tag{9}$$

$$r_{aj} = \frac{\left(\rho_{aj} - \boldsymbol{\eta}_a{}' - 1\Sigma_{xx}^{-1}\boldsymbol{\eta}_j\right)}{\left[\left(1 - \boldsymbol{\eta}_a'\Sigma_{xx}^{-1}\boldsymbol{\eta}_a\right)\left(1 - \boldsymbol{\eta}_j'\Sigma_{xx}^{-1}\boldsymbol{\eta}_j\right)\right]^{1/2}}, \tag{10}$$

where ρ_{aj} is the correlation of Y_a and Y_j, for $a, j = 1, \cdots, s$, $a < j$ and $v(i) = 2, \cdots, m(i)$. Then clearly from (7) that

$$\Phi_s\left(\alpha_1^*, \cdots, \alpha_s^*, \boldsymbol{R}_{yy,x}\right) = \Phi_s\left(\frac{c_{1,v(1)}}{\sigma_1} + \boldsymbol{b}_1'\boldsymbol{x}, \cdots, \frac{c_{s,v(s)}}{\sigma_s} + \boldsymbol{b}_s'\boldsymbol{x}; \boldsymbol{R}\right), \tag{11}$$

where $\boldsymbol{R}$ is the correlation matrix with off-diagonal element r_{aj}. Since $c_{a,v(a)}/\sigma_a$ is invariant to a scalar multiplication by $c_{a,v(a)}$ and σ_a, and hence, the parameters and the model are not identified. To solve this problem, we fixed $c_{a,2}$, $a = 1, \cdots, s$, to some preassigned values. The choice of these preassigned values will not greatly affect the estimation. At this stage, the new parameter vector is equal to $(\boldsymbol{\sigma}_x', \boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\boldsymbol{b}', \boldsymbol{\sigma}_y', \boldsymbol{r}', \boldsymbol{c}')'$, with $\boldsymbol{\sigma}_x, \boldsymbol{b}, \boldsymbol{\sigma}_y, \boldsymbol{r}$ and $\boldsymbol{c}$ being the parameter vectors that define the unknown distinct parameters in Σ_{xx}, $\boldsymbol{b}_1$, the diagonal elements of Σ_{yy}, r_{aj}, and $c_{a,v(a)}$ for $a, j = 1, \cdots, s$; $j > a$ and $v(a) = 2, \cdots, m(i)$. From (6) and (11), the likelihood function (4) in terms of this new set of parameter is expressed as

$$L(\boldsymbol{\sigma}_x, \boldsymbol{\beta}) = L_1(\boldsymbol{\sigma}_x) + L_2(\boldsymbol{b}, \boldsymbol{\sigma}_y, \boldsymbol{r}, \boldsymbol{c}), \tag{12}$$

where

$$L_1(\boldsymbol{\sigma}_x) = 2^{-1}\left\{rT\log(2\pi) + T\log|\Sigma_{xx}| + \sum_{i=1}^{T}\boldsymbol{x}_i\sigma_{xx}^{-1}\boldsymbol{x}_i\right\}, \tag{13}$$

$$L_2(\boldsymbol{b}, \boldsymbol{\sigma}_y, \boldsymbol{rc}) = -\sum_{k(1)=1}^{m(1)}\cdots\sum_{k(s)=1}^{m(s)}\sum_{i(\boldsymbol{k})=1}^{n(\boldsymbol{k})}\left\{\log(-1)^s\sum_{j(1)=0}^{1}\cdots\sum_{j(s)=0}^{1}(-1)^{\sum_{u=1}^{s}j(u)}\right.$$
$$\left.\cdot\,\Phi\left(\frac{c_{1,v(1)}}{\sigma_1} + \boldsymbol{b}_1'\boldsymbol{x}_{i(\boldsymbol{k})}, \cdots, \frac{c_{s,v(s)}}{\sigma_s} + \boldsymbol{b}_s'\boldsymbol{x}_{i(\boldsymbol{k})}; \boldsymbol{R}\right)\right\} \tag{14}$$

are functions that depend on fewer number of separable parameters. The unrestricted ML estimate, $\widehat{\Sigma}_{xx}$, of Σ_{xx} can be easily obtained by minimizing

(13). The ML estimates $\hat{b}, \hat{\sigma}_y, \hat{r}$, and $\hat{c}$ are obtained by minimizing (14) via the Fletcher-Powell algorithm (see, e.g. Luenberger, 1984). Furthermore, the ML estimates $(\hat{\eta}, \hat{\sigma}_y, \hat{\rho}, \hat{c})$ of $\eta, \sigma_y, \rho, c)$ can be obtained by the inverse transformations of (8) to (10). Finally, by a one-to-one transformation that converts correlations to covariances and vice versa, the ML estimates $\widehat{\Sigma}_{yx}$ and $\widehat{\Sigma}_{yy}$ of Σ_{yx} and Σ_{yy} can be obtains from $(\hat{\eta}, \hat{\sigma}_y, \hat{\rho})$. Thus, the first stage estimate $\widehat{\Sigma}$ of Σ_0, composed of $\widehat{\Sigma}_{xx}$, $\widehat{\Sigma}_{yx}$ and $\widehat{\Sigma}_{yy}$, can be obtained.

Let $\boldsymbol{\sigma}$ be the vector that consists the non-duplicating elements in Σ_0 and $\hat{\boldsymbol{\sigma}}$ be the corresponding vector that consists the ML estimates in $\widehat{\Sigma}$. We will assume that every element $\sigma_{ij}(\boldsymbol{\theta})$ of Σ and its first three orders partial derivatives with respect to $\boldsymbol{\theta}$ are continuous in a neighborhood of $\boldsymbol{\theta}_0$. Then, $\hat{\boldsymbol{\sigma}}$ is consistent and its joint asymptotic distribution is multivariate normal with mean vector $\boldsymbol{\sigma}$ and some covariance matrix Ω. At the second stage, the parameter vector $\boldsymbol{\sigma}$ in the covariance matrix $\Sigma(\boldsymbol{\sigma})$ is estimated by minimizing the GLS function

$$Q(\boldsymbol{\sigma}) = \frac{1}{2}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}(\boldsymbol{\theta}))' \boldsymbol{W}^{-1} (\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}(\boldsymbol{\theta})), \tag{15}$$

where $\boldsymbol{\sigma}(\theta)$ is the vector defined by the lower symmetric part of $\Sigma(\theta)$ and $\boldsymbol{W}$ is an appropriate weight matrix that converges to Ω in probability. Expressions for $\boldsymbol{W}$ can be obtained from the large sample approximation of various information matrices. Let $\hat{\boldsymbol{\theta}}$ be the vector that minimizes $Q(\boldsymbol{\theta})$. It can be shown by similar arguments as in Ferguson (1958, Section 3.2) that: (i) the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ is normal with mean vector $\boldsymbol{\theta}$ and covariance matrix $\{(\partial\boldsymbol{\sigma}/\partial\boldsymbol{\theta})'\boldsymbol{W}^{-1}(\partial\boldsymbol{\sigma}/\partial\boldsymbol{\theta})\}^{-1}$, and (ii) the asymptotic distribution of $2Q(\hat{\boldsymbol{\theta}})$ is chi-squared with degrees of freedom $(s + r)(s + r + 1)/2 - q$. Basic statistical inference for structural equation models, such as goodness-of-fit test of the model, test of the null hypothesis concerning $\boldsymbol{\theta}$, etc., can be performed via (i) and (ii). Computationally, $\hat{\boldsymbol{\theta}}$ can be obtained efficiently by the Gauss-Newton algorithm, see, for example, Lee & Jennrich, (1979).

4. The PML-GLS 2-stage Procedure.

The FML-GLS 2-stage procedure requires computing multiple integrals of the form $\Phi_s(\alpha_1^*, \cdots, \alpha_s^*; \boldsymbol{R}^*)$. Clearly, if s is large, it takes a long computer time to achieve the solution. In this section, a computationally more efficient PML-GLS 2-stage procedure is proposed. To solve the identification problem, here, we assume $\mathrm{diag}(\Sigma_{yy}) = \boldsymbol{I}$; and for convenience we also assume $\mathrm{diag}(\Sigma_{xx}) = \boldsymbol{I}$.

The first stage of this procedure is devoted to obtain the partition ML estimates (see, e.g. Poon & Lee, 1987) of the threshold parameters and elements in Σ without any correlatin structure. Elements in Σ are respectively the polyserial correlations ρ_a between $\boldsymbol{X}$ and Y_a and the polychoric correlation ρ_{ab} between Y_a and Y_b. The basic idea is to partition the whole general model into smaller submodels. For each $a = 1, \cdots, s$, ρ_a is estimated based on the observed random sample from the polyserial submodel corresponding to $(\boldsymbol{X}', Z_a)$ with only one polytomous variable Z_a. The partition ML estimate of ρ_{ab} is obtained from the polychoric submodel corresponding to the two dimensional $m(a) \times m(b)$ contingency table, with frequencies given by observations in Z_a and Z_b. Hence, we only need to compute single and double integrals in this procedure.

First, consider the partition ML estimation of $\boldsymbol{\rho}_a$ based on the random sample corresponding to $(\boldsymbol{X}', Z_a)$. Let $\boldsymbol{\alpha}_a = (\alpha_{a,2}, \cdots, \alpha_{a,m(a)})'$, and $n_{k(a)}$, $n_{k(a),k(b)}, \cdots, n_{\boldsymbol{k}}$ be the numbers of observations respectively corresponding to $Z_a = k(a), Z_a = k(a), Z_b = k(b)), \cdots, \boldsymbol{Z} = \boldsymbol{k}$, etc; and $\boldsymbol{x}_{k(a),j}$, $j = 1, \cdots, n_{k(a)}$ be the observed $\boldsymbol{x}$ vectors corresponding to $Z_a = k(a)$. For this submodel, the negative log likelihood function can be obtained as a special case of (4) with $s = 1$ and is given by

$$\begin{aligned} L_a(\boldsymbol{\rho}_x, \boldsymbol{\alpha}_a, \boldsymbol{\rho}_a) =& \frac{rT}{2}\log(2\pi) + \frac{T}{2}\log|\Sigma_{xx}| \\ &+ \frac{1}{2}\sum_{k(a)=1}^{m(a)}\sum_{j=1}^{n_{k(a)}} \boldsymbol{x}'_{k(a),j}\Sigma_{xx}^{-1}\boldsymbol{x}_{k(a),j} \\ &- \sum_{k(a)=1}^{m(a)}\sum_{j=1}^{n_{k(a)}} \log\left\{\Phi\left(\frac{\alpha_{a,k(a)+1} - \boldsymbol{\rho}_a\Sigma_{xx}^{-1}\boldsymbol{x}_{k(a),j}}{\left(1-\boldsymbol{\rho}'_a\Sigma_{xx}^{-1}\boldsymbol{\rho}_a\right)^{1/2}}\right)\right. \\ &\left. - \Phi\left(\frac{\alpha_{a,k(a)} - \boldsymbol{\rho}_a\Sigma_{xx}^{-1}\boldsymbol{x}_{k(a),j}}{\left(1-\boldsymbol{\rho}'_a\Sigma_{xx}^{-1}\boldsymbol{\rho}_a\right)^{1/2}}\right)\right\}, \end{aligned} \tag{16}$$

where $\boldsymbol{\rho}_x$ is the column vector obtained from the $r(r-1)/2$ non-duplicated lower triangular elements of Σ_{xx} sequentially row by row, and Φ is the standard univariate normal distribution function. To simplify the minimization of this function, the following one to one transformations as given in Lee and Poon (1986) are used:

$$\tau_{a,k(a)} = \alpha_{a,k(a)}\left(1 - \boldsymbol{\rho}'_a\Sigma_{xx}^{-1}\boldsymbol{\rho}_a\right)^{-1/2}, \tag{17}$$

and

$$\boldsymbol{\zeta}_a = -\Sigma_{xx}^{-1}\boldsymbol{\rho}_a\left(1 - \boldsymbol{\rho}'_a\Sigma_{xx}^{-1}\boldsymbol{\rho}_a\right)^{-1/2}. \tag{18}$$

Let $\boldsymbol{\tau}_a = (\tau_{a,2}, \cdots, \tau_{a,m(a)})'$. It follows from (16), (17) and (18) that $L_a(\boldsymbol{\rho}_x, \boldsymbol{\alpha}_a, \boldsymbol{\rho}_a)$ can be expressed as

$$L_a(\boldsymbol{\rho}_x, \boldsymbol{\tau}_a, \boldsymbol{\zeta}_a) = F_x(\boldsymbol{\rho}_x) + F_a(\boldsymbol{\tau}_a, \boldsymbol{\zeta}_a),$$

where

$$F_a(\boldsymbol{\tau}_a, \boldsymbol{\zeta}_a) = -\sum_{k(a)=1}^{m(a)}\sum_{j=1}^{n_{k(a)}} \log\left\{\Phi\left(\tau_{a,k(a)+1} + \boldsymbol{\zeta}'_a\boldsymbol{x}_{k(a),j}\right)\right.$$

$$- \Phi\big(\tau_{a,k(a)} + \zeta_a' \boldsymbol{x}_{k(a),j}\big)\big\}$$

$$= - \sum_{k(1)=1}^{m(1)} \cdots \sum_{k(s)=1}^{m(s)} \sum_{i=1}^{n_{\boldsymbol{k}}} \log \big\{\Phi\big(\tau_{a,k(a)+1} + \zeta_a' \boldsymbol{x}_{ki}\big) - \Phi\big(\tau_{a,k(a)} + \zeta_a' \boldsymbol{x}_{\boldsymbol{k},i}\big)\big\}, \tag{19}$$

and

$$F_x(\boldsymbol{\rho}_x) = \frac{rT}{2}\log(2\pi) + \sum_{k(1)=1}^{m(1)} \cdots \sum_{k(s)=1}^{m(s)} \sum_{i=1}^{n_{\boldsymbol{k}}} \frac{1}{2}\big\{\log|\Sigma_{xx}| + \boldsymbol{x}_{\boldsymbol{k},i}' \Sigma_{xx}^{-1} \boldsymbol{x}_{\boldsymbol{k},i}\big\}. \tag{20}$$

Since the jacobian of the one to one transformations (17) and (18) is nonsingular, minimizing L_a with respect to $\{\boldsymbol{\rho}_x, \boldsymbol{\alpha}_a, \boldsymbol{\rho}_a\}$ is equivalent to minimizing F_a and F_x with respect to $(\tilde{\boldsymbol{\tau}}_a, \tilde{\boldsymbol{\zeta}}_a$ and $\tilde{\boldsymbol{\rho}}_x$. Clearly, $\tilde{\boldsymbol{\rho}}_x$ is asymptotically equivalent to the vector that contains the corresponding lower triangular off-diagonal elements of the sample correlation matrix obtained from the continuous observations. Computationally, the minimum of $F_a(\boldsymbol{\tau}_a, \boldsymbol{\zeta}_a)$ can be achieved by the efficient Newton-Raphson algorithm, see Lee and Poon (1986). This minimization problem is relatively simple because only one polytomous variable and the univariate normal distribution function is involved. For all $a = 1, \cdots, s$, the partition ML estimate $\{\tilde{\boldsymbol{\alpha}}_a, \tilde{\boldsymbol{\rho}}_a\}$ of $\{\boldsymbol{\alpha}_a, \boldsymbol{\rho}_a\}$ is obtained from $\{\tilde{\boldsymbol{\tau}}_a, \tilde{\boldsymbol{\zeta}}_a\}$ and $\tilde{\boldsymbol{\rho}}_x$ via thee inverse transformation of (17) and (18).

Now, for $a, b = 1, \cdots, s,\ a > b$, consider the estimation of the polychoric correlation, ρ_{ab}, of the bivariate submodel corresponding to Y_a and Y_b. The negative logarithm likelihood function is given by

$$F_{ab}(\boldsymbol{\alpha}_a, \boldsymbol{\alpha}_b, \rho_{ab}) = - \sum_{k(a)=1}^{m(a)} \sum_{k(b)=1}^{m(b)} n_{k(a),k(b)} \log\{Pr(Z_a = k(a), Z_b = k(b))\},$$

where $\{k(a), k(b)\}$ are respectively the ath and bth element of an observation $\boldsymbol{z}_{\boldsymbol{k}}$ in the s-dimensional sample space. Since each random observation $\{k(a), k(b)\}$ in the bivariate submodel is corresponding to an observation in the s-dimensional multivariate model, we have

$$F_{ab}(\boldsymbol{\alpha}_a, \boldsymbol{\alpha}_b, \rho_{ab}) = - \sum_{k(1)=1}^{m(1)} \sum_{k(b)=1}^{m(b)} n_{\boldsymbol{k}} \log\{Pr(Z_a = k(a), Z_b = k(b))\}. \tag{21}$$

The partition ML estimate, $(\tilde{\boldsymbol{\alpha}}_a,\ \tilde{\boldsymbol{\alpha}}_b,\ \tilde{\rho}_{ab})'$, is the vector that minimizes $F_{ab}(\boldsymbol{\alpha}_a,\ \boldsymbol{\alpha}_b, \rho_{ab})$. Notice that this minimization problem is again relatively simple because only bivariate distribution functions are involved.

It should be noted that this partition ML approach gives two types of threshold estimates, one from the estimation of the polyserial correlations and the other from the estimation of the polychoric correlations. As pointed out by Poon and Lee 91987), the differences among these estimates usually are very small. Moreover, the second stage estimation of the parameter $\boldsymbol{\theta}$ in the correlation matrix is basically not affected by this phenomenon since it does not involve the threshold estimates in the estimation. Of course, one can take thee average of these estimates to be the final unique estimate of the threshold, if desired.

Let $\boldsymbol{\sigma}^*$ be the parameter vector of elements in Σ that consists all the polychoric correlations ρ_{ab}, and polyserial correlations ρ_a, and $\tilde{\boldsymbol{\sigma}}^*$ be the first stage estimate of $\boldsymbol{\sigma}$ in this PML-GLS 2-stage approach. It can be shown that (see, Lee Poon and Bentler, 19991) $\tilde{\boldsymbol{\sigma}}^*$ is consistent and its joint asymptotic distribution is multivariate normal with mean vector $\boldsymbol{\theta}^*$ and some covariance matrix Ω^*.

The second stage of the approach involves the estimation of the structural parameter $\boldsymbol{\theta}$ in the correlation structure based on the following GLS function:

$$Q^*(\boldsymbol{\theta}) = \frac{1}{2}[\tilde{\boldsymbol{\sigma}}^* - \boldsymbol{\sigma}^*(\boldsymbol{\theta})]'\boldsymbol{V}^{-1}[\tilde{\boldsymbol{\sigma}}^* - \boldsymbol{\sigma}^*(\boldsymbol{\theta}^*)], \tag{22}$$

where $\boldsymbol{\sigma}^*(\boldsymbol{\theta})$ is the vector $\boldsymbol{\sigma}^*$ with elements now considered as functions of the structural parameter $\boldsymbol{\theta}$, and $\boldsymbol{V}$ is an appropriate weight matrix that converges to Ω^* in probability. Again expressions for $\boldsymbol{V}$ can be obtained from the large sample approximation of the appropriate information matrices and Hessian matrices (see, Lee, Poon and Bentler, 1991). Let $\tilde{\boldsymbol{\theta}}$ be the vector that minimizes $Q^*(\boldsymbol{\theta})$. Then, it can be shown that (i) the asymptotic distribution of $\tilde{\boldsymbol{\theta}}$ is norma; with mean vector $\boldsymbol{\theta}$ and covariance matrix $\{(\partial\boldsymbol{\sigma}^*/\partial\boldsymbol{\theta})\}'\boldsymbol{V}^{-1}(\partial\boldsymbol{\sigma}^*/\partial\boldsymbol{\theta})\}^{-1}$, and (ii) The asymptotic distribution of $2Q^*(\boldsymbol{\theta})$ is chi-squared with degrees of freedom $(s+r)(s+r-1)/2-q$.

5. Simulation Study. A simulation study was conducted to study the performance of the estimates. Three sample sizes $T = 100, 200$, and 500 were considered and the dimensions r and s were chosen to be 6 and 2, respectively. The structure of Σ was taken from a confirmatory factor analysis model:

$$\Sigma = \boldsymbol{BCB}' + \boldsymbol{E},$$

were $\boldsymbol{B}$ is the factor loading matrix, $\boldsymbol{C}$ and $\boldsymbol{E}$ are covariance matrices of the factors and error measurements, respectively. For each sample size, random observations $(\boldsymbol{x}_1', \boldsymbol{y}_1')'$ with distribution $N[\mathbf{0}, \Sigma]$ were simulated using the IMSL (1987) subroutine DRNMVN. The true values of $\boldsymbol{B}, \boldsymbol{C}$, and $\boldsymbol{E}$ are:

$$\boldsymbol{B}' = \begin{bmatrix} 0.8 & 0.8 & 0.8 & 0.8 & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 0.8 & 0.8 & 0.8 & 0.8 \end{bmatrix},$$

$$\boldsymbol{C} = \begin{bmatrix} 1.0^* & 0.6 \\ 0.6 & 1.0^* \end{bmatrix}, \quad \text{and} \quad \boldsymbol{E} = 0.36\,\boldsymbol{I}_8,$$

where I_8 is the 8×8 identity matrix. The continuous data $\boldsymbol{y}_i$ were transformed to polytomous data $\boldsymbol{z}_i$ via $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = (-1.0, 0.0)$. The off diagonal elements of $\boldsymbol{E}$ and parameters with an asterisk were treated as known parameters and were not estimated. Since a correlation structure is considered in this paper, there are totally nine independent parameters in $\boldsymbol{B}$ and $\boldsymbol{C}$ to be estimated. Estimates of the diagonal parameters of $\boldsymbol{E}$ can be obtained as functions of the estimates of these nine parameters. Based on these simulated data, the two-stage estimates of the structural parameters were obtained. The subroutine given by Schervich (1984) was used to compute the multiple integrals of the normal distribution functions. For each sample size, the process was continued until 50 replications of the FML-GLS 2-stage estimates and the PML-GLS 2-stage estimates were completed. The means and the root mean square errors (RMSE) between the estimates and the true population values are reported in Table 1. Based on the RMSE values, it seems that the PML-GLS 2-stage method can produce reasonably accurate estimates that essentially have the same degree of accuracy as the FML-GLS 2-Stage estimates. For each sample size, the 50 goodness-of-fit test statistic values were analyzed via the SPSS (1988) program to see whether they deviate from the theoretical expected chi-squared distribution. The p-values of the 2-sided Kolmogorov test based on the PLM-GLS 2-Stage test statistic values for $T = 100$, 200, and 500 are 0.12, 0.72 and 0.53, respectively. The corresponding p-values of the FML-GLS 2 Stage approach are 0.91, 0.47 and 0.28, respectively. These values indicate the expected result that with moderate sample sizes, the distribution of the goodness-of-fit test statistics is reasonably chi-squared. Results obtained from similar analysis on univariate normality of the parameter estimates are also satisfactory.

6. Discussion. In this paper, two 2-stage procedures, namely FML-GLS and PML-GLS, for analyzing correlation structure with continuous and polytomous variables are developed. The asymptotic distribution of the estimators and a goodness-of-fit statistics are also presented. Hence, statistical analysis on the covariance and correlation structures is possible based on our results. Computationally, it is shown that since it only involves the evaluation of single and double integrals, the PML-GLS procedure is a feasible method to use in practice.

Acknowledgement. This research was support in part by a Hong Kong UGPC research grant for the project "Analysis of Structural Equation Models with Non-sstandard Data" and a research grant DA01070 from U.S. Public Health Service. We are indebted to a referee for very valuable comments and suggestions for imporvement of this papers.

		PML – GLS Estimates			FML – GLS Estimates		
	True Parameters	T=100	T=200	T=500	T=100	T=200	T=500
Mean	$B_{11}=0.8$	0.87	0.84	0.81	0.77	0.79	0.79
	$B_{21}=0.8$	0.87	0.85	0.82	0.78	0.79	0.79
	$B_{31}=0.8$	0.86	0.83	0.81	0.78	0.79	0.80
	$B_{41}=0.8$	0.85	0.83	0.82	0.77	0.79	0.79
	$B_{52}=0.8$	0.84	0.83	0.81	0.80	0.79	0.80
	$B_{62}=0.8$	0.83	0.82	0.81	0.80	0.79	0.80
	$B_{72}=0.8$	0.81	0.81	0.80	0.76	0.78	0.80
	$B_{82}=0.8$	0.81	0.80	0.80	0.75	0.80	0.80
	$C_{21}=0.6$	0.69	0.67	0.63	0.65	0.62	0.61
Mean	$B_{11}=0.8$	0.09	0.06	0.03	0.07	0.04	0.03
	$B_{21}=0.8$	0.09	0.05	0.03	0.07	0.04	0.03
	$B_{31}=0.8$	0.08	0.05	0.02	0.08	0.04	0.03
	$B_{41}=0.8$	0.07	0.05	0.03	0.07	0.05	0.03
	$B_{52}=0.8$	0.07	0.04	0.02	0.05	0.04	0.03
	$B_{62}=0.8$	0.07	0.05	0.03	0.06	0.04	0.02
	$B_{72}=0.8$	0.06	0.04	0.03	0.09	0.06	0.03
	$B_{82}=0.8$	0.07	0.05	0.03	0.09	0.06	0.03
	$C_{21}=0.6$	0.13	0.09	0.04	0.09	0.06	0.05

REFERENCES

ANDERSON, T. W. (1988). Multivariate linear relations. In Tarms Pukkila and Simo Puntanen (Ed.) *Proceedings of the second international Tampere conference in statistics*, (pp. 9–30) University of Tampere, Finland.

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

BOCK, R. D. and LIBERMAN, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, **75**, 179–197.

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

BROWNE, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika* **74**, 375–384.

CHRISTOFFERSSON, A. (1975), Factor analysis of dichotomized variables. *Psy-*

chometrika **40**, 5–32.

FERGUSON, T. S. (1958). A method of generating best asymptotically normal estimates with applications to the estimation of bacterial densities. *Annals of Mathematical Statistics*, **29**, 1046–1062.

IMSL LIBRARY (1987). *FORTRAN subroutine for statistical analysis.* Houston, Texas.

JÖRESKOG, K. G. (1970). A general method for analysis of covariance structures, *Biometrika* **57**, 239–251.

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

LEE, S. Y. & JENNRICH, R.I. (1979). A study of algorithm for covariances structure analysis with specific comparisons using factor analysis. *Psychometrika* **43**, 99–113.

LEE, S. Y., & POON, W. Y. (1986). Maximum likelihood estimation of polyserial correlation. *Psychometrika* **51**, 113–121.

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

LEE, S. Y., POON, W. Y. & BENTLER, P. M.(1990b). 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.(1991). A two-stage estimation of structural equation models with continuous and polytomous variables. Technical Report 60, Department of Statistics, The Chinese University of Hong Kong.

LUENBERGER, D. G. (1984). Linear and Nonlinear Programming (2nd Ed). MA: Addison-Wesley.

OLSSON, U., DRASGOW, F. & DORANS, N. J. (1982). The polyserial correlation coefficient. *Psychometrika* **47**, 337–347.

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

SCHERVISH, M. J. (1984). Multivariate normal probabilities with error bound. *Applied Statistics* **33**, 81–94.

SPSS (1988). SPSS-X User's Guide (3ed), Chicago, IL: SPSS, Inc.

DEPARTMENT OF STATISTICS
THE CHINESE UNIVERSITY OF HONG KONG
HONG KONG

DEPARTMENT OF STATISTICS
THE CHINESE UNIVERSITY OF HONG KONG
HONG KONG

DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF CALIFORNIA
LOS ANGELES
USA