Imposed quasi-normality in covariance structure analysis

Koning, Ruud H.; Neudecker, H.; Wansbeek, T.

Published in:
EPRINTS-BOOK-TITLE

Publication date:
1993

STATISTICAL MODELLING AND LATENT VARIABLES
K. Haagen, D.J. Bartholomew and M. Deistler (Editors)
1993 Elsevier Science Publishers B.V.

IMPOSED QUASI-NORMALITY IN COVARIANCE STRUCTURE ANALYSIS

Ruud Koning[a] , Heinz Neudecker[b] and Tom Wansbeek[c]

[a]Department of Econometrics, P.O. Box 800, 9700 AV Groningen, The Netherlands

[b]Department of Econometrics, Roeterstraat 11, 1018 WB Amsterdam, The Netherlands

[c]Department of Econometrics, P.O. Box 800, 9700 AV Groningen, The Netherlands

Abstract

In the analysis of covariance structures, the distance between an observed covariance matrix S of order $k \times k$ and $\Sigma(\theta) \equiv \mathcal{E}(S)$ is minimized by searching over the θ-space. The criterion leading to a best asymptotically normal (BAN) estimator of θ is found by minimizing the difference between vec$S$ and vec$\Sigma(\theta)$ in a metric that is based on the variance of s. So optimality requires inversion of a $k^2 \times k^2$ matrix, which is a formidable task when the model contains many equations and the matrix to be inverted is structured insufficiently. The latter can happen when the underlying distribution is non-normal. We present two computationally attractive alternatives that result in estimators for θ which are approximately BAN. Also, we compare the performance of various weight matrices by means of a Monte-Carlo simulation.

1. Introduction

When dealing with latent variables in cross-sectional models, the *analysis of covariance structures* is the most frequently used approach. In covariance structure analysis, the distance between a $k \times k$ sample covariance matrix S and its theoretical counterpart $\Sigma(\theta) = \mathcal{E}(S)$, induced by the underlying subject matter theory, is minimized by searching over the θ-space. A general criterion function to be minimized is the chi-squared criterion

$$q(\theta) = (s - \sigma(\theta))' U (s - \sigma(\theta)), \tag{1}$$

with $s = \text{vec}S, \sigma(\theta) = \text{vec}\Sigma(\theta)$ and U a $k^2 \times k^2$ weighting matrix. We will suppress the argument of Σ for brevity. It has been shown (Ferguson (1958), Browne (1984)) that an optimal choice for U is V^+, which is the Moore-Penrose inverse of

$$V = \mathcal{E}((s - \sigma)(s - \sigma)'), \tag{2}$$

or any consistent estimator W, say, of this matrix. (The need for employing a generalized inverse comes from the redundancy of elements in s and σ due to the symmetry of S and Σ.) With this choice of metric, the resulting estimator $\hat{\theta}$ is a best asymptotically normal (BAN) estimator. Moreover, with n denoting the sample size, the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$ is independent of the data generating process.

This is all standard but it should be noted that estimation of θ based on equation (1) and equation (2) may pose a computational problem when k is (moderately) large since the (Moore-Penrose) inverse of a matrix of $\mathcal{O}(k^2)$ has to be computed. In applications of covariance structure analysis, values of k in the order of magnitude of 15 or 20 or even larger can occur. When V is sufficiently structured, which occurs in particular when the underlying distribution is normal, the computational problem can be circumvented, but in general there is no such way out. For that case we propose an approximation of V or W that has the simple and computationally convenient structure that holds under normality. We hence try to find a compromise between *statistical efficiency*, obtained by using V^+ or W^+ as a weighting matrix in equation (1), and *computational expediency* of the structure under normality. In the next section we describe our approach in more detail. In section 3 we discuss a simpler, related approximation that has less theoretical appeal but is interesting from a mathematical point of view. In section 4 we report some Monte Carlo experiments to compare these two approaches with some benchmark cases. Section 5 contains some algebraic details of our approach, and section 6 concludes.

2. An approximation to the optimal weight matrix

BAN estimation of covariance structures may pose two problems. First, the matrix of fourth-order moments V is unknown (since it may depend on θ and possibly other parameters) and, secondly, this matrix can be very large in applications. The first problem is solved easily since any consistent estimator can be used without affecting the BAN-property. One way to find a consistent estimator is to consider first an *unbiased* estimator and to omit from that terms of lower order of n, the number of observations. Since we will use the unbiased estimator for comparison purposes in our Monte Carlo results further on, we will now describe it.

Suppose we have n independent observations $y_1, \ldots, y_n$ of a random k-vector y, which has a distribution with mean μ, variance Σ, and fourth-order moment

$$\Xi = \mathcal{E}\left((y-\mu)(y-\mu)' \otimes (y-\mu)(y-\mu)'\right).$$

We assume that the eighth-order moments of the distribution of y are finite. As usual in covariance structure analysis, we are primarily interested in estimating θ and consider all other parameters of the distribution of y to be nuisance parameters. If we collect all observations in $Y' \equiv (y_1, \ldots, y_n)$, the usual (unbiased and consistent) estimator for Σ is

$$S = \frac{1}{n-1} Y'GY,$$

where $G \equiv I_n - \frac{1}{n}\iota_n\iota_n'$, ι_n being an n-vector of ones. The *centering matrix* G transforms into deviations of means.

A useful matrix is the *commutation matrix*, P_{kk} (see Magnus and Neudecker (1979, 1988), Wansbeek (1989)). The commutation matrix of order k is defined as

$$P_{kk} \equiv \sum_{i,j=1}^{k} E_{ij} \otimes E_{ji},$$

where E_{ij} is a $k \times k$-matrix of zeros, except for its (i,j)th element, which is 1. Often, we will use $N_k \equiv \frac{1}{2}(I_{k^2} + P_{kk})$. Furthermore, note that a simple estimator for Ξ is

$$X \equiv \frac{1}{n}\sum_{j=1}^{n} Y'GE_{jj}GY \otimes Y'GE_{jj}GY.$$

Theorem 1 *Let y be distributed with mean μ, variance Σ and fourth-order moment Ξ. The variance of $s - \sigma$ is*

$$V = \mathcal{E}\left((s-\sigma)(s-\sigma)'\right) = \frac{1}{n}\Xi + \frac{2}{n(n-1)}N_k(\Sigma \otimes \Sigma) - \frac{1}{n}\sigma\sigma'.$$

An unbiased estimator for V is

$$W = \alpha\left(\frac{n^2}{n-1}X - 2N_k(S \otimes S) - ss'\right) - \frac{1}{n-2}ss', \tag{3}$$

where

$$\alpha \equiv \frac{n-1}{n(n-2)(n-3)}.$$

Proof See Koning, Neudecker and Wansbeek (1992). □

A variety of consistent estimators can be derived from this theorem. When the underlying distribution is *normal*, the second problem mentioned in the beginning of this section, viz. the size of V for large k, is also solved easily, since in that case the expression for V simplifies considerably:

$$V = \frac{2}{n-1}N_k(\Sigma \otimes \Sigma) \tag{4}$$

When equation (4) holds, a consistent estimator for V can be obtained by substituting a consistent estimator $\hat{\Sigma}$ for Σ in equation (4), that is, take W to be

$$W = \frac{2}{n-1}N_k\left(\hat{\Sigma} \otimes \hat{\Sigma}\right)$$

and the criterion can be written as

$$q(\theta) = (s-\sigma)'\left[\frac{2}{n-1}N_k\left(\hat{\Sigma} \otimes \hat{\Sigma}\right)\right]^{+}(s-\sigma) = \frac{1}{2}(n-1)\mathrm{tr}\left(\left(\hat{\Sigma} - \Sigma\right)\hat{\Sigma}^{-1}\right)^2. \tag{5}$$

We used the property that

$$\left(N_k(A \otimes A)\right)^{+} = N_k\left(A^{-1} \otimes A^{-1}\right)$$

for any nonsingular matrix A. One choice for $\hat{\Sigma}$ is S, and this defines the *weighted least squares* estimator in covariance structure analysis. Maximum likelihood is obtained by taking for $\hat{\Sigma}$ its value for some choice of θ and iterating until convergence. The appealing consequence of assuming normality is that a consistent estimator for V is obtained easily and that only $k \times k$ matrices have to be handled in (5). However, it is possible to

estimate V consistently without making distributional assumptions (see Browne (1984), Koning, Neudecker and Wansbeek (1992)). But then it is generally impossible to rewrite criterion (1) to criterion (5). If one is reluctant to make the assumption of normality, the computational problem remains.

We propose an alternative to exact BAN estimation of covariance structures. Suppose the matrix W can be *approximated* by $2N_k(A \otimes A)$, where W is a consistent estimator for V and A is a symmetric, nonsingular $k \times k$ matrix as yet to be determined. In this case $q(\theta)$ can (suppressing constants) be rewritten to

$$q(\theta) = \operatorname{tr}\left(\left(\hat{\Sigma} - \Sigma\right) A^{-1}\right)^2. \tag{6}$$

This expression for $q(\theta)$ is also very simple from a computational point of view, since it avoids inverting matrices of $\mathcal{O}(k^2)$, which in principle requires $\mathcal{O}(k^6)$ computing time. The computational gain would be lost, however, if the calculation of A would be $\mathcal{O}(k^6)$ again. However, we propose a method where this turns out not to be the case. We calculate A in such a way that

$$h(a) = \operatorname{tr}\left(2N_k(A \otimes A) - W\right)^2 = (w - t(a))'(w - t(a)), \tag{7}$$

is minimized, where W is any consistent estimator for V (e.g., based on the observed fourth-order moments), $w \equiv \operatorname{vec}W, a \equiv \operatorname{vec}A$ and $t(a) \equiv \operatorname{vec}2N_k(A \otimes A)$.

To calculate a from equation (7), we essentially have to minimize a fourth-degree polynomial in A. To solve this minimization problem, we use the Gauss-Newton method. We start the iterations with $A = S$ and then 'update' A in a way that is explained in detail in section 5. For obvious reasons, we call this approach *imposed quasi-normality*.

3. A related approximation

From equation (7) it is apparent that the first-order condition for A is a set of third-order equations in the elements of A, involving only third- and first-order terms. In some situations, most notably the estimation of factor loadings (conditional on unique variances) in factor analysis, such equations have the computationally attractive format of an eigenvalue problem. In the present case, this is not the case, but it is of some interest to note that there is a modified version that does lead to an eigenvalue problem. This arises when we approximate W by $A \otimes A$. In this case, criterion (1) can also be written as in equation (6). The matrix A is a minimizer of:

$$q^*(a) = \operatorname{tr}(A \otimes A - W)^2. \tag{8}$$

To present the first-order condition we need a matrix $\tilde{W}$ defined as follows. Each $k \times k$ block of W is stacked as a vector, and these vectors are grouped in a matrix. A formal definition is that $\tilde{W}$ is the $k^2 \times k^2$-matrix whose vec satisfies

$$\operatorname{vec}\tilde{W} \equiv (I_k \otimes P_{kk} \otimes I_k) \operatorname{vec}W \equiv B_k \operatorname{vec}W,$$

with $B_k \equiv I_k \otimes P_{kk} \otimes I_k$. For a more detailed discussion of the tilde-operator, we refer to Koning, Neudecker and Wansbeek (1991). Important properties of B_k are

$$B_k \operatorname{vec}(A \otimes A) = \operatorname{vec}(aa')$$

and

$$B_k \mathrm{vec}\,((A \otimes A)P_{kk}) = \mathrm{vec}\,((A \otimes A)P_{kk})\,.$$

To illustrate the tilde operator, take $k = 2$ and let W be

$$W = \begin{pmatrix} 1 & 2 & 2 & 3 \\ 2 & 4 & 4 & 5 \\ 2 & 4 & 4 & 5 \\ 3 & 5 & 5 & 6 \end{pmatrix},$$

then

$$\tilde{W} = \begin{pmatrix} 1 & 2 & 2 & 4 \\ 2 & 3 & 4 & 5 \\ 2 & 4 & 3 & 5 \\ 4 & 5 & 5 & 6 \end{pmatrix}.$$

Note that W has the structure of a fourth-order matrix (but for simplicity we used the numbers $1, \ldots, 6$ to indicate different elements so that it is not positive semidefinite).

Using the tilde operator in equation (8) we obtain

$$\begin{aligned} q^*(a) &= \mathrm{tr}(A \otimes A - W)^2 \\ &= (\mathrm{vec}(A \otimes A - W))' B_k B_k (\mathrm{vec}(A \otimes A - W)) \\ &= (\mathrm{vec}(aa' - \tilde{W}))' (\mathrm{vec}(aa' - \tilde{W})) \\ &= \mathrm{tr}(aa' - \tilde{W})^2. \end{aligned}$$

This yields the first-order condition

$$\tilde{W}a = (a'a)a. \tag{9}$$

Equation (9) is an eigenvalue problem, and it is simple to show that the eigenvector corresponding to the largest eigenvalue has to be chosen. The scaling of the vector follows from the largest eigenvalue, which equals $a'a$. Due to the fact that only the largest eigenvalue and corresponding eigenvector of $\tilde{W}$ are needed, the so-called power method (see Dahlquist and Björck (1974)) can be used. This method involves matrix multiplications only. Therefore, the matrix A minimizing $q^*(a)$ can be calculated easily.

Of course, some equations are redundant in equation (9), due to the symmetry of A. This redundancy can be removed by using the *duplication* and *elimination* matrices, denoted by D_k and L_k, respectively (Magnus and Neudecker (1980, 1988)). If we let

$$v \equiv L_k a,$$

i.e., v is the vector of length $k(k+1)/2$ with nonduplicated elements of a, then $D_k v = a$ and on premultiplying both sides of equation (9) by L_k we obtain

$$L_k \tilde{W} D_k v = (v' D_k' D_k v) v. \tag{10}$$

Equation (10) is an eigenvalue problem in the nonduplicated elements of a only. In terms of our earlier example on $\tilde{W}$ there holds

$$L_k \tilde{W} D_k = \begin{pmatrix} 1 & 2 & 4 \\ 2 & (3+4)/2 & 5 \\ 4 & 5 & 6 \end{pmatrix}.$$

So premultiplication of $\tilde{W}$ by L_k leads to the deletion of redundant rows and postmultiplication by D_k leads to a corresponding reduction of columns. For computational purposes, one should use equation (10) instead of equation (9).

A disadvantage of this approach in this section to approximating W is that it lacks a theoretical foundation. The logical structure of the estimator for the matrix of fourth-order moments V is lost. For example, $P_{kk}V = V$, but $P_{kk}(A \otimes A)$ will not be equal to $A \otimes A$, in general. Approximation by $2N_k(A \otimes A)$ does not suffer from this problem. In the Monte Carlo experiments we have nevertheless included the approximation by $A \otimes A$.

4. Simulations

Up till now we have been concerned with the structure of metrics only. The aim, however, of covariance structure analysis is to estimate θ. Determining the effect of various weighting matrices on the efficiency of the estimator $\hat{\theta}$ is a matter that seems hard to analyze analytically. Therefore we performed some simulation experiments. The setup of these experiments is quite simple and they are primarily meant to obtain a first idea of possible outcomes. We took $k = 2$ and considered the case of a linear parametric structure.

The data were generated according to the covariance structure

$$\Sigma = \begin{pmatrix} \theta_1 & \theta_1 \\ \theta_1 & \theta_1 + \theta_2 \end{pmatrix}, \tag{11}$$

so

$$\sigma = H\theta,$$

where

$$H \equiv \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}.$$

Let S be the sample covariance matrix. The estimator for θ that minimizes

$$(s - \sigma)'\Delta(s - \sigma)$$

is

$$\hat{\theta} = (H'\Delta H)^{-1} H'\Delta s,$$

with Δ a weighting matrix. For all our choices of Δ, $\hat{\theta}$ is consistent.

We used five different weighting matrices:

$$\Delta_1 = W^+,$$
$$\Delta_2 = (I_{k^2} + P_{kk})(A^{-1} \otimes A^{-1}),$$
$$\Delta_3 = A^{-1} \otimes A^{-1},$$
$$\Delta_4 = (I_{k^2} + P_{kk})(S^{-1} \otimes S^{-1}),$$
$$\Delta_5 = I_{k^2}$$

Table 1
Simulation results, model I.

Weighting matrix	$\hat{\theta}_1$	$\hat{\theta}_2$	Variance metric	
W^+	3.91	8.83	148.96	15783.79
	(0.34)	(1.64)		
$(I_{k^2} + P_{kk})(A^{-1} \otimes A^{-1})$	3.91	8.83	181.68	15791.41
(av. # its. 5.55)	(0.34)	(1.65)		
$A^{-1} \otimes A^{-1}$	3.96	9.05	57.78	15722.83
	(0.56)	(6.87)		
$(I_{k^2} + P_{kk})(S^{-1} \otimes S^{-1})$	3.91	8.82	89.15	10120.36
	(0.33)	(1.61)		
I_{k^2}	4.02	9.06	-	-
	(0.51)	(2.33)		

The weighting matrix Δ_1 is based on the unbiased estimator W in (3); Δ_2 is based on the approximation of W by $(I_{k^2} + P_{kk})(A \otimes A)$, Δ_3 is based on the approximation of W by $A \otimes A$. The weighting matrix Δ_4 is an optimal weighting matrix if the data generating process is multivariate normal. Finally, the weighting matrix Δ_5 yields the OLS estimator for θ.

The simulation results are presented in tables 1–3. We generated the data according to three different processes:

$$y_1 = 2z_1$$
$$y_2 = y_1 + 3z_2,$$

(model I), where z_1 and z_2 are independent standard normal variates;

$$y_1 = 4\sqrt{3}u_1$$
$$y_2 = y_1 + 6\sqrt{3}u_2,$$

(model II), with u_1 and u_2 independent uniformly $(0, 1)$ distributed, and

$$y_1 = 2\sqrt{3}z_1u_1$$
$$y_2 = y_1 + 6\sqrt{3}u_2,$$

(model III). In all three models, (11) holds and the implied parameter values are $\theta_1 = 4$ and $\theta_2 = 9$. The models were estimated with 100 observations and 1000 replications.

In the tables 1–3, we give the *variances* over the simulations of the estimators in parentheses. Under the header 'Variance metric' we give the variance of the weighting matrix of the $(1, 1)$-element of Δ (first column) and the variance of the $(4, 4)$-element of Δ. Since the metric is constant over the replications for $\Delta_5 = I_{k^2}$, we do not give variances there. The algorithm derived in section 5 converged in all cases. The mean number of iterations until convergence is given below the weighting matrix $(I_{k^2} + P_{kk})(A^{-1} \otimes A^{-1})$.

Table 2
Simulation results, model II.

Weighting matrix	$\hat{\theta}_1$	$\hat{\theta}_2$	Variance metric	
W^+	3.98	8.89	2.46	1792.16
	(0.13)	(0.67)		
$(I_{k^2} + P_{kk})(A^{-1} \otimes A^{-1})$	3.97	8.86	5.29	1836.19
(av. # its. 6.26)	(0.17)	(0.85)		
$A^{-1} \otimes A^{-1}$	3.98	8.88	1.03	1810.81
	(0.20)	(1.14)		
$(I_{k^2} + P_{kk})(S^{-1} \otimes S^{-1})$	3.94	8.78	33.10	6035.18
	(0.14)	(0.69)		
I_{k^2}	4.03	8.99	-	-
	(0.30)	(1.24)		

Table 3
Simulation results, model III.

Weighting matrix	$\hat{\theta}_1$	$\hat{\theta}_2$	Variance metric	
W^+	3.88	8.91	1823.76	14776.65
	(0.73)	(0.73)		
$(I_{k^2} + P_{kk})(A^{-1} \otimes A^{-1})$	3.90	8.84	1936.49	15092.45
(av. # its. 6.04)	(0.73)	(0.76)		
$A^{-1} \otimes A^{-1}$	3.98	8.96	1362.16	14256.55
	(1.83)	(0.90)		
$(I_{k^2} + P_{kk})(S^{-1} \otimes S^{-1})$	3.96	8.81	209.91	8484.60
	(0.71)	(0.73)		
I_{k^2}	4.06	9.02	-	-
	(0.90)	(1.36)		

For model I, we see that the results for weighting matrix $\Delta_1 = W^+$ and $\Delta_4 = (I_{k^2} + P_{kk})(S^{-1} \otimes S^{-1})$ are approximately equal, as far as the estimation of θ is concerned. This was to be expected: for a normal data generating process, we have $\mathcal{E}(W) = \frac{1}{n-1}(I_{k^2} + P_{kk})(\Sigma \otimes \Sigma)$. For the other models, the estimators based on Δ_1 should be more efficient than estimators based on Δ_4 or any of the other metrics. Apart from two marginal exceptions, this appears to be the case.

For all models, the average OLS estimate is close to the true value of θ. Its variance, however, is higher than that of most other estimators. This is as it should be, although it is a matter of judgement whether the loss of precision when using OLS is so severe that it outweighs the computational simplicity of this method for large models. The estimator with the second lowest variance is the one using weight matrix Δ_4, which is optimal if the data generating process is normal. The estimator based on this metric outperforms the estimator that uses some information of the theoretically optimal weighting matrix, viz. Δ_2. One also sees that this metric is outperformed by Δ_4 in terms of variance.

5. Derivation of the Gauss-Newton algorithm

To calculate A (or equivalently, $a \equiv \mathrm{vec}A$) from equation (7), we use the Gauss-Newton method. Then v $(= L_k a)$ is calculated using the recurrence relation

$$v_{i+1} = v_i + (G_i' G_i)^{-1} G_i'(w - g_i), \tag{12}$$

where the subscript i denotes the value of v in the i-th iteration step (for simplicity of notation we omit this subscript from now on), and

$$g \equiv g(v) \equiv t(D_k v) = t(a),$$
$$G = \frac{\partial g}{\partial v'}$$

We first calculate G. Using differential notation:

$$\begin{aligned} \mathrm{d}t &= 2\mathrm{vec}N_k\,(A \otimes \mathrm{d}A + \mathrm{d}A \otimes A) = 4\mathrm{vec}N_k\,(\mathrm{d}A \otimes A)\,N_k \\ &= 4\,(N_k \otimes N_k)\,\mathrm{vec}\,(\mathrm{d}A \otimes A) = 4\,(N_k \otimes N_k)\,B\,(\mathrm{vec}\mathrm{d}A \otimes \mathrm{vec}A) \\ &= 4\,(N_k \otimes N_k)\,B\,(I_{k^2} \otimes a)\,\mathrm{d}a, \end{aligned} \tag{13}$$

from which we conclude

$$G = \frac{\partial g}{\partial v'} = 4\,(N_k \otimes N_k)\,B\,(I_{k^2} \otimes a)\,D_k,$$

since $a = D_k v$, hence $\partial a / \partial v' = D_k$. Now we derive $G'G$. From the third member of equation (13) we derive

$$\begin{aligned} (\mathrm{d}t)'\mathrm{d}t &= 16\,[\mathrm{vec}N_k(\mathrm{d}A \otimes A)N_k]'\,[\mathrm{vec}N_k(\mathrm{d}A \otimes A)N_k] \\ &= 16\mathrm{tr}(\mathrm{d}A \otimes A)N_k(\mathrm{d}A \otimes A)N_k \\ &= 4\mathrm{tr}\,(\mathrm{d}A \otimes A)^2 + 8\mathrm{tr}P(\mathrm{d}A \otimes A)^2 + 4\mathrm{tr}(\mathrm{d}A)A \otimes A(\mathrm{d}A) \\ &= 4\mathrm{tr}\,(\mathrm{d}A)^2 \cdot \mathrm{tr}A^2 + 8\mathrm{tr}A(\mathrm{d}A)^2 A + 4(\mathrm{tr}A\mathrm{d}A)^2 \\ &= 4(\mathrm{d}a)'a'aI_{k^2}\mathrm{d}a + 8(\mathrm{d}a)'(I_k \otimes A^2)\mathrm{d}a + 4(\mathrm{d}a)'aa'\mathrm{d}a \end{aligned}$$

$$
\begin{aligned}
&= 4(\mathrm{d}a)' \left(a'aI_{k^2} + I_k \otimes A^2 + A^2 \otimes I_k + aa'\right) \mathrm{d}a \\
&= 4(\mathrm{d}a)' \bar{A} \mathrm{d}a \\
&= 4(\mathrm{d}v)' D_k' \bar{A} D_k \mathrm{d}v,
\end{aligned}
$$

with $\bar{A}$ implicitly defined. So $G'G = 4D_k'\bar{A}D_k$. Since $D_kL_kN_k = N_k$, $L_kN_kD_k = I$, and $\bar{A}N_k = N_k\bar{A}$, its inverse is $L_kN_k\bar{A}^{-1}N_kL_k'$, and we are left with the question of finding an expression for $\bar{A}^{-1}$. Let the spectral decomposition of A be

$$A = T\Lambda T'$$

and let λ denote the k-vector of eigenvalues grouped on the diagonal of Λ. Then

$$
\begin{aligned}
\bar{A} &= (T \otimes T) \left[a'aI_{k^2} + I_k \otimes \Lambda^2 + \Lambda^2 \otimes I_k + (T \otimes T)'aa'(T \otimes T)\right] (T \otimes T)' \\
&\equiv (T \otimes T) \left[\Delta + (\mathrm{vec}\Lambda)(\mathrm{vec}\Lambda)'\right] (T \otimes T)',
\end{aligned}
$$

with the diagonal matrix Δ implicitly defined. Its typical element is $a'a + \lambda_i^2 + \lambda_j^2$. Note that $a'a = \lambda'\lambda$. Define

$$\phi \equiv \frac{1}{1 + \sum_{i=1}^{k} \frac{\lambda_i^2}{\lambda'\lambda + 2\lambda_i^2}}$$

and

$$\mu \equiv \Delta^{-1}\mathrm{vec}\Lambda,$$

then

$$\bar{A}^{-1} = (T \otimes T) \left[\Delta^{-1} - \phi\mu\mu'\right] (T \otimes T)'.$$

So far for $(G'G)^{-1}$. We now consider the other term in the Gauss-Newton algorithm:

$$
\begin{aligned}
&G'(w - g) \\
&\quad = 4D_k' \left(I_{k^2} \otimes a'\right) B \left(N_k \otimes N_k\right) \left(\mathrm{vec}W - 2\mathrm{vec}N_k \left(A \otimes A\right)\right) \\
&\quad = 4D_k' \left(I_{k^2} \otimes a'\right) B\mathrm{vec}N_k W N_k - 8D_k' \left(I_{k^2} \otimes a'\right) B\mathrm{vec}N_k^2 \left(A \otimes A\right) N_k \\
&\quad = 4D_k' \left(I_{k^2} \otimes a'\right) B\mathrm{vec}W - 4D_k' \left(I_{k^2} \otimes a'\right) B\mathrm{vec}(A \otimes A) \\
&\quad -4D_k' \left(I_{k^2} \otimes a'\right) B\mathrm{vec}\left(A \otimes A\right) P_{kk} \\
&\quad = 4D_k'\tilde{W}a - 4D_k' \left(I_{k^2} \otimes a'\right) (a \otimes a) - 4D_k' \left(I_{k^2} \otimes a'\right) \mathrm{vec}\left(A \otimes A\right) P_{kk} \\
&\quad = 4D_k' \left(\tilde{W}a - (a'a)a - \mathrm{vec}A^3\right) \\
&\quad \equiv 4D_k'\mathrm{vec}U,
\end{aligned}
$$

with U implicitly defined. Obviously,

$$
\begin{aligned}
\mathrm{vec}U &= (\tilde{W} - (a'a)I_{k^2} - (A \otimes A))\mathrm{vec}A \\
&\equiv C\mathrm{vec}A \\
&= \mathrm{vec}(\sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}C_{ij}),
\end{aligned}
$$

with C_{ij} the (i,j)th block of C.

The update term in equation (12) can now be given if we let M denote the diagonal $k \times k$-matrix containing the elements of μ,

$$\zeta \equiv \mathrm{tr} MT'UT = \mu' \mathrm{vec} T'UT,$$

and V is the $k \times k$-matrix containing the elements of Δ^{-1}, so

$$v_{ij} = \frac{1}{\lambda'\lambda + \lambda_i^2 + \lambda_j^2}.$$

The update term for a $(= D_k v)$ then is

$$\begin{aligned} & D_k (G'G)^{-1} G'(w - g(v)) \\ & \quad = D_k L_k N_k \bar{A}^{-1} N_k L_k' D_k' \mathrm{vec} U \\ & \quad = N_k (T \otimes T) \left[\Delta^{-1} - \phi\mu\mu'\right] (T \otimes T)' \mathrm{vec} U \\ & \quad = N_k (T \otimes T) \left[\Delta^{-1} - \phi\mu\mu'\right] \mathrm{vec} T'UT \\ & \quad = \mathrm{vec}(T\left[V * T'UT - \zeta\phi M\right] T'), \end{aligned}$$

where $*$ denotes the Hadamard (element-by-element) product. The update term has to be added to the current value of A $(= T\Lambda T')$, so given this current value of A (and hence of T, Λ, V, U etc.) the new value of A is given by

$$A = T\left(\Lambda + V * T'UT - \zeta\phi M\right) T'.$$

This has to be repeated until convergence.

6. Conclusion

In this paper we proposed a method for approximating the weight matrix in covariance structure analysis that has the computational simplicity that is induced when the underlying distribution is normal. It requires an iterative solution of a set of third-order equations. The derivations rely on a number of recently developed, useful matrix operators. The iterative algorithm converged quickly in the (limited) numerical experiments that we performed. These experiments also suggested no sizeable loss of efficiency relative to the optimal weighting matrix.

A number of questions remain for future research. The performance of our approach is as yet untested in larger simulation or real-life cases. When k is large the computational burden of having to compute the spectral decomposition of a $k \times k$-matrix may become troublesome, and an OLS approach may then become advisable, notwithstanding its inefficiency, which (as expected) was also brought out by the experiments. Another finding that requires further analysis is the good performance of the fourth metric in our experiments, which is the weighted least squares one, which is to simply take S for A. The evidence in the experiments raises some puzzles for which we do not yet have an answer. In the second set of experiments, we found the variances of the metric based on this approach to be huge, but in the third set these variances were by contrast surprisingly small. In both (non-normal) cases, however, the variance of $\hat{\theta}$ was for all practical purposes the

same as the one based on the unbiased estimator W, the first metric in the experiments. In terms of estimating V this estimator is superior from a theoretical point of view but this quality does apparently not carry over to estimating θ, which is a nonlinear function of W. This suggests that some of the good properties of W get lost along the line.

References

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

Dahlquist, G. and A. Björck (1974), *Numerical Methods*, Prentice Hall, Englewood Cliffs.

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

Koning, R.H., H. Neudecker and T. Wansbeek (1991), "Block Kronecker products and the vecb operator", *Linear Algebra and its Applications*, **149**, 165–184.

Koning, R.H., H. Neudecker and T. Wansbeek (1992), "Unbiased estimation of fourth-order matrix moments", *Linear Algebra and its Applications*, **160**, 163–174.

Magnus, J.R. and H. Neudecker (1979), "The commutation matrix: some properties and applications", *The Annals of Statistics*, **7**, 381–394.

Magnus, J.R. and H. Neudecker (1980), "The elimination matrix: some lemmas and applications", *SIAM Journal on Algebraic and Discrete Methods*, **1**, 422–449.

Magnus, J.R. and H. Neudecker (1988), *Matrix Differential Calculus*, John Wiley & Sons, New York.

Wansbeek, T. (1989), "Permutation matrix-II", in N. L. Johnson and S. Kotz, editors, *Encyclopedia of Statistical Sciences Supplementary Volume*, New York, John Wiley & Sons.

STATISTICAL MODELLING AND LATENT VARIABLES
K. Haagen, D.J. Bartholomew and M. Deistler (Editors)

Structural Equation Models with Hierarchical Data

Sik-Yum Lee & Wai-Yin Poon

Departmen of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong

Abstract

This paper considers the two-level analysis of structural equation models with unbalanced sampling designs. First, statistical estimation based on the direct maximum likelihood and the generalized least squares approaches is developed. In addition, convenient two-stage procedures are developed to analyze models with small sample of level-two or level-one units. In these procedures, the between-group and the within-group covariance matrices are first estimated without imposing any structures, then the structural parameters are estimated based on the asymptotic properties of the first stage estimates. Computationally, it is shown that the standard programs, LISREL VII or EQS 3, can be used to obtain the solution. Generalizations to arbitrary and elliptical distributions are also established. Numerical examples are presented to illustrate the results developed.

1. INTRODUCTION

Structural equation modelling (Joreskog, 1978; Bentler, 1980) is a widely used multivariate method in assessing behavioral, medical and social data. Traditionally, this method assumed independence among observations. However, in many applications, investigators frequently encounter data that have a multilevel nature, for example, those obtained by samples of randomly drawn students from randomly drawn schools. Under this kind of situation, the assumption of independence is not realistic, because individuals within a group are expected to share certain influencing factors and hence produce correlated observations. In recent years, a lot of attention has been devoted to developing models and methods for analyzing various kinds of data with

multilevel structures; see, for instance, Aitkin & Longford (1986), Goldstein (1986), Goldstein & McDonald (1988), Rubin (1981), and papers in Bock (1989). Clearly, the application of structural equation models to multilevel data will need to take into account the correlated structure of the data. McDonald & Goldstein (1989) studied a two level structural equation model and derived the likelihood function, and a goodness-of-fit test for the case of a balanced sampling design. In this paper, we will discuss some recent methods (Lee, 1990; Lee & Poon, 1992, 1991) for analysis of two level structural equation models with unbalanced sampling designs.

The following notation will be used in this paper. Suppose $\underline{A}$ is any p by p symmetric matrix, vec($\underline{A}$) represents the p^2 by 1 column vector formed by stacking the columns of $\underline{A}$, and vecs($\underline{A}$) represents the column vector obtained from the $p^* = p(p + 1)/2$ non-duplicated elements of $\underline{A}$. Let $\underline{K}_p$ be the p^2 by p^* transition matrix such that $\text{vecs}(\underline{A}) = \underline{K}_p'\text{vec}(\underline{A})$ and $\text{vec}(\underline{A}) = \underline{K}_p^{-}{}'\text{vecs}(\underline{A})$, where $\underline{K}_p^{-} = (\underline{K}_p'\underline{K}_p)^{-1}\underline{K}_p'$; and $\underline{M}_p = \underline{K}_p\underline{K}_p^{-}$. For properties of $\underline{K}_p$ and $\underline{M}_p$, see Browne (1974); and Nel (1980). The right Kronecker product of matrices is denoted by $\otimes$.

2. The Two-level Structural Equation Model

Suppose $\underline{x}_{gi}$ is a p by 1 observed random vector, g = 1, ..., G and i = 1, ..., N_g, such that

$$\underline{x}_{gi} = \underline{v}_g^* + \underline{v}_{gi} \ , \tag{1}$$

where $\underline{v}_g^*$ is a random vector varying at the group level, while $\underline{v}_{gi}$ is another random vector varying at the individual level. We assume that $\underline{v}_g^*$ and $\underline{v}_{gi}$ are independent, $\underline{v}_g^*$ and $\underline{v}_h^*$ are mutually independent for $h \neq g$, and $\underline{v}_{gi}$, $\underline{v}_{gj}$, $\underline{v}_{hi}$ and $\underline{v}_{hj}$ are also mutually independent. Note that $\underline{x}_{gi}$ and $\underline{x}_{gj}$ are dependent due to the existence of $\underline{v}_g^*$. It is assumed that $\underline{v}_g^*$ is identically distributed as $N[\underline{0}, \Sigma_{Bo}]$ and $\underline{v}_{gi}$ is identically distributed as $N[\underline{0}, \Sigma_{gwo}]$, where the between group covariance matrix Σ_{Bo}

$= \Sigma_B(\underline{\theta}_o)$ and the within group covariance matrix $\Sigma_{gwo} = \Sigma_{gw}(\underline{\theta}_o)$ can have any general identified structures defined with an unknown q by 1 population parameter vector $\underline{\theta}_o$. The covariance structure of $\underline{x}_{gi}$ is given by

$$\Sigma_{gwB}(\underline{\theta}_o) = \Sigma_{Bo} + \Sigma_{gwo} \ . \qquad (2)$$

Let $\underline{\theta}$ be the vector of unknown parameters associated with $\underline{\theta}_o$, and Σ_B, Σ_{gw} are matrix functions of $\underline{\theta}$ corresponding to the covariance structures. It is assumed that q is finite and the covariance structures satisfy the mild regularity conditions as given by Shapiro (1986, 1987). For example, a two-level factor analysis model is given by

$$\underline{x}_{gi} = \Lambda\xi_g + \epsilon_g + \Lambda\eta_{gi} + \delta_{gi} \ ,$$

where ξ_g, ϵ_g, η_{gi} and δ_{gi} are latent vectors at different levels, such that if $g \leq r$, η_{gi} and δ_{gi} are distributed as $N[\underline{0},\Phi_1]$ and $N[\underline{0},\Psi_1]$ respectively, and if $g > r$, their covariances matrices are equal to Φ_2 and Ψ_2 respectively; ξ_g and ϵ_g are distributed as $N[\underline{0},\Phi_B]$ and $N[\underline{0},\Psi_B]$, respectively, and Λ is the cross level factor loading matrix. The structures of the covariance matrices are given by

$$\Sigma_B = \Lambda\Phi_B\Lambda' + \Psi_B \qquad \Sigma_{gw} = \Lambda\Phi_1\Lambda' + \Psi_1 \ , \qquad g = 1, \ldots, r,$$

$$= \Lambda\Phi_2\Lambda' + \Psi_2 \ , \qquad g = r+1, \ldots, G,$$

and $\underline{\theta}$ consists of the unknown parameters in Λ, Φ_B, Ψ_B, Φ_1, Φ_2, Ψ_1, and Ψ_2. If $\Phi_1 = \Phi_2$ and $\Psi_1 = \Psi_2$ then all Σ_{gw} are equal and invariant over groups.

3. MAXIMUM LIKELIHOOD AND GENERALIZED LEAST SQUARES ESTIMATES

Consider the overall data vector $\underline{z}_g' = (\underline{x}_{g1}', \ldots, \underline{x}_{gN_g}')$, which may have different length for different g values. The distribution of $\underline{z}_g$ is $N[\underline{0}, \Sigma_{gz}]$, where

$$\Sigma_{gz} = \Sigma_{gz}(\underline{\theta}) = \underline{J}_g \otimes \Sigma_B + \underline{I}_g \otimes \Sigma_{gw} \ ,$$

with $\underline{J}_g$ is the square matrix of unit elements of order N_g, and $\underline{I}_g$ is the identity matrix of order N_g. By standard reasoning, the maximum likelihood (ML) estimate $\hat{\underline{\theta}}$ of $\underline{\theta}_0$ is the vector that minimizes the discrepancy function

$$F(\underline{\theta}) = \sum_g (\log|\Sigma_{gz}| + \underline{z}_g'\Sigma_{gz}^{-1}\underline{z}_g), \tag{3}$$

where Σ denotes the summation over g. It can be shown that (Graybill, 1983, chapter 8),

$$|\Sigma_{gz}| = |\Sigma_{gw}|^{N_g-1}|\Sigma_{gw} + N_g\Sigma_B| \ , \text{ and } \Sigma_{gz}^{-1} = (\underline{I}_g \otimes \Sigma_{gw}^{-1}) - (\underline{J}_g \otimes \Sigma_g^*) \ ,$$

where $\Sigma_g^* = N_g^{-1}(\Sigma_{gw}^{-1} - \Sigma_g^{-1})$ with $\Sigma_g = \Sigma_{gw} + N_g\Sigma_B$. Substitute these results in (3), it can be shown that $F(\underline{\theta}) = F_1(\underline{\theta}) + F_2(\underline{\theta})$, where

$$F_1(\underline{\theta}) = \sum_g [(N_g - 1)\log|\Sigma_{gw}| + N_g^{-1}\,\mathrm{tr}\{\Sigma_{gw}^{-1}\sum_{i\neq j}\sum (\underline{y}_{gii} - \underline{y}_{gij})\}] \ , \tag{4}$$

$$F_2(\underline{\theta}) = \sum_g \{\log|\Sigma_g| + N_g^{-1}\,\mathrm{tr}(\Sigma_g^{-1}\sum\sum_{ij} \underline{y}_{gij})\} \ , \tag{5}$$

with $\underline{y}_{gii} = \underline{x}_{gi}\underline{x}_{gi}'$, $\underline{y}_{gij} = \underline{x}_{gi}\underline{x}_{gj}'$, and $\sum_{i\neq j}\sum$ denotes the summation over i and j with $i \neq j$. Based on similar arguments as in Anderson & Rubin (1956) or Browne (1974), it can be proved that $\hat{\underline{\theta}}$ is a consistent estimator of $\underline{\theta}_0$. The asymptotic distribution of $\hat{\underline{\theta}}$ is given by the following theorem.

THEOREM 1. The asymptotic distribution of $N^{1/2}(\hat{\underline{\theta}} - \underline{\theta}_0)$ is normal with zero mean vector and covariance matrix $2\underline{H}_0^{-1}$, where

$$\underline{H} = G^{-1}\sum_g \{\Delta_{gw}(\Sigma_{gw}^{-1} \otimes \Sigma_{gw}^{-1})\Delta_{gw}' + N_g^{-1}\Delta_g(\Sigma_g^{-1} \otimes \Sigma_g^{-1})\Delta_g'\} \ , \tag{6}$$

with $\underline{H}_0 = \underline{H}(\underline{\theta}_0)$, $\Delta_{gw} = \partial\Sigma_{gw}/\partial\underline{\theta}$ and $\Delta_g = \partial\Sigma_g/\partial\underline{\theta}$.

From theorem 1, the standard error estimate of $\hat{\underline{\theta}}$ can be obtained from the diagonal elements of $2\underline{H}(\hat{\underline{\theta}})^{-1}$. Moreover, since N_g can be different, the results are true for unbalanced sampling designs. It should also be noted that for single level structural equation models, $\underline{v}_g^*$ is not exist, $\Sigma_B = \underline{0}$, $\Delta_g = \Delta_{gw}$ and $\Sigma_g^{-1} = \Sigma_{gw}^{-1}$. Hence the last

term in (6) becomes $N_g^{-1}\Delta_{gw}(\Sigma_{gw}^{-1} \otimes \Sigma_{gw}^{-1})\Delta_{gw}'$, and tends to zero as N_g tends to infinity. This implies that under this special case, the asymptotic distribution of $N^{1/2}(\hat{\underline{\theta}} - \underline{\theta}_0)$ is the same as the ordinary multiple group structural equation models.

Another important estimation method in structural equation models is the generalized least squares approach. Hence, we will develop the basic statistical properties of this approach for two-level structural equation models, and investigate its relationship with the theory in maximum likelihood approach under the assumption of normality. First, the following lemma is presented.

LEMMA 1. The random vectors $\text{vecs}\sum_{i\neq j}\sum(\underline{y}_{gii} - \underline{y}_{gij} - \Sigma_{gw})$ and $\text{vecs}(\sum\sum_{ij}\underline{y}_{gij} - N_g\Sigma_g)$ are uncorrelated.

Based on Lemma 1, the following generalized least squares (GLS) objective function is defined as

$$Q(\underline{\theta}|\underline{W}) = Q_1(\underline{\theta}|\underline{W}) + Q_2(\underline{\theta}|\underline{W}) \ , \tag{7}$$

where

$$Q_1(\underline{\theta}|\underline{W}) = \sum_g \{\text{vecs}\sum_{i\neq j}\sum(\underline{y}_{gii}-\underline{y}_{gij}-\Sigma_{gw})\}'\underline{W}_{g1}\{\text{vecs}\sum_{i\neq j}\sum(\underline{y}_{gii}-\underline{y}_{gij}-\Sigma_{gw})\}, \tag{8}$$

$$Q_2(\underline{\theta}|\underline{W}) = \sum_g \{\text{vecs}(\sum\sum_{ij}\underline{y}_{gij} - N_g\Sigma_g)\}'\underline{W}_{g2}\{\text{vecs}(\sum\sum_{ij}\underline{y}_{gij} - N_g\Sigma_g)\} \ , \tag{9}$$

in which $\underline{W}_{g1}$ and $\underline{W}_{g2}$ are p* by p* positive definite matrices which converge in probability to some positive definite matrices $\underline{W}_{g1}^*$ and $\underline{W}_{g2}^*$ respectively. The generalized least squares estimate $\tilde{\underline{\theta}}$ of $\underline{\theta}_0$ is the vector that minimizes $Q(\underline{\theta}|\underline{W})$. Again, $\tilde{\underline{\theta}}$ is a consistent estimator of $\underline{\theta}_0$. The asymptotic distribution of $\tilde{\underline{\theta}}$ is given by the following theorem.

THEOREM 2. The asymptotic distribution of $N^{1/2}(\tilde{\underline{\theta}} - \underline{\theta}_0)$ is normal with zero mean vector and covariance matrix

$$\underline{C} = 2[E\{\partial^2 Q(\underline{\theta}|\underline{W})/\partial\underline{\theta}\partial\underline{\theta}'\}/N]^{-1}\underline{H}^*[E\{\partial^2 Q(\underline{\theta}|\underline{W})/\partial\underline{\theta}\partial\underline{\theta}'\}/N]^{-1}|_{\underline{\theta}=\underline{\theta}_0}, \tag{10}$$

where $N = N_1 + N_2 + \ldots N_G$, and

$$\underline{H}^* = \frac{4}{G}\sum\{N_g^6\Delta_{gw}\underline{K}_p\underline{W}_{g1}^*(\Sigma_{gw}\otimes\Sigma_{gw})\underline{M}_p\underline{W}_{g1}^*\underline{K}_p'\Delta_{gw}' + N_g^3\Delta_{g}\underline{K}_p\underline{W}_{g2}^*(\Sigma_g\otimes\Sigma_g)\underline{M}_p\underline{W}_{g2}^*\underline{K}_p'\Delta_g'\}.$$

generalized least squares estimation. One of the best choice is to take $\underline{W}_{g1}$ and $\underline{W}_{g2}$ respectively equal to some consistent estimates of the inverse of the covariance matrices of $\text{vecs}\sum_{i\neq j}(\underline{y}_{gii} - \underline{y}_{gij} - \Sigma_{gw})$ and $\text{vecs}\sum\sum_{ij}(\underline{y}_{gij} - N_g\Sigma_g)$, see Ferguson (1958), and Browne (1984). Under the assumption of normality, the covariance matrix of $\text{avecs}\sum_{i\neq j}(\underline{y}_{gii} - \underline{y}_{gij} - \Sigma_{gw})$ is equal to $2(N_g - 1)\, N_g^{\,2}\underline{K}_p'(\Sigma_{gw}\otimes\Sigma_{gw})\underline{K}_p$ and the covariance matrix of $\text{vecs}(\sum\sum_{ij}\underline{y}_{gij} - N_g\Sigma_g)$ is $2N_g^{\,2}\underline{K}_p'(\Sigma_g\otimes\Sigma_g)\underline{K}_p$. Thus, the optimal choice of the weight matrices is

$$\bar{\underline{W}}_{g1} = \frac{1}{2}\,(N_g - 1)^{-1}N_g^{-2}\underline{K}_p^{-}(\bar{\Sigma}_{gw}^{-1}\otimes\bar{\Sigma}_{gw}^{-1})\underline{K}_p^{-\prime}, \tag{11}$$

$$\bar{\underline{W}}_{g2} = \frac{1}{2}\,N_g^{-2}\underline{K}_p^{-}(\bar{\Sigma}_g^{-1}\otimes\bar{\Sigma}_g^{-1})\underline{K}_p^{-\prime}, \tag{12}$$

where $\bar{\Sigma}_{gw}$ and $\bar{\Sigma}_g$ are consistent estimates of Σ_{gwo} and Σ_{go} respectively. Under these choices of the weight matrices, $\underline{H}^*(\underline{\theta}_o)$ converges in probability to $\underline{H}_o$, $\underline{C} = 2\underline{H}_o^{-1}$, and the GLS estimation has the following nice property.

THEOREM 3. The asymptotic covariance matrix of $\tilde{\underline{\theta}}$ is bounded below by $2\underline{H}_o^{-1}$ in the sense that $\underline{C} - 2\underline{H}_o^{-1}$ is a positive definite matrix. This bound is attained if the weight matrices are taken to be the optimal matrices as defined by (11) and (12); in this case, the asymptotic distribution of $N^{1/2}(\underline{\theta} - \underline{\theta}_o)$ is normal with zero mean vector and covariance matrix $2\underline{H}_o^{-1}$.

The following theorem establishes the relationship between the maximum likelihood estimator and the generalized least squares estimator.

THEOREM 4. Under the optimal choice of the weight matrices, $\tilde{\underline{\theta}}$ is asymptotically equivalent to $\hat{\underline{\theta}}$.

Computationally, in general, some iterative algorithms have to be used to obtain $\hat{\underline{\theta}}$ or $\underline{\theta}$. It is well-known that (Lee & Jennrich, 1979) the scoring algorithm and the Gauss-Newton algorithm are efficient and

dependable procedures. The basic step of the algorithms are respectively defined by:

Scoring : $\Delta\underline{\theta} = -\ (E\partial^2 F/\partial\underline{\theta}\partial\underline{\theta}')^{-1}(\partial F/\partial\underline{\theta})$ (13)

and

Gauss-Newton : $\Delta\underline{\theta} = -\ (E\partial^2 Q/\partial\underline{\theta}\partial\underline{\theta}')^{-1}(\partial Q/\partial\underline{\theta})$. (14)

Under the optimal choice of the weight matrices, it follows from similar arguments as in Lee and Jennrich (1979) that if $\bar{\Sigma}_{gw}$ and $\bar{\Sigma}_{g}$ were respectively set equal to Σ_{gw} and Σ_{g} with $\underline{\theta}$ updated from iteration to iteration, then $\partial F/\partial\underline{\theta} = \partial Q/\partial\underline{\theta}$ and $E(\partial^2 F/\partial\underline{\theta}\partial\underline{\theta}') = E(\partial^2 Q/\underline{\partial}\partial\underline{\theta}')$ during the minimization process. Hence the scoring algorithm is an iteratively reweighted Gauss-Newton algorithm. Consequently, results reported in Jennrich & Moore (1975), Lee (1979), and Green (1984) are also valid in this situation.

4. A TWO-STAGE ESTIMATION PROCEDURE

In this section, we will develop a more convenient two-stage estimation procedure which can produce acceptable solutions via standard programs such as LISREL VII (Joreskog & Sorbom, 1988) or EQS (Bentler, 1989). The first stage estimates Σ_B and Σ_{gw} without imposing any structures, while the second stage estimates the structural parameters $\underline{\theta}_0$ based on the defined structures and the statistical properties of the first stage estimates.

We first consider one arbitrary individual group g. It follows from similar arguments in Section 3 that the maximum likelihood estimates $\tilde{\Sigma}_B$ and $\tilde{\Sigma}_{gw}$ of Σ_{Bo} and Σ_{gwo} are the matrices that minimize the discrepancy function

$$F(\Sigma_B,\Sigma_{gw}) = (N_g-1)\log|\Sigma_{gw}| + N_g^{-1}tr(\Sigma_{gw}^{-1}\underline{T}_w) + \log|\Sigma_g| + N_g^{-1}tr(\Sigma_g^{-1}\underline{T}), \quad (15)$$

where $\underline{T}_w = \sum\sum_{i\neq j} (\underline{y}_{gii} - \underline{y}_{gij})$ and $\underline{T} = \sum\sum_{ij} \underline{y}_{gij}$. Now, consider the following one-to-one transformation,

$$\Sigma_{g1} = \Sigma_{gw}, \text{ and } \Sigma_{g2} = \Sigma_g = \Sigma_{gw} + N_g\Sigma_B \quad (16)$$

and its inverse,

$$\Sigma_{gw} = \Sigma_{g1}, \text{ and } \Sigma_B = (\Sigma_{g2} - \Sigma_{g1})/N_g. \tag{17}$$

By means of this one-to-one function, we transform $\{\Sigma_{gw}, \Sigma_B\}$ to a new set of independent parameters $\{\Sigma_{g1}, \Sigma_{g2}\}$. The likelihood function defined by (15) can be expressed as $F_1(\Sigma_{g1}) + F_2(\Sigma_{g2})$, where

$$F_1(\Sigma_{g1}) = (N_g - 1)\log|\Sigma_{g1}| + N_g^{-1}\mathrm{tr}(\Sigma_{g1}^{-1}\,\underline{T}_w), \tag{18}$$

and

$$F_2(\Sigma_{g2}) = \log|\Sigma_{g2}| + N_g^{-1}\mathrm{tr}(\Sigma_{g2}^{-1}\,\underline{T}). \tag{19}$$

The first stage maximum likelihood estimates $\tilde{\Sigma}_{g1}$ and $\tilde{\Sigma}_{g2}$ can then be separately obtained by minimizing $F_1(\Sigma_{g1})$ and $F_2(\Sigma_{g2})$, respectively. Thus, we have

$$\tilde{\Sigma}_{g1} = \underline{T}_w/N_g(N_g - 1)\ , \tag{20}$$

and

$$\tilde{\Sigma}_{g2} = \underline{T}/N_g. \tag{21}$$

It follows from (20), (21), and (17) that the maximum likelihood estimates of Σ_{gwo} and Σ_{Bo} are respectively equal to

$$\tilde{\Sigma}_{gw} = \underline{T}_w/N_g(N_g - 1), \tag{22}$$

and

$$\tilde{\Sigma}_{gB} = \underline{T}_B/N_g(N_g - 1), \tag{23}$$

where $\underline{T}_B = \sum_{i\neq j}\sum \underline{y}_{gij}$.

Repeating the analysis for all other independent groups, we get $\tilde{\Sigma}_{gw}$ and $\tilde{\Sigma}_{gB}$, $g = 1, \ldots, G$. The final estimate of Σ_{Bo} is taken to be

$$\tilde{\Sigma}_B = \sum_g \tilde{\Sigma}_{gB}/G. \tag{24}$$

Clearly, $E(\tilde{\Sigma}_B) = \Sigma_{Bo}$, and $E(\tilde{\Sigma}_{gw}) = \Sigma_{gwo}$, hence $\tilde{\Sigma}_B$ and $\tilde{\Sigma}_{gw}$ are unbiased estimates. Moreover, it can be shown that

$$\tilde{\Sigma}_{gw} \overset{a}{=} \sum_i \underline{v}_{gi}\underline{v}_{gi}'/N_g - \sum_{i\neq j}\sum \underline{v}_{gi}\underline{v}_{gj}'/N_g(N_g-1)$$

$$\stackrel{a}{=} \sum_i (\underline{v}_{gi} - \bar{\underline{v}}_g)(\underline{v}_{gi} - \bar{\underline{v}}_g)'/N_g \ . \tag{25}$$

here '$\stackrel{a}{=}$' denotes convergence in probability and $\bar{\underline{v}}_g = \sum_i \underline{v}_{gi}/N_g$. Since $\underline{v}_{gi}$ are identically and independently distributed as $N[\underline{0}, \Sigma_{gwo}]$, the following theorem is valid.

THEOREM 5. Suppose $\{\underline{x}_{gi}; g=1, \ldots, G, i=1, \ldots, N_g\}$ is a sample of dependent observations as defined in (1), and $\tilde{\Sigma}_{gw}$ is an estimate of Σ_{gwo} defined in (22) for g=1, ..., G. Then for g=1, ..., G, $N_g^{1/2}\text{vecs}(\tilde{\Sigma}_{gw} - \Sigma_{gwo})$ is asymptotically and mutually independently distributed as $N(\underline{0}, 2\underline{K}_p'(\Sigma_{gwo} \otimes \Sigma_{gwo})\underline{K}_p)$.

It should be noted that the proof of theorem 5 does not depend on the values of N_g and G, hence this theorem and the subsequent analyses are true even with unbalanced designs and/or a small number of groups available at the group level.

From theorem 5, it can be seen that the first-stage estimate $\tilde{\Sigma}_{gw}$ has the same asymptotic distribution as a sample covariance matrix that based on identically and independently distributed observations from $N[\underline{0}, \Sigma_{gwo}]$, and $\tilde{\Sigma}_{gw}$ and $\tilde{\Sigma}_{hw}$ are asymptotically independent for $g \neq h$. Hence, in analyzing the within-group covariance structures with these G independent samples at the group level, the approach of normal theory covariance structure analysis with multiple groups (see, e.g., Joreskog, 1971; Lee & Tsui, 1982) can be applied with $\tilde{\Sigma}_{gw}$ plays exactly the same role of the sample covariance matrix. Here, similar to Joreskog (1971), and Lee and Tsui (1982), we also require a large N_g, which may be unequal for unbalanced designs, in order to obtain the desirable asymptotic statistical properties for inference of the model. It should be noted that if one is only interested in analyzing the within-group structures, the number of groups at the group level can be small.

The analysis of the between-group covariance structure in the present model is difficult if G is small. Therefore, similar to the classical theory of structural equation models, small sample inference

is still an important open area for future research. However, if G is large and tends to infinity in the same order as N_g, it can be shown that the between-group covariance structure can be analyzed alone or simultaneously with the within-group structures. From (23) and (24),

$$\tilde{\Sigma}_B = G^{-1} \sum_g (\underline{v}_g^* \underline{v}_g^{*\prime} + \underline{v}_g^* \bar{\underline{v}}_g^{\prime} + \bar{\underline{v}}_g \underline{v}_g^{*\prime} + \bar{\underline{v}}_g \bar{\underline{v}}_g^{\prime})$$

$$\stackrel{a}{=} G^{-1} \sum_g \underline{v}_g^* \underline{v}_g^{*\prime} , \tag{26}$$

where $\underline{v}_g^*/G^{1/2}$ is independently and identically distributed as $N(\underline{0}, \Sigma_{Bo}/G)$. Thus, from the central limit theorem, the following theorem is valid.

THEOREM 6. Suppose $\{\underline{x}_{gi}; g=1, \ldots, G, i=1, \ldots N_g\}$ is a sample of dependent observations as defined in (1), and $\tilde{\Sigma}_B$ is an estimate of Σ_{Bo} defined in (24). Then the asymptotic distribution of $G^{1/2}\text{vecs}(\tilde{\Sigma}_B - \Sigma_{Bo})$ is $N[\underline{0}, 2\underline{K}_p^{\prime}(\Sigma_{Bo} \otimes \Sigma_{Bo})\underline{K}_p]$ as G tends to infinity.

Hence, the asymptotic distribution of $\tilde{\Sigma}_B$ is equal to the distribution of a sample covariance matrix as obtained from a sample of independent random observations of a multivariate normal distribution. Moreover, since asymptotically, $\tilde{\Sigma}_{gw}$ depends only on $\underline{v}_{gi}$ while $\tilde{\Sigma}_B$ depends only on $\underline{v}_g^*$, and $\underline{v}_{gi}$ and $\underline{v}_g^*$ are independent, hence, $\tilde{\Sigma}_{gw}$ and $\tilde{\Sigma}_B$ are asymptotically independent and the following theorem is valid.

THEOREM 7. Asymptotically, $\tilde{\Sigma}_B$, $\tilde{\Sigma}_{1w}$, ..., $\tilde{\Sigma}_{Gw}$ are mutually independent.

Therefore, under the situation that G is sufficiently large, the between-group covariance structure can be analyzed. If there are no cross level parameters, the analysis can be carried out with $\tilde{\Sigma}_B$ alone, based on results in theorem 6 and the classical normal theory of covariance structure analysis. If cross level parameters are involved, the analysis can be performed together with $\tilde{\Sigma}_{gw}$, g=1, ..., G, by the normal theory covariance structure analysis of multiple groups, based on justification provided by theorem 7. Here, we have G+1 groups, with $\tilde{\Sigma}_B$

and $\tilde{\Sigma}_{gw}$, g=1, ..., G, play the role of the sample covariance matrices. Computationally, standard programs such as LISREL VII (Joreskog & Sorbom, 1988) and EQS (Bentler, 1989) can be used to obtain the generalized least squares (GLS) or the analogous maximum likelihood (AML) solutions. Therefore, statistical analyses of the general two-level model with unbalanced designs can be performed in a manner similar to the multiple group analysis of Joreskog (1971). For analyses of more complicated nonlinear functional relationships among the parameters, the approach given in Lee and Tsui (1982), and Lee (1985) can be employed.

Finally, it should be pointed out that if there are no group level effects, then $\underline{y}_g^* = \underline{0}$, $\Sigma_B = \underline{0}$, $\underline{x}_{gi} = \underline{v}_{gi}$, and the model and its analysis reduce back to ordinary single level multiple groups covariance structure analysis.

To illustrate the results developed and attain a rough idea about the accuracy of the estimation method, an artificial example will be reported. There are two main purposes of this example; (i) to illustrate the between and the within aspects of the model, and (ii) to demonstrate that ignoring the existence of the group level effect might lead to incorrect solutions. Suppose G = 50, $\Sigma_{Bo} = \Lambda_{Bo}\Phi_{Bo}\Lambda_{Bo}' + \Psi_{Bo}$, and $\Sigma_{gwo} = \Lambda_{go}\Phi_{go}\Lambda_{go}' + \Psi_{go}$ for all g, where

$$\Lambda_{Bo}' = \begin{bmatrix} 0.6 & 0.6 & 0^* & 0^* \\ 0^* & 0^* & 0.7 & 0.7 \end{bmatrix}, \quad \Phi_{Bo} = \begin{bmatrix} 1.0^* & 0.7 \\ 0.7 & 1.0^* \end{bmatrix}, \quad \Psi_{Bo} = 0.7\underline{I},$$

$$\Lambda_{go}' = \begin{bmatrix} 0.6 & 0.6 & 0^* & 0^* \\ 0^* & 0^* & 0.7 & 0.7 \end{bmatrix}, \quad \Phi_{go} = \begin{bmatrix} 1.0^* & 0.4 \\ 0.4 & 1.0^* \end{bmatrix}, \quad \Psi_{go} = 0.5\underline{I},$$

where the non-diagonal elements of Ψ_{Bo} and Ψ_{go}, and parameters with asterisk are treated as fixed known parameters. Fifty independent random vectors $\underline{y}_g^*$ were generated from $N(\underline{0}, \Sigma_{Bo})$, one for each group g. To consider an unbalanced design, $N_i = 100$, $N_j = 75$, $N_k = 50$ independent random observations $\underline{v}_{gi}$ were generated according to $N(\underline{0}, \Sigma_{gwo})$, i = 1, ..., 10; j = 11, ..., 30; k = 31, ..., 50. Thus, a total of 3500 dependent observations $\underline{x}_{gi}$ of the form (1) were obtained. These $\underline{x}_{gi}$ were

considered as the observed random vectors for analyses.

Based on the above development, fifty matrices $\tilde{\Sigma}_{gw}$ were computed according to (22) and $\tilde{\Sigma}_B$ was computed according to (24). These matrices were then simultaneously analyzed using the multiple group (multi-sample) section of LISREL VII (Joreskog & Sorbom, 1988, chapter 9) under the following specifications:

$$\Lambda_B = \Lambda_1 = \ldots = \Lambda_{50} = \Lambda, \ \Phi_1 = \ldots = \Phi_{50}, \text{ and } \Psi_1 = \ldots = \Psi_{50}. \quad (27)$$

The GLS and AML estimates of the unknown parameters and their estimated standard errors are reported in Table 1.

TABLE 1
GLS and AML Estimates for Example in Section 4

Parameters	GLS	AML	Parameters	GLS	AML
$\Lambda(1,1)$	0.56 (0.03)	0.61 (0.03)			
$\Lambda(2,1)$	0.55 (0.03)	0.59 (0.03)			
$\Lambda(3,2)$	0.61 (0.03)	0.66 (0.03)			
$\Lambda(4,2)$	0.67 (0.03)	0.70 (0.03)			
$\Phi_B(2,1)$	0.92 (0.19)	0.95 (0.18)	$\Phi_g(2,1)$	0.44 (0.03)	0.41 (0.03)
$\Psi_B(1,1)$	0.44 (0.13)	0.47 (0.13)	$\Psi_g(1,1)$	0.46 (0.03)	0.41 (0.03)
$\Psi_B(2,2)$	0.50 (0.14)	0.59 (0.15)	$\Psi_g(2,2)$	0.52 (0.03)	0.46 (0.03)
$\Psi_B(3,3)$	0.60 (0.16)	0.62 (0.17)	$\Psi_g(3,3)$	0.53 (0.03)	0.47 (0.03)
$\Psi_B(4,4)$	0.73 (0.20)	0.81 (0.21)	$\Psi_g(4,4)$	0.46 (0.04)	0.38 (0.04)

Note: In all Tables, standard error estimates are in parentheses.

The chi-squared goodness-of-fit values of the GLS approach and the AML approach are equal to 461.57 and 450.49, respectively, each with 496 degrees of freedom. Hence, we got the expected result that the proposed model fits the observed sample data.

Now suppose one mistakenly analyzes the data without taking into account the group level effect by treating all the data as from one

population with covariance matrix $\Sigma = \Lambda\Phi\Lambda' + \Psi$. The covariance structure is then fitted based on the ordinary sample covariance matrix $\underline{S}$ of $\underline{x}_{gi}$. Estimates obtained from the GLS and ML approach in LISREL VII are reported in Table 2.

TABLE 2
GLS and AML Estimates for Example in Section 4 Based on $\underline{S}$

Parameters	GLS	AML	Parameters	GLS	AML
$\Lambda(1,1)$	0.71 (0.03)	0.71 (0.03)	$\Psi(1,1)$	0.99 (0.04)	1.00 (0.04)
$\Lambda(2,1)$	0.78 (0.03)	0.77 (0.03)	$\Psi(2,2)$	1.04 (0.04)	1.05 (0.04)
$\Lambda(3,2)$	0.85 (0.03)	0.85 (0.03)	$\Psi(3,3)$	1.20 (0.05)	1.22 (0.05)
$\Lambda(4,2)$	1.11 (0.04)	1.10 (0.04)	$\Psi(4,4)$	1.10 (0.07)	1.12 (0.07)
$\Phi(2,1)$	0.71 (0.03)	0.71 (0.03)			

Obviously, these are questionable estimates. Therefore, in practice, if data come from different groups, before pooling all data together for analyses, it is important to check whether the group level effect exists.

5. ANALYSES UNDER ARBITRARY AND ELLIPTICAL DISTRIBUTIONS

In this section, we will assume that G is sufficient large so that the between-group covariance structure can be analyzed together with the within-group covariance structures. Now, we consider the situation where the distributions of $\underline{v}_g^*$ and $\underline{v}_{gi}$ are not necessarily multivariate normal but have finite eighth-order moments and same assumptions about independence as before. The estimation of the overall structural parameter vector $\underline{\ell}_0$, will also be based on $\tilde{\Sigma}_{gw}$ and $\tilde{\Sigma}_B$ as given by (22) and (24), respectively. From the central limit theorem, it can be shown from (25) and (26) that the asymptotic distributions of $N_g^{1/2}\text{vecs}(\tilde{\Sigma}_{gw}-\Sigma_{gwo})$ and $G^{1/2}\text{vecs}(\tilde{\Sigma}_B-\Sigma_{Bo})$ are multivariate normal with

mean vector zero, and some covariance matrices Γ_{go} and Γ_{Bo}, respectively. Moreover, $\tilde{\Sigma}_B$, $\tilde{\Sigma}_{1w}$, ..., $\tilde{\Sigma}_{Gw}$ are also asymptotically independent. The GLS estimate $\underline{\tilde{\theta}}_A$ of $\underline{\theta}_o$ is defined as the vector that minimizes the following discrepancy function

$$Q(\underline{\theta}) = 2^{-1}[C\{\text{vecs}(\tilde{\Sigma}_B-\Sigma_B)\}'\underline{W}_B^{-1}\{\text{vecs}(\tilde{\Sigma}_B-\Sigma_B)\} + \sum_g C_g\{\text{vecs}(\tilde{\Sigma}_{gw}-\Sigma_{gw})\}'\underline{W}_g^{-1}\{\text{vecs}(\tilde{\Sigma}_{gw}-\Sigma_{gw})\}], \quad (28)$$

where $\underline{W}_B$, $\underline{W}_g$ are positive definite matrices that converges in probability to Γ_{Bo} and Γ_{go}, respectively, $C = G/N$ and $C_g = N_g/N$, with $N = G + N_1 + \ldots + N_G$. Under mild regularity conditions, it can be shown by similar arguments as in Ferguson (1958) and Browne(1984) that the asymptotic distribution of $\underline{\tilde{\theta}}_A$ is multivariate normal, and the asymptotic distribution of the goodness-of-fit statistic $2NQ(\underline{\tilde{\theta}}_A)$ is chi-squared with degrees of freedom $(G+1)p^*-q$, where q is the number of parameters in $\underline{\theta}_o$.

To define $Q(\underline{\theta})$, we require consistent estimates, $\underline{W}_B$ and $\underline{W}_g$, of Γ_{Bo} and Γ_{go}. It is well-known (Browne, 1984) that a typical element of Γ_{go} is given by

$$\Gamma_{go}(rs,uv) = \sigma_{go}(rsuv) - \Sigma_{gwo}(r,s)\Sigma_{gwo}(u,v), \quad (29)$$

where $\sigma_{go}(rsuv)$ is the fourth-order central moment of $\underline{v}_{gi}$. Since the eighth-order moments are finite, the sample fourth-order moment of $\underline{v}_{gi}$, $\tilde{\sigma}_g(rsuv)$ is a consistent estimate of $\sigma_{go}(rsuv)$. Since $\underline{v}_{gi}$ is unobservable, $\tilde{\underline{\sigma}}_g(rsuv)$ cannot be obtained directly from $\underline{v}_{gi}$. However, since $\underline{x}_{gi} - \bar{\underline{x}}_g = \underline{v}_{gi} - \bar{\underline{v}}_g$, where $\bar{\underline{x}}_g$ is the sample mean of $\underline{x}_{gi}$, we have

$$\tilde{\sigma}_g(rsuv)=N_g^{-1}\sum_i\{\underline{x}_{gi}(r)-\bar{\underline{x}}_g(r)\}\{\underline{x}_{gi}(s)-\bar{\underline{x}}_g(s)\}\{\underline{x}_{gi}(u)-\bar{\underline{x}}_g(u)\}\{\underline{x}_{gi}(v)-\bar{\underline{x}}_g(v)\}.$$

Thus, it follows from (29) that one possible choice of $\underline{W}_g$ in (28) is

$$\underline{W}_g(rs,uv) = \tilde{\sigma}_g(rsuv) - \tilde{\Sigma}_{gw}(r,s)\tilde{\Sigma}_{gw}(u,v). \quad (30)$$

It follows from similar reasoning as in Browne (1984) that if N_g is sufficiently large, $\underline{W}_g$ will be positive definite with probability 1. By exactly the same reasoning, a consistent estimate $\underline{W}_B(rs,uv)$ of $\Gamma_{Bo}(rs,uv)$ is given by

$$\underline{W}_B(rs,uv) = \tilde{\sigma}_B(rsuv) - \tilde{\Sigma}_B(r,s)\tilde{\Sigma}_B(u,v), \quad (31)$$

where

$$\tilde{\sigma}_B(rsuv)=G^{-1}\sum_g \{\bar{\underline{x}}_g(r)-\bar{\underline{x}}(r)\}\{\bar{\underline{x}}_g(s)-\bar{\underline{x}}(s)\}\{\bar{\underline{x}}_g(u)-\bar{\underline{x}}(u)\}\{\bar{\underline{x}}_g(v)-\bar{\underline{x}}(v)\},$$

with $\bar{\underline{x}} = \Sigma_g \bar{\underline{x}}_g/G$. For sufficiently large G, $\underline{W}_B$ will be positive definite with probability 1.

Therefore, based on (28), (30), (31) and the asymptotic distributions of $\tilde{\Sigma}_B$ and $\tilde{\Sigma}_{gw}$, it can be seen that the analysis of the between-group and the within-group covariance structures of the dependent two-level observations $\underline{x}_{gi}$ can be performed via the asymptotically distribution-free approach of the classical multiple group structural equation models with very slight modifications. Computationally, standard programs such as LISREL VII and EQS can be applied to obtain the solution by minimizing the discrepancy function $Q(\underline{\ell})$. Again, it should be noted that the above analysis is still valid even with unbalanced designs.

The implementation of the computer program to obtain the asymptotically distribution-free estimates involves heavy computation and large storage for the weight matrices $\underline{W}_g$ or $\underline{W}_B$. However, like classical structural equation models, if the distribution belongs to an elliptical class (Muirhead, 1982), the computation and storage involved will be greatly reduced.

Let the independent properties among $\underline{v}_g^*$ and $\underline{v}_{gi}$ be the same as before, and that their distributions belong to an elliptical class with zero mean vector and covariance matrix Σ_{Bo} and Σ_{gwo}, respectively.

Since the underlying distributions are elliptical, $G^{1/2}\text{vecs}(\tilde{\Sigma}_B - \Sigma_{Bo})$ and $N_g^{1/2}\text{vecs}(\tilde{\Sigma}_{gw} - \Sigma_{gwo})$ will converge in distribution to multivariate normal distributions with zero mean vector and some covariance matrices which only depend on the second moments and the kurtosis parameters κ_B and κ_g of the respective distributions. It can be shown by similar reasoning as in Bentler (1983) and Browne (1984) that the GLS function is

$$Q_E(\ell) = C[2^{-1}(\hat{\kappa}_B+1)^{-1}\text{tr}\{(\tilde{\Sigma}_B - \Sigma_B)\underline{U}_B^{-1}\}^2 - \hat{\delta}_B\{\text{tr}(\tilde{\Sigma}_B - \Sigma_B)\underline{U}_B^{-1}\}^2]$$

$$+ \sum_g C_g[2^{-1}(\hat{\kappa}_g+1)^{-1}\text{tr}\{(\tilde{\Sigma}_{gw} - \Sigma_{gw})\underline{U}_g^{-1}\}^2 - \hat{\delta}_g\{\text{tr}(\tilde{\Sigma}_{gw} - \Sigma_{gw})\underline{U}_g^{-1}\}^2], \qquad (32)$$

where $\underline{U}_B$ and $\underline{U}_g$ are positive definite matrices that converge in probability to Σ_{Bo} and Σ_{gwo} respectively, $\hat{\kappa}_B$ and $\hat{\kappa}_g$ are consistent estimates of κ_B and κ_g, and

$$\hat{\delta}_i = \hat{\kappa}_i/\{4(\hat{\kappa}_i+1)^2 + 2p\hat{\kappa}_i(\hat{\kappa}_i+1)\}, \qquad (33)$$

for i=B, and i=1, ..., G. To define the GLS function, we need to provide $\underline{U}_B$, $\underline{U}_g$, $\hat{\kappa}_B$ and $\hat{\kappa}_g$. Clearly, possible choices of $\underline{U}_B$ and $\underline{U}_g$ are $\tilde{\Sigma}_B$ and $\tilde{\Sigma}_{gw}$, respectively. Moreover, based on results from Mardia (1970), it can be proved that consistent estimates of κ_B and κ_g that based on dependent observations $\underline{x}_{gi}$ are respectively given by

$$\hat{\kappa}_B = \{Gp(p+2)\}^{-1} \sum_g \{(\bar{\underline{x}}_g - \bar{\underline{x}})' \underline{S}_{Bx}^{-1}(\bar{\underline{x}}_g - \bar{\underline{x}})\} - 1 , \qquad (34)$$

and

$$\hat{\kappa}_g = \{N_g p(p+2)\}^{-1} \sum_i \{(\underline{x}_{gi} - \bar{\underline{x}}_g)' \underline{S}_{gx}^{-1}(\underline{x}_{gi} - \bar{\underline{x}}_g)\} - 1 , \qquad (35)$$

where $\underline{S}_{Bx} = G^{-1} \sum_g (\bar{\underline{x}}_g - \bar{\underline{x}})(\underline{x}_g - \bar{\underline{x}})'$ and $\underline{S}_{gx} = N_g^{-1} \sum_i (\underline{x}_{gi} - \bar{\underline{x}}_g)(\underline{x}_{gi} - \bar{\underline{x}}_g)'$. Now, since the components of $Q_E(\ell)$ have the same statistical properties as their counter-parts in the single level models under elliptical distribution, the two-level GLS analysis of the between-group and within-group covariance structures can be carried out via slight modification of the existing theory. Again, the above development does

not require N_g to be equal, and hence can apply to unbalanced designs.

It should be noted that under the special case with $\kappa_B = \kappa_g = 0$, the results developed in this section will reduce back to the results obtained with the multivariate normal assumption. Moreover, again, if there are no second level group effects, the theory reduces back to classical multiple group covariance structure analysis with elliptical distributions.

6. ESTIMATION WITH SMALL N_g

To achieve the large sample properties in the previous sections, we have assumed that the number of level-one units in each group is large. In many applications, this condition may not be satisfied. Typical examples are individuals in households, animals in a litter, and workers in small office, etc. In this section, we will consider a slightly restrictive model and an estimation method to handle the situation where the number of groups at the group level is large but the numbers of level-one individuals can be small. We rearrange the data in J categories such that within each category, there are M_j groups which consist the same number of level-one units N_j. Let $G = M_1 + \ldots + M_J$ be the total number of groups at level-two. We will consider the situation where J is finite, N_j is a small number greater than 2, and G is large enough so that the central limit theorem can be applied. The model defined in (1) can be rewritten as

$$\underline{x}_{jki} = \underline{v}^*_{jk} + \underline{v}_{jki}. \quad i = 1,\ldots,N_j,\ k = 1,\ldots,M_j,\ j = 1,\ldots,J. \tag{36}$$

Now, since N_g is small, we have to assume that the within-group structures are invariant over groups, that is, $\Sigma_{gwo} = \Sigma_w(\underline{\theta}_o)$ for all g = 1, ..., G. Other assumptions of the model are the same as before. A similar two-stage estimation procedure for analysis of the present model will be discussed. Consider an arbitrary group in the j^{th} category, let

$$\underline{C}_{j,rs} = M_j^{-1} \sum_{k=1}^{M_j} \underline{x}_{jkr}\underline{x}_{jks}', \quad r, s = 1, \ldots, N_j,\ r \neq s\ ; \tag{37}$$

since $E(\underline{x}_{jkr}\underline{x}_{jks}') = E(\underline{v}_{jk}^*\underline{v}_{jk}^{*\prime}) = \Sigma_{Bo}$, we have $E(\underline{C}_{j,rs}) = \Sigma_{Bo}$. By the

law of large numbers, $\underline{C}_{j,rs}$ is a consistent estimate of Σ_{Bo}. Let

$$\underline{C}_j = \{N_j(N_j - 1)\}^{-1} \sum_{r \neq s} \sum \underline{C}_{j,rs} \tag{38}$$

be the pooled estimate for each pair (r, s). Since Σ_{Bo} is invariant over all the groups in all categories, the first stage estimate of Σ_{Bo} is taken as

$$\hat{\Sigma}_B = G^{-1} \sum_j M_j \underline{C}_j \tag{39}$$

which is the weighted average of the $\underline{C}_j$, $j = 1, \ldots, J$. Clearly, $\hat{\Sigma}_B$ is also an unbiased and consistent estimate of Σ_{Bo}. Moreover, from (37), (38), (39) and the law of large numbers, it can be shown that

$$\text{vecs}(\hat{\Sigma}_B) \overset{a}{=} G^{-1} \sum_{jk} \sum \text{vecs}(\underline{v}_{jk}^* \underline{v}_{jk}^{*\prime}). \tag{40}$$

Consider

$$G^{1/2} \text{vecs}(\hat{\Sigma}_B - \Sigma_{Bo}) \overset{a}{=} G^{-1/2} \sum_{jk} \sum \text{vecs}(\underline{v}_{jk}^* \underline{v}_{jk}^{*\prime} - \Sigma_{Bo}) , \tag{41}$$

since for $j = 1, \ldots, J$, $k = 1, \ldots, M_j$; $\text{vecs}(\underline{v}_{jk}^* \underline{v}_{jk}^{*\prime} - \Sigma_{Bo})$ is identically and independently distributed with mean vector zero and covariance matrix $2\underline{K}_p'(\Sigma_{Bo} \otimes \Sigma_{Bo})\underline{K}_p$, by the central limit theorem, the following theorem is valid.

THEOREM 8. The asymptotic distribution of $G^{1/2} \text{vecs}(\hat{\Sigma}_B - \Sigma_{Bo})$ is multivariate normal with zero mean vector and covariance matrix $2\underline{K}_p'(\Sigma_{Bo} \otimes \Sigma_{Bo})\underline{K}_p$.

It should be noted that the proof of theorem 8 is true even with unbalanced designs and small numbers of level-one units. From theorem 8, it can be seen that $\text{vecs}(\hat{\Sigma}_B)$ has exactly the same asymptotic distribution as the sample covariance matrix that obtained from a sample of multivariate normal random vectors with covariance matrix Σ_{Bo}.

Now consider the estimation of Σ_{wo}. Since $\text{cov}(\underline{v}_{jk}^*, \underline{v}_{jki}) = \underline{0}$, it follows from (36) that the covariance matrix of $\underline{x}_{jki}$ is equal to

$$\Sigma_o = \Sigma_{Bo} + \Sigma_{wo} \ . \tag{42}$$

Consider an arbitrary category j and for i = 1, ..., N_j, an estimate of Σ_o is given by

$$\underline{R}_{ji} = M_j^{-1} \sum_k \underline{x}_{jki}\underline{x}_{jki}' \ . \tag{43}$$

The average estimate based on the j^{th} category is given by

$$\underline{R}_j = N_j^{-1} \sum_i \underline{R}_{ji} = N_j^{-1} \sum_i M_j^{-1} \sum_k \underline{x}_{jki}\underline{x}_{jki}' \ . \tag{44}$$

Clearly, $\underline{R}_j$ is an unbiased and consistent estimate of Σ_o. Hence, from (42) and (43), an unbiased and consistent estimate of Σ_{wo} is given by

$$\underline{U}_j = \underline{R}_j - \underline{C}_j \ . \tag{45}$$

The final first-stage estimate of Σ_{wo} is then taken to be the weighted average of $\underline{U}_j$:

$$\hat{\Sigma}_w = L^{-1} \sum_j N_j M_j \underline{U}_j \ , \tag{46}$$

where $L = \sum_j N_j M_j$. Clearly $\hat{\Sigma}_w$ is also an unbiased and consistent estimate of Σ_{wo}.

From (38), (44), (45), (46), and the law of large numbers, it can be shown that

$$\text{vecs}(\hat{\Sigma}_w) \overset{a}{=} L^{-1} \sum_{jrk} \text{vecs}(\underline{v}_{jkr}\underline{v}_{jkr}'). \tag{47}$$

Since $\text{vecs}(\underline{v}_{jkr}\underline{v}_{jkr}')$, r = 1, ..., N_j, k = 1, ..., M_j, j = 1, ..., J is identically and independently distributed with mean vector $\text{vecs}(\Sigma_{wo})$ and covariance matrix $2\underline{K}_p'(\Sigma_{wo} \otimes \Sigma_{wo})\underline{K}_p$. By the central limit theorem, the following theorem is true.

THEOREM 9. The asymptotic distribution of $L^{1/2}\text{vecs}(\hat{\Sigma}_w - \Sigma_{wo})$ is multivariate normal with zero mean vector and covariance matrix $2\underline{K}_p'(\Sigma_{wo} \otimes \Sigma_{wo})\underline{K}_p$

It should be pointed out that theorem 9 is true even with

unbalanced designs and small numbers of level-one units. It is clear from this theorem that the estimate $\text{vecs}(\hat{\Sigma}_w)$ also has exactly the same asymptotic distribution as the sample covariance matrix that obtained from a sample of multivariate normal distribution with covariance matrix Σ_{wo}. Moreover, since $\underline{v}^*_{jk}$ and $\underline{v}_{jki}$ are independent, the following theorem is also valid.

THEOREM 10. Asymptotically, $\hat{\Sigma}_B$ and $\hat{\Sigma}_w$ are independent.

Based on results in theorems 9 and 10, the between-group and the within-group structures can be analyzed separately by the normal theory of structural equation models with multiple groups, with $\hat{\Sigma}_B$ and $\hat{\Sigma}_w$ play the role of the sample covariance matrix. Computationally, standard programs LISREL VII and EQS can be used in obtaining the GLS and AML solutions. To save space, the standard analysis is not presented.

In order to illustrate the above results in analyzing between-group and within-group covariance structures with the standard program LISREL VII, and give a rough idea about the accuracy of the estimates, the artificial example in Lee & Poon (1992) is presented. In this example, $p = 4$, $J = 3$, $N_1 = 4$, $N_2 = 5$, $N_3 = 6$, $M_1 = 100$, $M_2 = 50$, and $M_3 = 50$. Random vectors $\underline{v}_{jk}^*$ and $\underline{v}_{jki}$ were respectively simulated from $N[\underline{0}, \Sigma_{Bo}]$ and $N[\underline{0}, \Sigma_{wo}]$ with

$$\Sigma_{Bo} = \Lambda_{Bo}\,\Phi_{Bo}\,\Lambda_{Bo}' + \Psi_{Bo}, \qquad \Sigma_{wo} = \Lambda_{wo}\,\Phi_{wo}\,\Lambda_{wo}' + \Psi_{wo}, \tag{48}$$

where Φ_{Bo} and Φ_{wo} are 2 by 2 symmetric matrices and Ψ_{Bo} and Ψ_{wo} are diagonal matrices. Only the $\underline{x}_{jki}$ obtained via (36) was treated as the available observed random vector. The population values of the parameter matrices are given by

$$\Lambda_{Bo}' = \begin{bmatrix} 0.8 & 0.8 & 0^* & 0^* \\ 0^* & 0^* & 0.8 & 0.8 \end{bmatrix}, \quad \Phi_{Bo} = \begin{bmatrix} 1.0^* & 0.6 \\ 0.6 & 1.0^* \end{bmatrix},$$

$$\Lambda_{wo}' = \begin{bmatrix} 0.4 & 0.4 & 0^* & 0^* \\ 0^* & 0^* & 0.4 & 0.4 \end{bmatrix}, \quad \Phi_{wo} = \begin{bmatrix} 1.0^* & 0.3 \\ 0.3 & 1.0^* \end{bmatrix},$$

and $\Psi_{Bo} = 0.6\underline{I}$, $\Psi_{wo} = 0.7\underline{I}$, where the non-diagonal zero elements of Ψ_{Bo} and Ψ_{wo}, and parameters with an asterisk were treated as fixed

parameters not to be estimated. The first stage estimates $\hat{\Sigma}_B$ and $\hat{\Sigma}_W$ were obtained from $\underline{x}_{jki}$ by equations (39) and (46). Based on these estimates of the covariance matrices, the generalized least squares and the maximum likelihood type estimates of the structural parameters in the between-group covariance matrix and the within-group covariance matrix were first separately obtained via the LISREL VII program. We observe that the estimates obtained are quite close to the true population values, and as expected, the GLS parameter estimates and standard error estimates are extremely close to the AML type estimates. Hence, only the AML estimates are presented in Table 3.

TABLE 3
GLS Estimates for Example in Section 6

Parameters	GLS	Parameters	GLS
$\Lambda_B(1,1)$	0.72 (0.10)	$\Lambda_W(1,1)$	0.44 (0.09)
$\Lambda_B(2,1)$	0.81 (0.10)	$\Lambda_W(2,1)$	0.33 (0.07)
$\Lambda_B(3,2)$	0.82 (0.11)	$\Lambda_W(3,2)$	0.36 (0.08)
$\Lambda_B(4,2)$	0.59 (0.10)	$\Lambda_W(4,2)$	0.48 (0.10)
$\Phi_B(2,1)$	0.57 (0.10)	$\Phi_W(2,1)$	0.47 (0.11)
$\Psi_B(1,1)$	0.67 (0.12)	$\Psi_W(1,1)$	0.67 (0.08)
$\Psi_B(2,2)$	0.51 (0.14)	$\Psi_W(2,2)$	0.73 (0.05)
$\Psi_B(3,3)$	0.27 (0.17)	$\Psi_W(3,3)$	0.80 (0.06)
$\Psi_B(4,4)$	0.90 (0.12)	$\Psi_W(4,4)$	0.61 (0.09)

We find that the chi-square values for the goodness-of-fit test of the between-group structures are almost identically equal to 0.67; and the corresponding chi-square values for the fit of the within-group structures are also almost identically equal to 0.57. With one degree of freedom, these values indicate the expected result that the proposed covariance structures fit the sample data.

Suppose we wish to analyze the data further in order to test the following null hypotheses across levels :

$$\Lambda_{Bo} = \Lambda_{wo} \ (= \Lambda_o). \tag{49}$$

Under this situation, the covariance structures are defined by

$$\Sigma_{Bo} = \Lambda_o \, \Phi_{Bo} \, \Lambda_o' + \Psi_{Bo}, \quad \Sigma_{wo} = \Lambda_o \, \Phi_{wo} \, \Lambda_o' + \Psi_{wo}, \tag{50}$$

with overlapping parameter matrix Λ_o. Based on the results developed above, the model was analyzed as a normal theory multiple group covariance structure analysis problem, with input matrices $\hat{\Sigma}_B$ and $\hat{\Sigma}_w$ to the LISREL VII program. Again, we observed that the GLS estimates are close to the AML estimates which are presented in Table 4.

TABLE 4
AML Estimates for Example in Section 6

Parameters	AML	Parameters	AML
$\Lambda_B(1,1)$	0.48 (0.07)	$\Lambda_w(1,1)$	0.48 (0.07)
$\Lambda_B(2,1)$	0.43 (0.06)	$\Lambda_w(2,1)$	0.43 (0.06)
$\Lambda_B(3,2)$	0.45 (0.07)	$\Lambda_w(3,2)$	0.45 (0.07)
$\Lambda_B(4,2)$	0.45 (0.06)	$\Lambda_w(4,2)$	0.45 (0.06)
$\Phi_B(2,1)$	0.89 (0.19)	$\Phi_w(2,1)$	0.42 (0.08)
$\Psi_B(1,1)$	0.82 (0.11)	$\Psi_w(1,1)$	0.66 (0.07)
$\Psi_B(2,2)$	0.81 (0.10)	$\Psi_w(2,2)$	0.69 (0.06)
$\Psi_B(3,3)$	0.56 (0.09)	$\Psi_w(3,3)$	0.72 (0.07)
$\Psi_B(4,4)$	0.94 (0.11)	$\Psi_w(4,4)$	0.66 (0.06)

The AML chi-square value for the goodness-of-fit of the whole model with restriction (49) is 45.78 with 6 degrees of freedom. The chi-square value for the null hypothesis (49) is equal to 44.54 with 4 degrees of freedom. Hence, as expected, the null hypotheses (49) is rejected at type I error level 0.05

Based on the similar approach given in Section 5, the analyses can

be extended to models with arbitrary and elliptical distributions, see, for example Lee & Poon (1992). To save space, details are not presented here.

Acknowledgement : This research was support in part by a Hong Kong UPGC research grant for the project "Analysis of Structural Equation Models with Non-Standard Data" and a research grant DA01070 from the National Institute on Drug Abuse.

7. REFERENCES

1 Aitken, M., & Longford, N. Statistical modelling issues in school effectiveness studies. J.R. Statist. Soc. A, 149, 1-43 (1986).

2 Anderson, T. W., & Rubin, H. Statistical inference in factor analysis. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 5, University of California Press, Berkeley, 111-150 (1956).

3 Bentler, P. M. Multivariate analysis with latent variables : Causal modeling. Annual Review of Psychology, 31, 419-456 (1980).

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

5 Bentler, P. M. EQS: Structural Equation Program Manual. Los Angles: BMDP Statistical Software (1989).

6 Bock, R. D. Multilevel Analysis of Educational Data. Academic Press (1989).

7 Browne, M. W. Generalized least squares estimators in the analysis of covariance structures. S. Afr. Statist. J., 8, 1-24 (1974).

8 Browne, M. W. Asymptotically distribution-free methods in the analysis of covariance structure. Br.J. Math. Stat. Psychol., 37, 62-83 (1984).

9 Ferguson, T. W. A method of generating best asymptotically normal estimates with applications to the estimation of bacterial densities. Ann. Statist., 29, 1046-1062 (1958).

10 Goldstein, H. Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika, 73, 43-56 (1986).

11 Goldstein, H., & McDonald, R. P. A general model for the analysis of multilevel data. Psychometrika, 53, 455-468 (1988).

12 Graybill, F. A. Matrices with applications in statistics, 2nd ed., Wadsworth (1983).

13 Green, P. J. Iteratively reweighted least squares for maximum likelihood estimation and some robust and resistant alternatives (with Discussion). J. R. Stat. Soc. B, 46, 149-192 (1984).

14 Jennrich, R. I., & Moore, R. A. Maximum likelihood estimation by means of nonlinear least squares. Statistical Computing Section, Proceedings of the American Statistical Association, 57-65 (1975).

15 Joreskog, K. G. Simultaneous factor analysis in several populations. Psychometrika, 57, 409-426 (1971).

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

17 Joreskog, K. G., & Sorbom, D. LISREL 7: A Guide to the Program and Applications. Chicago, IL: SPSS, Inc, (1988).

18 Lee, S.Y. Constrained estimation in covariance structure analysis. Biometrika, 66, 539-545 (1979).

19 Lee, S. Y. On testing functional constraints in structural equation models. Biometrika, 72, 125-131 (1985).

20 Lee, S. Y. Multilevel analysis of structural equation models. Biometrika, 77, 763-772 (1990).

21 Lee, S. Y. & Jennrich, R. I. A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. Psychometrika, 43, 99-113 (1979).

22 Lee, S. Y., & Poon W. Y. Two-level analysis of covariance structures for unbalanced designs with small level-one samples. British Journal of Mathematical and Statistical Psychology, in press (1992).

23 Lee, S. Y., & Poon W. Y. Analyses of two-level structural equation models with unbalanced designs. Technical report, Chinese University of Hong Kong (1991).

24 Lee, S. Y., & Tsui, K. L. Covariance structure analysis in several populations. Psychometrika, 47, 297-308 (1982).

25 Mardia, K. V. Measures of multivariate skewness and kurtosis with applications. Biometrika, 57, 519-530 (1970).

26 McDonald, R.P. & Goldstein, H. Balanced versus unbalanced designs for linear structural relations in two-level data. Br. J. Math. Statist. Psychol., 42, 215-232 (1989).

27 Muirhead, R. J. Aspects of Multivariate Statistical Theory, New York; Wiley (1982).

28 Nel, D. G. On matrix differentiation in statistics. S. Afr. Statist. J., 14, 137-193 (1980).

29 Rubin, D. B. Estimation in parallel randomized experiments. Journal of Educational Statistics, 6, 337-401 (1981).

30 Shapiro, A. Asymptotic theory of overparameterized structural models. J. Am. Statist. Assoc., 81, 142-149 (1986).

31 Shapiro, A. Robustness properties of the MDF analysis of moment structures. South African Statistical Journal, 21, 39-62 (1987).