《社會科學計量方法發展與應用》　楊文山主編
中央研究院中山人文社會科學研究所專書 (41)，頁139-183
民國86年9月，臺灣，臺北

Covariance Structure Analysis with Intraclass Dependent Observations*

Li-Jen Weng** P. M. Bentler***

I. Introduction

Independence among observations is one of the basic assumptions in covariance structure analysis. However, as was noted, for example, by Freedman (1985), some of the individuals in a conventional cluster sample may have known and interacted with each other. Under such circumstances, independence is unlikely. Violation of the independence assumption may bias the estimates. Moreover, no matter what distributional assumptions are made for the variables, all current statistical theory developed for covariance structure analysis requires the assumption of independence among observations. Once independence is questionable, the associated tests may become invalid.

* The authors would like to thank two anonymous reviewers for their helpful comments on an earlier version of this paper. The research was supported in part by USPHS grants DA 01070 and DA 00017.
** Department of Psychology, National Taiwan University
*** University of California, Los Angeles

In the present study, an attempt is made to approach the problem of dependent observations in covariance structure analysis. A dependence structure among all the observations is assumed. To simplify the problem, the same dependence structure is assumed for all the observed variables. Under the assumption of normality, the matrix normal distribution (Arnold, 1981: 313; de Waal, 1982; Nel, 1977) gives the joint distribution of the elements of a data matrix where both the variables and the observations can be correlated. It is assumed throughout this research that the data matrix follows a matrix normal distribution. The crucial problem for covariance structure analysis is to find a consistent estimator of the covariance matrix among the variables with the dependent effects among observations filtered.

The present research focuses on one case. The dependent structure studied has practical implications, especially for family, genetic, or classroom research. This dependence structure is block-diagonal intraclass, implying independence among groups. Members in each group are assumed to correlate equally with one another. Data collected from twins, couples, or siblings are examples. Many researchers have worked on the asymptotic properties of maximum likelihood estimators (MLE's) with dependent observations (e.g., Amemiya, 1985; Bar-Shalom, 1971; Bhat, 1974; Crowder, 1976; Heijmans & Magnus, 1986a, 1986b, 1986c; Weiss, 1971, 1973). However, the study of special cases is necessary because the published theorems are usually very general and are not guaranteed to be easily applicable to all cases.

Amemiya (1985) presented theorems on consistency and asymptotic normality of extremum estimators. By extremum estimators he means "estimators obtained by either maximizing or minimizing a certain function defined over the parameter space." Observations are not required to be independent or identically distributed. These theorems are applied to establish the consistency and the asymptotic normality of the MLE's obtained in this study by verifying the assumed conditions. Special attention is given to the dependence structure among observations because not all dependence structures satisfy the sufficient conditions set forth.

This paper is organized as follows. The notation and the matrix normal distribution are first introduced. The asymptotic properties of the maximum likelihood estimators are presented. A two-stage procedure is proposed to use the MLE of the covariance matrix obtained from dependent observations in covariance structure analysis. Finally, examples from simulated and real data are presented. The simulation studies investigate the effects of ignoring dependence among observations while the data are dependent. The real data example analyzes the factor structure among six personality scales based on 77 couples. The implications and the limitations of the model are discussed. Directions for future work are suggested.

II. Notation and the Matrix Normal Distribution

The following notation, unless indicated elsewhere, is used throughout the paper. The $n \times p$ data matrix X represents the observed values from a random sample of n observations on p variables. The vec operator stacks rows of a matrix into a long column. The symbol λ (X) represents an eigenvalue of matrix X. The $n \times p$ matrix μ represents the expected values of data matrix X. In other words, $E(x_{ij}) = \mu_{ij}$, $X = [x_{ij}]$ and $\mu = [\mu_{ij}]$. The symmetric positive definite $p \times p$ matrix Σ represents the covariance matrix between the columns of the data matrix

X (i.e., the variables). The symmetric positive definite matrix R ($n \times n$) with all the diagonal elements equal 1.0 represents the correlations between rows of the data matrix X (i.e., observations). Matrix R specifies the dependence structure among the observations.

A matrix X with a moment generating function

$$M_X(t) = \exp[\mathrm{tr}(\mu't) + 2^{-1}\mathrm{tr}(t'Rt\Sigma)],$$

where t is of the same order as X, is said to follow a matrix normal distribution with parameters μ, R, and Σ (Arnold, 1981). $X \sim N_{n,p}(\mu, R, \Sigma)$. The joint density function is

$$f(X) = (2\pi)^{-np/2}|R|^{-p/2}|\Sigma|^{-n/2}\exp\{-2^{-1}\mathrm{tr}[R^{-1}(X-\mu)\Sigma^{-1}(X-\mu)']\},$$

$-\infty < x_{ij} < \infty$, $X = [x_{ij}]$. The usual assumption of independent observations leads to a special case of the above density function with R = I, the identity matrix.

The matrix normal distribution implies that the covariance between any two data points depends not only on the covariation between the associated variables but also on the correlation between the observation units. In mathematical form,

$$\mathrm{cov}(x_{ij}, x_{\kappa t}) = \rho_{i\kappa}\sigma_{jt},$$

Were ρ_{ik} is an element of R and σ_{jt} is an element of Σ (Arnold, 1981: 311). Therefore, when observations are independent as assumed in most statistical methods, $\rho_{ik} = 0$ for all $i \neq k$, and data points between any two observation units yield zero covariance, i.e. $\mathrm{cov}(x_{ij}, x_{\kappa l}) = 0$ for all $i \neq k$. All the observations are identically distributed. The ordinary formula for computing the sample covariance works. However, if observations are dependent, i.e., $\rho_{ik} \neq 0$ for some $i \neq k$, computation of the sample covariance matrix using the ordinary formula and ignoring the dependence structure may yield an estimator which confounds covariation between variables with correlation between observations.

III. Asymptotic Properties of the Maximum Likelihood Estimators $\hat{\rho}$ and $\hat{\Sigma}$

The dependence structure studied is block-diagonal intraclass, implying independence among groups. Each group may consist of different number of observations. Members in each group are assumed to correlate equally with one another. Data collected from classrooms or siblings are examples. Donner and Koval (1980) estimated the intraclass correlation among siblings in univariate cases. They did not prove any asymptotic properties of the maximum likelihood estimators where the observations were dependent.

Let n_g be the number of observations in group g, G be the total number of groups, and n be the sum of all n_g's. The group size n_g is assumed to be fixed. Further, assume $\lim_{n \to \infty} G/n = \delta$, $0 < \delta < 1$. Let m_k be the number of groups of size k; in other words, $m_k = \{\text{number of } g : n_g = k\}$. Assume $\lim_{G \to \infty} \frac{m_k}{G} = f_k$, $0 \leq f_k \leq 1$ and $\sum_{k=1}^{K} f_k = 1$ for some finite positive integer K, $K = \max_g n_g < \infty$. Suppose an industrial psychologist is interested in studying the working attitudes of employees. The employees are organized into work groups of two or three. There are an equal number of these two types of work groups. In the study, there will be 100 samples of group size 2, and 100 samples of group size 3. Thus n = 500, G = 200, m_2 = 100, m_3 = 100, k = 2 or 3, K = 3, and in the limit $\delta = 0.4$, and $f_2 = f_3 = 0.5$.

For each positive integer n, X(n) is an n $\times$ p random matrix to be observed. Without loss of generality, X(n) is assumed to follow a matrix normal distribution with zero means, row correlation matrix $R(\rho)$, and column covariance matrix Σ ; X(n) ~ $N_{n,p}(0,\ R(\rho),\ \Sigma)$. Here $R(\rho)$, representing the dependence structure, is block-diagonal. Each $n_g \times n_g$ diagonal block in $R(\rho)$ has 1.0 on the diagonal and ρ elsewhere. Denote the vector containing the p^* distinct elements in the lower triangle of Σ_0 as σ_0. Let $\theta_0 = (\sigma_0, \rho_0)$ be the true parameter vector of order q; q = $p^* + 1$, with $p^* = p(p + 1)/2$. The total number of parameters q is independent of the sample size n. The domain of $\theta = (\sigma,\ \rho)$, or the parameter space Θ, consists of all possible values that θ can take. σ is the vector with the p^* distinct elements in the lower triangle of Σ. The parameter space Θ is a subset of q-dimensional Euclidean space; Θ = $\{(\sigma,\ \rho):\ \lambda_2 I_p < \ \Sigma < \lambda_1 I_p, -\varepsilon < \rho < \gamma,$ for $0 < \lambda_2 < \lambda_1$ and some $\varepsilon < K^{-1}$ and $\gamma < 1\}$. For matrices A and B of the same order, A > B in the Loewner sense of inequality, if and only if A — B is positive definite (see, e.g., Browne, 1974: 10; Browne & Shapiro, 1988: 207; Kano & Shapiro, 1987, Theorem 1). The definition of the parameter space implies that the eigenvalues of Σ are bounded between λ_2 and λ_1.

<u>Theorem</u>. The maximum likelihood estimators of Σ and ρ, $\hat{\theta} = (\hat{\sigma},\ \hat{\rho})$, under the conditions specified above are consistent and

asymptotically normal with plim $\hat{\theta} = \theta_0$, and $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N[0, A(\theta_0)^{-1}]$, where $A(\theta_0) = \lim E n^{-1}(\partial^2 L_n / \partial\theta\partial\theta')_{\theta_0}$

$$= \frac{1}{2}\begin{bmatrix} K_p^-(\Sigma_0^{-1} \otimes \Sigma_0^{-1})K_p^{-'} & \left[\delta \sum_{k>1}^{K} f_k \frac{k-1}{1+(k-1)\rho_0} - \frac{1-\delta}{1-\rho_0}\right] K_p^- \operatorname{Vec}\Sigma_0^{-1} \\ \text{Sym.} & \left[\delta \sum_{k>1}^{K} f_k \left(\frac{k-1}{1+(k-1)\rho_0}\right)^2 + \frac{1-\delta}{(1-\rho_0)^2}\right] p \end{bmatrix}.$$

The matrix K_p is of order $p^2 \times p^*$ with typical element $[K_p]_{ij,gh} = 2^{-1}(\delta_{ig}\delta_{jh} + \delta_{ih}\delta_{jg})$, $i \le p, j \le p, g \le h \le p$ and δ_{ij} represents Kronecker's delta. And $K_p^- = (K_p' K_p)^{-1} K_p'$ is a left inverse of K_p of order $p(p+1)/2 \times p^2$. See, for example, Browne (1974).

The asymptotic variance of $\sqrt{n}(\hat{\rho} - \rho_0)$, denoted as $\psi(\rho_0)$, is

$$2p^{-1}\delta^{-1}\left\{ \sum_{k>1}^{K} f_k \left[\frac{k-1}{1+(k-1)\rho_0}\right]^2 - \delta\left[\sum_{k>1}^{K} f_k \frac{k-1}{1+(k-1)\rho_0}\right]^2 \right.$$

$$\left. +2\frac{1-\delta}{1-\rho_0} \sum_{k>1}^{K} f_k \frac{k-1}{1+(k-1)\rho_0} + \frac{1-\delta}{(1-\rho_0)^2} \right\}^{-1}.$$

So, an estimate of the asymptotic variance of $\hat{p}$ is given by $\psi(\rho_0)/n$.

Amemiya's (1985) theorem 4.1.6 on extremum estimators is applied to prove the consistency and asymptotic normality of the MLE's of ρ and Σ. The verification of the required conditions is given in the Appendix for the sake of completeness.

The maximum likelihood estimator $\hat{\theta}$ is asymptotically efficient because its asymptotic variance-covariance matrix reaches the Cramer-Rao lower bound (see, e.g., Amemiya, 1985, Definition 4.2.1). The likelihood ratio test of $\rho_0 = 0$ is defined as $2(F_0 - F_1)$; where F_0 is the function value of the negative log-likelihood under the null hypothesis H_0 of no dependence and F_1 is the function value under the alternative hypothesis H_1 of non-zero ρ_0. This statistic is asymptotically distributed as a chi-square variate with 1 degree of freedom (see, e.g., Amemiya, 1985, Section 4.5.1). This test can be used to evaluate the appropriateness of the dependence assumption among observations.

IV. A Two-Stage Procedure for Covariance Structure Analysis

A Consistent estimator of the covariance matrix is one basic component in covariance structure analysis. $\hat{\Sigma}$ and $\hat{\rho}$ in the model have been shown to be consistent and asymptotically normal. Weng and Bentler (1987) discussed the use of $\hat{\Sigma}$ in covariance structure analysis when the data follow a matrix normal distribution and the dependence structure R is known. In the present research with an unknown parameter in the dependence structure, if the sample size is sufficiently large and one takes ρ at its estimated value $\hat{\rho}$, $\hat{\Sigma}$ can be used in any computer package for covariance structure analysis. Using $\hat{\Sigma}$ instead of S in the analysis yields appropriate estimates. S represents the usual sample covariance matrix estimate under independence.

A two-stage procedure is proposed for cases with sufficient sample sizes. $\hat{\Sigma}$ and $\hat{\rho}$ are estimated and evaluated at the first stage. At the second stage, covariance structure analysis is performed with an appropriately chosen input matrix. If the result of the likelihood ratio test indicates that ρ_0 is not significantly different from zero, observations can be treated as independent, and the usual analysis procedures using S follow. If ρ_0 is significantly different from zero, one can use $\hat{\Sigma}$ in analysis, while fixing ρ at $\hat{\rho}$. With ρ fixed at its estimate, the observed data can be transformed to independently and identically distributed vectors, while still preserving the original covariance matrix among the variables. Standard asymptotic statistical theories for covariance structure analysis are applicable.

Suppose X follows a matrix normal distribution, $N_{n,p}(0, R, \Sigma)$, with $R = R(\rho)$. Decompose R as AA', Where A is a square matrix of order n. The transformed matrix, $Y = A^{-1}X$, is distributed as $N_{n,p}(0, I, \Sigma)$. With $\hat{\rho}$ being consistent, we have $\hat{R} \xrightarrow{P} R$ and $\hat{A} \xrightarrow{P} A$, where $\hat{R} = R(\hat{\rho})$ and $\hat{A} = A(\hat{\rho})$. Asymptotically, $Y^* = \hat{A}^{-1}X$ has the same distribution as Y. Note that each row in Y is independently, identically distributed as N $(0, \Sigma)$. Since the transformed data have the same covariance matrix as the original data, analyses can be carried out by using the transformed independent data instead of the original dependent data. Existing programs for covariance structure analysis can be used. Statistical theories based on the i.i.d. assumption are applicable.

The two-stage procedure has several advantages. First, one can evaluate the degree of dependence among observations prior to any analysis. If the degree of dependence is negligible, the data can be treated as independent and the usual analysis procedure follows. Second, when dependence exists among observations, with ρ fixed at its estimate and with sufficient sample size, the transformation approach discussed above is legitimate. Calculation of the sample covariance matrix among observed variables is straightforward without the need for developing new estimator for it. Third, as long as $\hat{\Sigma}$ is obtained, any standard package program for covariance structure analysis can be used to get model parameter estimates and other statistics. However, when sample size is not large enough, the transformed data may not be independently, identically distributed because of the sampling error of $\hat{\rho}$. The sampling error of $\hat{\rho}$ may affect parameter estimates and associated standard errors, as well as the chi-square statistics.

Alternatively, one may simultaneously estimate the intraclass correlation and the model parameters. However, if the model is rejected with the one-stage estimation procedure, it is difficult to detect which part of the model breaks down. Moreover, the statistical properties of estimators under the one-stage procedure have to be proved otherwise. Statistical theory in covariance structure analysis under the independence assumption is not applicable. The two-stage procedure has the advantage of decomposing the problem into two components and examining them separately. As long as the transformation approach is appropriate, existing computer programs for covariance structure analysis can be used to obtain parameter estimates and available statistical theory is applicable.

Note that $\hat{\rho}$, the MLE of ρ, is not the only estimator that can be used in the two-stage procedure. Any consistent estimator is appropriate. One may be able to obtain estimators simpler than MLE.

V. Simulations and Example

This section presents the results from two simulation studies and one real data analysis. The purpose of the first simulation is to compare S and $\hat{\Sigma}$, where S represents the usual sample covariance matrix under independence and $\hat{\Sigma}$ represents the MLE of the population covariance matrix under the assumption of a matrix normal distribution with an intraclass dependence structure. The second simulation investigates the effect of using S and $\hat{\Sigma}$ in a tow-factor analytic model. The real data example looks into the factor structure of six personality measures on 77 couples.

(I) Simulation 1

Two cases of six variables are studied. The total sample size is fixed at 200, which is considered acceptable for six variables. In Case I, m=2 and G=100. In Case II, m=10 and G=20,. One hundred replications were performed for each case. The population intraclass correlation (ρ_0) ranges from 0.0 to 0.9. The population covariance matrix (Σ_0) is:

$$\Sigma_0 = \begin{bmatrix} 9.000 & 4.410 & 4.410 & 1.323 & 1.323 & 1.323 \\ 4.410 & 9.000 & 4.410 & 1.323 & 1.323 & 1.323 \\ 4.410 & 4.410 & 9.000 & 1.323 & 1.323 & 1.323 \\ 1.323 & 1.323 & 1.323 & 9.000 & 4.410 & 4.410 \\ 1.323 & 1.323 & 1.323 & 4.410 & 9.000 & 4.410 \\ 1.323 & 1.323 & 1.323 & 4.410 & 4.410 & 9.000 \end{bmatrix}.$$

Two estimators of the covariance matrix, S and $\hat{\Sigma}$, are obtained. S = X'X/(n - 1). $\hat{\Sigma}$ is obtained from the Newton-Raphson optimization procedure (see, e.g., Fox, 1971, Section 2.10). The starting values are set as $\hat{\Sigma}$ = S and $\hat{\rho} = 0$. The iteration stops if the sum of the absolute values of the parameter changes is no greater than 0.001.

Table 1 RMSE[a] of s and $\hat{\sigma}$

ρ_0		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
m	I										
2	S	0.743	0.736	0.728	0.734	0.794	0.835	0.858	0.890	0.950	0.970
2	$\hat{\Sigma}$	0.740	0.731	0.715	0.706	0.748	0.736	0.739	0.749	0.747	0.749
10	S	0.713	0.745	0.853	0.971	1.127	1.268	1.525	1.686	1.845	2.103
10	$\hat{\Sigma}$	0.710	0.722	0.702	0.767	0.772	0.775	0.820	0.844	0.910	0.933

[a]Average RMSE(s) $= \frac{1}{100}\sum_{i=1}^{100}\sqrt{\sum_{j=1}^{p}\sum_{k=1}^{j}(s_{ijk}-\alpha_{0ijk})^2/p^*}$, and

Average RMSE($\hat{\sigma}$) $= \frac{1}{100}\sum_{i=1}^{100}\sqrt{\sum_{j=1}^{p}\sum_{k=1}^{j}(\hat{\sigma}_{ijk}-\sigma_{0ijk})^2/p^*}$.

m = Number of observations in each group

I = Input matrix used in the analysis

Table 1 presents the average root-mean-squared-error (RMSE) over 100 replications for S and $\hat{\Sigma}$. The average RMSE's for $\hat{\Sigma}$ are small and the average RMSE of S is, in general, greater than that of $\hat{\Sigma}$ for any given intraclass correlation. All the average RMSE's fall between 0.7 and 1.0 except for those of S in Case II. In Case I, the Difference between the average RMSE's for S and $\hat{\Sigma}$ is less then 0.1 for $0.0 < \rho_0 < 0.4$, and the difference gradually increases as ρ_0 increases. In Case II, where m equals 10, the average RMSE's of S are much greater than those of $\hat{\Sigma}$. The results suggest that, when the total sample size is fixed, the difference between S and $\hat{\Sigma}$ increases (1) as the degree of dependence represented by the intraclass correlation increases, and (2) as the number of observations in each group increases or equivalently as the number of groups decreases.

The RMSE and bias of $\hat{\rho}$ are presented in Table 2. The MLE of the intraclass correlation under a matrix normal distribution performs quite well with small RMSE and bias.

Table 2 RMSE[a] and Bias[b] of $\hat{\rho}$

$m\backslash\rho_0$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					RMSE					
2	0.038	0.044	0.040	0.038	0.036	0.028	0.029	0.023	0.014	0.009
10	0.014	0.023	0.032	0.034	0.033	0.034	0.040	0.034	0.024	0.013
					BIAS					
2	0.000	0.002	0.000	-.004	-.005	-.002	0.004	0.000	0.003	-.001
10	0.001	-.004	0.008	-.003	-.002	-.003	-.003	-.005	-.005	0.000

[a]$\mathrm{RMSE}(\hat{\rho})=\sqrt{\sum_{i=1}^{100}(\hat{\rho}_i-\rho_0)^2/100}$.

[b]$\mathrm{Bias}(\hat{\rho}) = \mathrm{E}(\hat{\rho}) - \rho_0 = \sum_{i=1}^{100}\hat{\rho}_i/100-\rho_0$.

m = Number of observations in each group

The likelihood ratio test (LRT) of the hypothesis: $H_0:\rho_0 = 0$ versus $H_1:\rho_0 \neq 0$ is performed in each sample. The likelihood ratio test when

H_0 is true has an asymptotic chi-square distribution with 1 degree of freedom. The probability for Type I error is chosen at the $\alpha = 0.5$ level. The average test values and the empirical power or probability of Type I error (for $\rho_0 = 0.0$) of the test during the 100 replications are presented in Table 3. The empirical power of the test is the frequency with which H_0 was correctly rejected over the total number of replications. The mean of the chi-square statistic is close to 1.0 at $\rho_0 = 0$, but far from 1.0 for $\rho_0 > 0$. The test appears to be very powerful. In Case I with m = 2, the empirical power of the test reaches 1.00 for $\rho_0 \geq 0.2$. In Case II with m = 10, the empirical power of the LRT reaches 1.00 for $\rho_0 \geq 0.1$. A power analysis is conducted to study the theoretical power of the test. The series of hypotheses tested are $H_0 : \rho_0 = 0.0$ versus $H_1 : \rho_0 = 0.1$, up to 0.9. For the sake of simplicity, the power analysis is based on the asymptotic normal distribution of $\hat{\rho}$. Power = Prob $(\rho^2 / (\psi(\rho_0) / n) > 3.841 \mid H_1$ is true). $\rho^2 / (\psi(\rho_0) / n)$ under H_1 is asymptotically distributed as a non-central chi-square variate with 1 degree of freedom and non-centrality parameter $\rho_0^2 / (\psi(\rho_0) / n)$, $\rho_0 = 0.1$ to 0.9 (see, e.g., Hogg & Craig, 1978, Section 8.4). The middle of Table 3 gives the theoretical power of the test for population intraclass correlation ranging from 0.1 to 0.9. The results of the power analysis support the high frequency of rejection in the simulation.

Smaller sample sizes should decrease the power of the test. Another simulation is conducted to look into the empirical power of the test with smaller samples in comparison with theoretical power. The same setup is used except for the sample size. In this case, the sample size reduces to 100 with m = 2 and G =50. The results are given in the last two rows of Table 3. The power is reduced as a result of a decrease in sample size. But, the power of the test is still very high and reaches 1.00 for $\rho_0 \geq 0.3$.

Table 3 Mean of the Chi-Square Statistic and Power[a] of the Test

$m\backslash\rho_0$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(n = 200)										
Mean of the Test Statistic										
2	0.86	7.91	24.83	54.49	99.99	167.59	266.35	396.56	609.82	977.42
10	0.96	33.64	118.35	211.47	338.92	497.64	701.55	961.46	1347.85	2074.14
Empirical Power of the Test										
2	0.02	0.69	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Theoretical Power of the Test										
2	0.05	0.70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10	0.05	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
(n = 100)										
Empirical Power of the Test										
2	0.07	0.46	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Theoretical Power of the Test										
2	0.05	0.42	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00

[a]For ρ_0=0.0 the number represents the probability for Type I error rather than the power of the test.

m = Number of observations in each group

n = Total sample size

(II) Simulation 2

The purpose of the second simulation is to compare the results of a confirmatory factor analysis using either S or $\hat{\Sigma}$ as the input matrix, while the data are matrix normally distributed with intraclass dependence structure. The analysis based on $\hat{\Sigma}$ corresponds to the proposed two-stage methodology. In this model $\Sigma_0 = \Lambda_0 \Phi_0 \Lambda_0^{'} + \Psi_0$. The two-factor model includes six variables with three indicators for each factor. The population factor loading matrix (Λ_0), the factor covariance matrix (Φ_0), and the residual matrix (Ψ_0) are as follows.

$$\Lambda_0 = \begin{bmatrix} 2.1 & 0 \\ 2.1 & 0 \\ 2.1 & 0 \\ 0 & 2.1 \\ 0 & 2.1 \\ 0 & 2.1 \end{bmatrix}$$

$$\Phi_0 = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}$$

$$\Psi_0 = \begin{bmatrix} 4.59 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4.59 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4.59 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4.59 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4.59 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4.59 \end{bmatrix}.$$

The two cases studied are the same as in Simulation 1. The sample size of 200 is considered sufficient for the two-factor model studied. Therefore, the transformation approach and the proposed two-stage methodology may be applied. One hundred replications were performed. The parameters of the factor model were estimated using the MLE option in EQS (Bentler, 1989).

Table 4 Mean and Standard Deviation[a] of $\hat{\lambda}$ and $\hat{\phi}$

P	m	I \ ρ_0	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
λ_{11}	2	S	2.0744	2.1479	2.0762	2.1204	2.1218	2.1007	2.1006	2.1230	2.1141	2.0330
			(.245)	(.239)	(.207)	(.207)	(.265)	(.257)	(.248)	(.280)	(.321)	(.305)
		$\hat{\Sigma}$	2.0689	2.1415	2.0719	2.1175	2.1100	2.0998	2.0813	2.1026	2.1179	2.0477
			(.244)	(.235)	(.204)	(.201)	(.243)	(.232)	(.220)	(.238)	(.255)	(.243)
	10	S	2.0975	2.1264	2.1194	2.0398	2.0841	2.1895	2.1013	2.0956	2.2128	2.0695
			(.215)	(.237)	(.309)	(.304)	(.343)	(.428)	(.425)	(.546)	(.634)	(.705)
		$\hat{\Sigma}$	2.0917	2.1064	2.1174	2.0498	2.0673	2.1233	2.0726	2.0733	2.1435	2.0812
			(.216)	(.224)	(.262)	(.214)	(.237)	(.242)	(.225)	(.216)	(.240)	(.245)
λ_{21}	2	S	2.1551	2.0528	2.0939	2.0814	2.1126	2.0808	2.0826	2.0502	2.1474	2.0941
			(.212)	(.239)	(.242)	(.258)	(.276)	(.231)	(.278)	(.307)	(.274)	(.318)
		$\hat{\Sigma}$	2.1502	2.0515	2.0941	2.0807	2.0996	2.0684	2.0659	2.0321	2.1268	2.0988
			(.212)	(.239)	(.233)	(.255)	(.264)	(.219)	(.234)	(.246)	(.232)	(.229)
	10	S	2.0775	2.1194	2.1685	2.1283	2.1258	2.0398	2.0434	2.0803	2.0559	2.1565
			(.236)	(.201)	(.313)	(.312)	(.405)	(.446)	(.471)	(.598)	(.694)	(.672)
		$\hat{\Sigma}$	2.0708	2.1064	2.1230	2.1125	2.1164	2.0445	2.0310	2.0514	2.1249	2.1154
			(.236)	(.198)	(.257)	(.224)	(.271)	(.231)	(.213)	(.239)	(.237)	(.264)
λ_{31}	2	S	2.0832	2.0666	2.1014	2.0625	2.1127	2.0928	2.1308	2.0873	2.1232	2.1492
			(.228)	(.242)	(.229)	(.267)	(.260)	(.245)	(.285)	(.285)	(.292)	(.310)
		$\hat{\Sigma}$	2.0767	2.0609	2.0916	2.0672	2.1193	2.0802	2.1135	2.0790	2.1211	2.1389
			(.226)	(.241)	(.223)	(.246)	(.225)	(.236)	(.246)	(.240)	(.239)	(.237)

Table 4 (continued)

P	m	I	ρ_0: 0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	10	S	2.1174	2.1109	2.0988	2.1102	2.1098	2.1314	2.0979	2.1218	2.0499	2.1340
			(.202)	(.226)	(.254)	(.324)	(.382)	(.412)	(.416)	(.527)	(.622)	(.705)
		$\hat{\Sigma}$	2.1124	2.1062	2.0984	2.1079	2.0880	2.1320	2.0930	2.1199	2.1236	2.0897
			(.201)	(.234)	(.208)	(.232)	(.235)	(.222)	(.258)	(.255)	(.278)	(.270)
λ_{42}	2	S	2.1056	2.1110	2.1182	2.0959	2.1151	2.1337	2.1184	2.0941	2.0939	2.1040
			(.236)	(.240)	(.232)	(.222)	(.241)	(.231)	(.255)	(.284)	(.296)	(.323)
		$\hat{\Sigma}$	2.1009	2.1044	2.1102	2.0921	2.1143	2.1280	2.1052	2.1084	2.0984	2.1072
			(.236)	(.237)	(.229)	(.211)	(.219)	(.223)	(.221)	(.240)	(.238)	(.232)
	10	S	2.1127	2.1259	2.0940	2.0684	2.0974	2.0808	2.1247	2.0797	2.0745	2.1511
			(.239)	(.238)	(.248)	(.260)	(.346)	(.377)	(.491)	(.549)	(.581)	(.764)
		$\hat{\Sigma}$	2.1077	2.1184	2.1046	2.0803	2.0966	2.0652	2.0829	2.1090	2.1567	2.0861
			(.238)	(.235)	(.206)	(.207)	(.216)	(.226)	(.286)	(.263)	(.241)	(.272)
λ_{52}	2	S	2.0726	2.1153	2.1014	2.0556	2.1043	2.1668	2.0893	2.0912	2.0815	2.1110
			(.208)	(.225)	(.233)	(.227)	(.241)	(.242)	(.276)	(.288)	(.254)	(.352)
		$\hat{\Sigma}$	2.0667	2.1114	2.0874	2.0501	2.1145	2.1624	2.0855	2.0970	2.0945	2.1290
			(.209)	(.223)	(.238)	(.223)	(.239)	(.227)	(.216)	(.208)	(.232)	(.256)
	10	S	2.1406	2.1123	2.0715	2.1383	2.1253	2.1640	2.1106	2.0342	2.1245	2.0540
			(.209)	(.251)	(.296)	(.307)	(.383)	(.482)	(.481)	(.477)	(.635)	(.680)
		$\hat{\Sigma}$	2.1343	2.1084	2.0678	2.1182	2.1211	2.1236	2.1091	2.1043	2.1169	2.0942
			(.210)	(.230)	(.230)	(.234)	(.229)	(.228)	(.260)	(.210)	(.273)	(.269)

Table 4 (continued)

P	m	I (ρ_0)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
λ_{62}	2	S	2.0918	2.1136	2.0935	2.1138	2.0901	2.1150	2.0592	2.0408	2.1189	2.1925
			(.247)	(.228)	(.232)	(.259)	(.249)	(.254)	(.237)	(.269)	(.339)	(.316)
		$\hat{\Sigma}$	2.0863	2.1044	2.0904	2.1175	2.0962	2.0910	2.0798	2.0524	2.1150	2.1543
			(.246)	(.227)	(.225)	(.244)	(.230)	(.234)	(.202)	(.231)	(.246)	(.255)
	10	S	2.1088	2.1148	2.0922	2.1241	2.1424	2.1256	2.1066	2.1020	2.0848	2.0693
			(.234)	(.237)	(.265)	(.295)	(.343)	(.410)	(.513)	(.484)	(.636)	(.761)
		$\hat{\Sigma}$	2.1017	2.1032	2.1041	2.1354	2.1345	2.0925	2.0703	2.0688	2.1226	2.0916
			(.235)	(.219)	(.224)	(.237)	(.203)	(.238)	(.227)	(.220)	(.263)	(.280)
ϕ_{21}	2	S	0.3062	0.2847	0.3067	0.3027	0.3026	0.2912	0.2977	0.2765	0.2969	0.3203
			(.083)	(.099)	(.095)	(.094)	(.093)	(.098)	(.119)	(.108)	(.116)	(.113)
		$\hat{\Sigma}$	0.3072	0.2861	0.3076	0.3033	0.3004	0.2919	0.2965	0.2720	0.2938	0.3118
			(.083)	(.097)	(.094)	(.092)	(.086)	(.085)	(.106)	(.096)	(.083)	(.090)
	10	S	0.2925	0.3139	0.2882	0.2979	0.3145	0.2864	0.2929	0.2828	0.3195	0.3124
			(.083)	(.088)	(.103)	(.120)	(.138)	(.187)	(.187)	(.229)	(.259)	(.330)
		$\hat{\Sigma}$	0.2926	0.3137	0.2905	0.3092	0.3037	0.3073	0.2983	0.3031	0.3078	0.3217
			(.083)	(.086)	(.088)	(.092)	(.083)	(.096)	(.097)	(.096)	(.096)	(.080)

[a]Standard deviation is in parentheses.

P = Parameter

m = Number of observations in each group

I = Input Matrix used in the analysis

Means and standard deviations of the six factor loading estimates and the estimated factor covariance over the 100 replications are given in Table 4. All the means of the parameter estimates are very close to the true values regardless of the population intraclass correlation, the number of observations in each group, or the input covariance matrix used in the analysis. As $\rho_0 = 0$, the factor loading estimates from S are always greater than those from $\hat{\Sigma}$ by approximately 0.005. This results from different denominators used for S and $\hat{\Sigma}$: (n-1) for S and n for $\hat{\Sigma}$. Therefore, S equals approximately (200/99)$\hat{\Sigma}$. Factor loadings estimates from S are greater than factor loadings from $\hat{\Sigma}$ approximately by a multiplier of $(200/199)^{1/2}$.

When we look at standard deviations of the parameter estimates, discrepancies among cases appear. The standard deviations of all factor loading estimates show similar patterns. The standard deviations of parameter estimates when S is used are usually larger than those with $\hat{\Sigma}$ being used, except for very small ρ_0, say under 0.4 for Case I and under 0.2 for Case II. This result seems to indicate that the parameter estimates using S have less empirical efficiency than using $\hat{\Sigma}$. When $\hat{\Sigma}$ is used, the standard deviations of the parameter estimates have small differences between the two cases. When S is used, the standard deviations in Case II are larger than those in Case I, and the difference increases as ρ_0 increases. The results seem to indicate that although the mean parameter estimates are close regardless of group size or the input matrix used, the use of S leads to estimates with greater dispersion, especially for data with large groups and high degrees of dependence, when the total sample size is held constant.

Table 5 Frequency of Model Rejection

ρ_0 / m I	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2 S	3	6	6	7	10	9	13	23	32	35
2 $\hat{\Sigma}$	3	6	6	4	5	2	3	7	7	3
10 S	6	12	16	43	57	75	86	94	98	98
10 $\hat{\Sigma}$	6	9	7	7	7	5	4	4	8	6

m = Number of observations in each group

I = Input matrix used in the analysis

Table 6 Mean and Standard Deviation[a] of the Chi-Square Statistic: $\chi^2(8)$

ρ_0 / m	I	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	S	8.526 (4.2)	8.757 (4.2)	8.224 (4.3)	8.267 (4.1)	9.938 (4.5)	8.928 (4.4)	10.741 (4.6)	12.268 (5.6)	13.080 (6.4)	13.743 (6.5)
	$\hat{\Sigma}$	8.520 (4.2)	8.648 (4.0)	7.939 (4.3)	7.687 (3.7)	8.311 (4.1)	7.314 (3.4)	7.550 (3.2)	7.914 (3.9)	8.260 (4.5)	7.683 (3.3)
10	S	7.695 (3.8)	9.239 (5.2)	10.708 (4.7)	14.812 (7.3)	18.339 (9.1)	25.201 (11.4)	34.415 (17.7)	42.673 (20.0)	55.675 (25.7)	71.865 (31.8)
	$\hat{\Sigma}$	7.734 (3.8)	8.660 (4.6)	8.154 (3.7)	8.659 (4.2)	8.002 (4.4)	7.841 (3.9)	7.938 (4.1)	7.861 (4.5)	8.375 (4.2)	8.225 (4.6)

[a]Standard deviation is in parentheses.

m = Number of observations in each group

I = Input matrix used in the analysis

The frequency with which the two-factor model was rejected and the mean and standard deviation of the chi-square test statistic are summarized in Table 5 and Table 6. The expected frequency of rejection is 5 with α = .05. The test associated with the model has 8 degrees of freedom. The mean and standard deviation of the statistic are expected to be 8 and 4, respectively.

The model tends to be rejected too often when S is used as the input matrix, especially in Case II. When $\hat{\Sigma}$ is used, the frequency of model rejection ranges from 2 to 9, and the chi-square test statistic behaves very well with a mean close to 8 and a standard deviation close to 4 in both cases. In Case I with S as the input matrix, the test performs fine for $\rho_0 < 0.4$, while its mean and standard deviation begin to increase as $\rho_0 \geq 0.4$. In Case II using S, the test did not perform well even with $\rho_0 = 0.1$, while the mean and standard deviation of the test statistic increases dramatically as ρ_0 increases. The results indicate that the test statistic is asymptotically distributed as a chi-square variate with 8 degrees of freedom if $\hat{\Sigma}$ is used as the input matrix, but it does not have an asymptotic chi-square distribution if S is used in the analysis.

(III) An Example

The sample consisted of 77 couples. The Bentler Psychological Inventory (BPI) (Comrey, Backer, & Glaser, 1973), which assesses 28 personality traits, was administered to the couples. An interested reader may refer to Bentler and Newcomb (1978) for detailed description of the sample, the data collection procedures, and the BPI. For demonstration purpose, six personality traits were selected based on husband-wife correlations on the scales (Bentler & Newcomb, 1978) and the factor structure among the personality traits (see, e.g. Stein, Newcomb, &

Bentler, 1987). Deliberateness, Diligence, Orderliness, Law Abidance, Liberalism, and Religious Commitment were chosen. The first three reflect the factor of Conscientiousness and the rest reflect the factor of Social Conformity.

A confirmatory two-factor analytic model with three indicators on each factor is employed. Separate analyses were first conducted on husband and wife samples to ensure the appropriateness of pooling together the data. The fits in both samples are acceptable. For husbands, $\chi^2(8)$ = 10.247, p=.248, the factor loading of liberalism on Social Conformity was not significant. For wives, $\chi^2(8)$ = 13.953, p=.083, and all the parameter estimates were significant. The data on husbands and wives were pooled together to form the couple sample of size 154. The husband and wife in each couple are assumed to be dependent. The sample size of 154 is considered at least moderate for the transformation approach and the two-stage methodology to be used.

The model using S as the input matrix for the couple data cannot be rejected; $\chi^2(8)$ = 12.079, p=.148 and all the parameter estimates are significant. Means were subtracted from the original data prior to the estimation of $\hat{\Sigma}$. The Newton-Raphson iteration method was used to obtain $\hat{\Sigma}$ and $\hat{\rho}$. $\hat{\rho}$ = 0.193 and the LRT, asymptotically distributed as a chi-square variate with 1 degree of freedom, yields 22.056, rejecting the hypothesis of no dependence among husbands and wives on the six personality traits. The two-factor model, using $\hat{\Sigma}$ as the input matrix, cannot be rejected, either; $\chi^2(8)$ = 12.039, p=.149, with all the parameter estimates being significant. The parameter estimates and their associated standard errors are summarized in Table 7. The high similarity in the results coincides with the simulation results with $\hat{\rho}$ being close to

0.2, since when $\rho_0 = 0.2$, the results for S and $\hat{\Sigma}$ in both simulation studies are essentially the same for Case I with m = 2.

Table 7 Parameter Estimates and Associated Standard Errors for the Couples Example

Input Matrix Used	S		$\hat{\Sigma}$	
Parameter (θ)	$\hat{\theta}$	S.E.	$\hat{\theta}$	S.E.
λ_{11} (DB)	2.577	(.656)	2.875	(.676)
λ_{21} (DG)	4.322	(.791)	4.201	(.769)
λ_{31} (OD)	4.029	(.772)	4.185	(.767)
λ_{42} (LA)	4.087	(.678)	4.177	(.696)
λ_{52} (LB)	-2.041	(.632)	-1.799	(.610)
λ_{62} (RC)	3.878	(.785)	3.647	(.755)
ϕ_{21}	0.511	(.122)	0.528	(.122)
ψ_{11} (DB)	33.463	(4.436)	34.936	(4.756)
ψ_{22} (DG)	30.806	(6.293)	32.893	(6.053)
ψ_{33} (OD)	33.245	(5.967)	32.849	(6.023)
ψ_{44} (LA)	11.453	(4.873)	10.682	(5.141)
ψ_{55} (LB)	36.945	(4.500)	33.589	(4.269)
ψ_{66} (RC)	38.627	(6.198)	36.466	(5.718)

DB = Deliberateness

DG = Diligence

OD = Orderliness

LA = Law Abidance

LB = Liberalism

RC = Religious Commitment

VI. Discussion

The presented model has practical applications in many areas of research. For example, an industrial psychologist may observe similarity in working attitudes among employees under the same supervisor. This similarity can be either the cause or the result of being assigned to the same supervisor. Suppose the psychologist is interested in the relationships among various working attitudes and intends to conduct a factor analysis on these attitude scales. One immediate difficulty encountered is the dependence existing among the employees under the same supervisors. The current research offers one solution to this problem.

The model is limited in some aspects. The limitations provide directions for future research. First, the matrix normal distribution assumes one dependence structure for all the variables. This constraint is unacceptable in certain research. Second, the degree of dependence among observations within groups are assumed to be identical. This restriction can be relaxed to allow for different intraclass correlations. In addition, other dependence structures can be studied for different research designs. Stadje (1984) pointed out that not every dependence structure yields consistent estimators. One has to prove the properties of the estimators under different dependence structures.

Transformation is a useful approach for certain classes of dependence structures. Suppose the dependence structure can be decomposed as follows: $R = KLK'$, with K known and L unknown diagonal. The intraclass dependence studied is a special case of this general model. If the eigenvectors of the dependence structure do not depend on any parameters, the data can be transformed to independent, but not identically distributed vectors. The distributions of the vectors depend on L and Σ. The proofs involving independent, but not identically distributed observations, are usually simpler than those involving dependent observations. Extra conditions are necessary for the parameter estimates to have desirable statistical properties.

Another line of research involves dependent but non-normally distributed data. The matrix elliptical distribution is one possibility. The transformation approach should give uncorrelated and not identically distributed vectors. Whether the transformed vectors are independent depends on the nature of the matrix-valued distributions. More work is needed.

References

Amemiya, T.

1985 *Advanced Econometrics.* Cambridge, MA: Harvard University Press.

Arnold, S. F.

1981 *The Theory of Linear Models and Multivariate Analysis.* New York: Wiley.

Bar-Shalom, Y.

1971 "On the Asymptotic Properties of the Maximum-likelihood Estimate Obtained from Dependent Observations," *Journal of the Royal Statistical Society Series B,* 33: 72-77.

Bentler, P. M.

1989 *EQS Structural Equations Program Manual.* Los Angeles: BMDP Statistical Software, Inc.

Bentler, P. M., & M. D. Newcomb

1978 "Longitudinal Study of Marital Success and Failure," *Journal of Consulting and Clinical Psychology* 46: 1053-1070.

Bhat, B. R.

1974 "On the Method of Maximum-Likelihood for Dependent Observations," *Journal of the Royal Statistical Society Series B,* 36: 48-53.

Bradley, R. A., & J. J. Gart

1962 "The Asymptotic Properties of ML Estimators When Sampling from Associated Populations," *Biometrika* 49: 205-214.

Browne, M. W.

1974 "Generalized Least Squares Estimators in the Analysis of Covariance Structures," *South African Statistical Journal* 8: 1-24.

Comrey, A. L., T. E. Backer & E. M. Glaser

1973 *A Sourcebook for Mental Health Measures.* Los Angeles: Human Interaction Research Institute.

Crowder, M.J.

1976 "Maximum Likelihood Estimation for Dependent Observations," *Journal of the Royal Statistical Society Series B,* 38: 45-53.

de Waal, D. J.

1982 "Matrix-Valued Distributions," in S. Kotz & N. Johnson (eds.), *Encyclopedia of Statistical Sciences,* 5: 326-333. New York: Wiley.

Donner, A. & J. J. Koval

1980 "The Estimation of Intraclass Correlation in the Analysis of Family Data," *Biometrics* 36: 19-25.

Fox, R. L.

1971 *Optimization Methods for Engineering Design.* Reading, MA: Addison-Wesley.

Freedman, D. A.

1985 "Statistics and the Scientific Method," in W. M. Mason & S. E.

Fienberg (eds.), *Cohort Analysis in Social Research*, pp.343-366. New York: Springer-Verlag.

Graybill, F. A.

1983 *Matrices with Applications in Statistics, 2nd Ed.* Belmont, CA: Wadsworth.

Heijmans, R. D. H., & J. R. Magnus

1986a "On the First-order Efficiency and Asymptotic Normality of Maximum Likelihood Estimators Obtained from Dependent Observations," *Statistic Neerlandica* 40: 169-188.

1986b "Consistent Maximum-Likelihood Estimation with Dependent Observations: The General (Non-normal) Case and the Normal Case," *Journal of Econometrics* 32: 253-285. Corrigenda in: *Journal of Econometrics* 35: 395.

1986c "Asymptotic Normality of Maximum Likelihood Estimators Obtained from Normally Distributed but Dependent Observations," *Econometric Theory* 2: 374-412.

Hoadley, B.

1971 "Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case," *The Annals of Mathematical Statistics* 42: 1977-1991.

Hogg, R. V., & A. T. Craig

1978 *Introduction to Mathematical Statistics, 4th Ed.* New York: Macmillan.

Nel, H. M.

1977 "On Distributions and Moments Associated with Matrix Normal Distribution," *Technical Report No.24,* Department of Mathematical Statistics, University of the Orange Free State, Bloemfontein, South Africa.

Philippou, A. N., & G. G. Roussas

1973 "Asymptotic Normality of the Maximum Likelihood Estimate in the Independent Not Identically Distributed Case," *Annals of the Institute of Statistical Mathematics* 27: 45-55.

Stadje, W.

1984 "A Note on Sample Means and Variances of Dependent Normal Variables," *Mathematische Operationsforschung und Statistik, series statistics* 15: 205-218.

Stein, J. A., M. D. Newcomb & P. M. Bentler

1987 "Personality and Drug Use: Reciprocal Effects across Four Years," *Personality and Individual Differences* 8: 419-430.

Weiss, L.

1971 "Asymptotic Properties of Maximum Likelihood Estimators in Some Nonstandard Cases," *Journal of the American Statistical Association* 66: 345-350.

1973 "Asymptotic Properties of Maximum Likelihood Estimators in Some Nonstandard Cases, II," *Journal of the American Statistical Association* 68: 428-430.

Weng, J. L.

1990 "Aspects of Covariance Structure Analysis with Dependent Observations," Ph. D. Dissertation, University of California, Los Angeles (Unpublished).

Weng, L.-J., & P. M. Bentler

1987 "Linear Structural Equation Modeling with Dependent Observations," *Proceedings of the Social Statistics Section, American Statistical Association:* 498-500.

Appendix

Theorem 4.1.6 of Amemiya (1985) is used to prove the consistency and the asymptotic normality of the maximum likelihood estimators $\hat{\theta}$. Let $L_n(\theta)$ denote the function to be minimized over the parameter space Θ. The following conditions have to be met.

(A) $\partial^2 L_n / \partial\theta\partial\theta'$ exists and is continuous in an open, convex neighborhood of θ_0.

(B) $n^{-1}(\partial^2 L_n / \partial\theta\partial\theta')_{\theta_n^*}$ converges to a finite positive definite matrix $A(\theta_0) = \lim En^{-1}(\partial^2 L_n / \partial\theta\partial\theta')_{\theta_0}$ in probability for any sequence θ_n^* such that plim $\theta_n^* = \theta_0$.

(C) $n^{-1/2}(\partial L_n / \partial\theta)_{\theta_0} \rightarrow N[0, B(\theta_0)]$, where $B(\theta_0) = \lim \ En^{-1} (\partial L_n / \partial\theta)_{\theta_0} \times (\partial L_n / \partial\theta')_{\theta_0}$.

(D) $n^{-1}L_n(\theta)$ converges to a nonstochastic function $L(\theta)$ in probability uniformly in θ in an open neighborhood of θ_0.

(E) plim $n^{-1}\partial^2 L_n / \partial\theta\partial\theta'$ exists and is continuous in a neighborhood of θ_0.

A transformation of the original data matrix X(n) is used to verify the stated conditions. Partition X(n) as $(X_1', X_2', X_3', \ldots, X_g', \ldots)'$, where X_g is the observed $n_g \times p$ data matrix for group g. The eigenvalues of each $n_g \times n_g$ diagonal block of $R(\rho)$ are $[1 + (n_g - 1)\rho]$ with multiplicity 1 and $(1 - \rho)$ with multiplicity $(n_g - 1)$. Tet T_g be a nonsingular $n_g \times n_g$ matrix with all the elements in the first column equal to $n_g^{-1/2}$ and the remaining columns being orthonormal without ρ and Σ involved. Define $Y_g = T_g' X_g$. Then, all the rows of Y_g are independent. The first row of each Y_g has a normal distribution with mean zero and covariance matrix $[1 + (n_g - 1)\rho]\Sigma$. The remaining rows also have a normal distribution with zero means and covariance matrix $(1 - \rho)\Sigma$. Denote $(Y_1', Y_2', Y_3', \ldots, Y_g', \ldots)'$ as Y(n). $Y'(n) = (y_1, y_2, y_3, \ldots, y_{n_1}, y_{n_1+1}, y_{n_1+2}, y_{n_1+3}, \ldots, y_{n_1+n_2}, y_{n_1+n_2+1}, y_{n_1+n_2+2}, \ldots, y_{n_1+n_2+n_3}, \ldots, y_{n_1+n_2+n_3+\cdots+n_g+\cdots}, \cdots)$. Each y_i is a column vector of order p.

All the rows of Y are independently, but non-identically distributed. Although there are theorems for consistemcy and asymptotic normality of the MLE from independently but not identically distributed observations (e.g., Bradley & Gart, 1962; Hoadley, 1971; Philippou & Roussas, 1973), Amemiya's theorems (1985) are applied to prove the consistency and asymptotic normality of the MLE's of ρ and Σ for the sake of simplicity.

Let us denote the y_i, whose distribution is $N(0, (1 + (n_g - 1)\rho)\Sigma)$, g = 1, 2, 3, ..., G, as u_g. Let the G $\times$ p matrix U represent the transpose of the collection of all the u_g's. Note that $(1 + (n_g - 1)\rho)^{-1/2} u_g \sim N(0, \Sigma)$ for every g. The remaining y_i's are independently, identically, normally distributed with zero means and covariance matrix $(1 - \rho)\Sigma$. Denote the transpose of the collection of all these y_i's as V, an (n - G) $\times$ p matrix. The i.i.d. property plays an important role in the demonstration of asymptotic properties of the MLE's, $\hat{\rho}$ and $\hat{\Sigma}$. The negative Log-likelihood for Y, or equivalently for U and V, is

$$L_n(Y, \theta) =$$

$$2^{-1}\left\{ np\ln(2\pi) + p \sum_{g=1}^{G} \ln[1 + (n_g - 1)\rho] + p(n - G)\ln(1 - \rho) + n \ln|\Sigma| \right.$$

$$\left. + \sum_{g=1}^{G} \frac{1}{1 + (n_g - 1)\rho} u_g' \Sigma^{-1} u_g + \frac{1}{1 - \rho} \mathrm{tr} \quad V \quad \Sigma^{-1} V' \right\}. \qquad (1)$$

The first and second derivatives of $L_n(\theta)$ are as follows. The derivative with respect to Σ refers only to the p^* distinct elements in the matrix. The product $A \otimes B$ represents the right Kronecker product with typical element $[a_{ij}B]$.

$$\frac{\partial L_n(\theta)}{\partial \sigma} = 2^{-1} K_p^{-} \left\{ n \ \mathrm{Vec}\ \Sigma^{-1} - \sum_{g=1}^{G} \frac{1}{1+(n_g-1)\rho} \mathrm{Vec}(\Sigma^{-1} u_g u_g' \Sigma^{-1}) - \frac{1}{1-\rho} \mathrm{Vec}(\Sigma^{-1} V' V \Sigma^{-1}) \right\} \quad (2)$$

$$\frac{\partial L_n(\theta)}{\partial \rho} = 2^{-1} \left\{ p \sum_{g=1}^{G} \frac{n_g - 1}{1+(n_g-1)\rho} - p(n-G)\frac{1}{1-\rho} - \sum_{g=1}^{G} \frac{n_g-1}{[1+(n_g-1)\rho]^2} u_g' \Sigma^{-1} u_g + \frac{1}{(1-\rho)^2} \mathrm{tr} V \Sigma^{-1} V' \right\} \quad (3)$$

$$\frac{\partial^2 L_n(\theta)}{\partial \sigma \partial \sigma'} = 2^{-1} K_p^{-} \left\{ -n(\Sigma^{-1} \otimes \Sigma^{-1}) + \sum_{g=1}^{G} \frac{1}{1+(n_g-1)\rho} (\Sigma^{-1} \otimes \Sigma^{-1} u_g u_g' \Sigma^{-1} + \Sigma^{-1} u_g u_g' \Sigma^{-1} \otimes \Sigma^{-1}) + \frac{1}{1-\rho} (\Sigma^{-1} \otimes \Sigma^{-1} V' V \Sigma^{-1} + \Sigma^{-1} V' V \Sigma^{-1} \otimes \Sigma^{-1}) \right\} K_p^{-'} \quad (4)$$

$$\frac{\partial^2 L_n(\theta)}{\partial\sigma\partial\rho} = 2^{-1} K_p^- \left\{ \sum_{g=1}^{G} \frac{n_g - 1}{[1+(n_g-1)\rho]^2} \mathrm{Vec}(\Sigma^{-1} u_g u_g' \Sigma^{-1}) \right.$$

$$\left. - \frac{1}{(1-\rho)^2} \mathrm{Vec}(\Sigma^{-1} V' V \Sigma^{-1}) \right\} \tag{5}$$

$$\frac{\partial^2 L_n(\theta)}{\partial\rho^2} = 2^{-1} \left\{ -p \sum_{g=1}^{G} \left[\frac{n_g - 1}{1+(n_g-1)\rho} \right]^2 - p(n-G)\frac{1}{(1-\rho)^2} \right.$$

$$\left. +2 \sum_{g=1}^{G} \frac{(n_g-1)^2}{[1+(n_g-1)\rho]^3} u_g' \Sigma^{-1} u_g + 2\frac{1}{(1-\rho)^3} \mathrm{tr} \; V \; \Sigma^{-1} \; V' \right\} \tag{6}$$

The definition of matrix K_p is given in the text.

Verification of Condition (A)

Condition (A) is met trivially.

Verification of Condition (B)

It is to be shown that for any sequence $\theta_n^* = (\rho_n^*, \sigma_n^*)$ such that plim $\theta_n^* = \theta_0$, $n^{-1}(\partial^2 L_n / \partial\theta\partial\theta')_{\theta_n^*}$ converges to a finite positive definite matrix $A(\theta_0) = \lim E n^{-1}(\partial^2 L_n / \partial\theta\partial\theta')_{\theta_0}$ in probability.

Recall that $(1+(n_g-1)\rho_0)^{-1/2}$ u_g's are i.i.d., so are the rows of V. Therefore, we have

$$\frac{1}{n-G} V'V \xrightarrow{p} (1-\rho_0)\Sigma_0 \, ;$$

$$E\left[\frac{1}{G}\sum_{g=1}^{G}\frac{1}{1+(n_g-1)\rho_0}u_g u_g'\right]=\Sigma_0 \text{ ; and}$$

$$E\left[\frac{1}{n-G}V'V\right]=(1-\rho_0)\Sigma_0.$$

It can be shown that under the assumption $\left|\rho_n^* - \rho_0\right| < \varepsilon^* < K^{-2}$,

$$\frac{1}{G}\sum_{g=1}^{G}\frac{1}{1+(n_g-1)\rho_n^*}u_g u_g' \xrightarrow{p} \Sigma_0 .$$

The results follow based on the above statements.

$$\text{plim } n^{-1}\left(\frac{\partial^2 L_n}{\partial\sigma\partial\sigma'}\right)_{\theta_n^*} = 2^{-1}K_p^-(\Sigma_0^{-1}\otimes\Sigma_0^{-1})K_p^{-\prime}$$

$$= \lim E\, n^{-1}\left(\frac{\partial^2 L_n}{\partial\sigma\partial\sigma'}\right)_{\theta_0}.$$

$$n^{-1}\left(\frac{\partial^2 L_n}{\partial\sigma\partial\rho}\right)_{\theta_n^*} = \frac{1}{2n}K_p^-\sum_{k>1}^{K}\frac{k-1}{[1+(k-1)\rho_n^*]^2}\text{Vec}\left[\Sigma_n^{*-1}\left(\sum_{\{g:n_g=k\}}u_g u_g'\right)\Sigma_n^{*-1}\right]$$

$$-\frac{1}{2n(1-\rho_n^*)^2}K_p^-\,\text{Vec}\,(\Sigma_n^{*-1}V'V\Sigma_n^{*-1})$$

$$\xrightarrow{p} 2^{-1}\delta\sum_{k>1}^{K}f_k\frac{k-1}{1+(k-1)\rho_0}K_p^-\,\text{Vec}\,\Sigma_0^{-1}$$

$$-2^{-1}(1-\delta)\frac{1}{1-\rho_0}K_p^-\,\text{Vec}\,\Sigma_0^{-1}$$

$$= 2^{-1}\left[\delta \sum_{k>1}^{K} f_k \frac{k-1}{1+(k-1)\rho_0} - \frac{1-\delta}{1-\rho_0}\right] K_p^{-} \operatorname{Vec} \Sigma_0^{-1}.$$

$$En^{-1}\left(\frac{\partial^2 L_n}{\partial\sigma\partial\rho}\right)_{\theta_0} = \frac{1}{2n} \sum_{k>1}^{K} m_k \frac{k-1}{1+(k-1)\rho_0} K_p^{-} \operatorname{Vec} \Sigma_0^{-1}$$

$$- \frac{n-G}{2n(1-\rho_0)} K_p^{-} \operatorname{Vec} \Sigma_0^{-1}$$

So, $\lim En^{-1}\left(\dfrac{\partial^2 L_n}{\partial\sigma\partial\rho}\right)_{\theta_0} = \operatorname{plim} n^{-1}\left(\dfrac{\partial^2 L_n}{\partial\sigma\partial\rho}\right)_{\theta_n^*}$.

$$n^{-1}\left(\frac{\partial^2 L_n}{\partial\rho^2}\right)_{\theta_n^*} = -\frac{p}{2n} \sum_{k>1}^{K} m_k \left[\frac{k-1}{1+(k-1)\rho_n^*}\right]^2 - \frac{n-G}{2n} p \frac{1}{(1-\rho_n^*)^2}$$

$$+ \frac{1}{n} \sum_{k>1}^{K} \frac{(k-1)^2}{[1+(k-1)\rho_n^*]^3} \operatorname{tr}\left[\left(\sum_{\{g:n_g=k\}} u_g u_g'\right) \Sigma_n^{*-1}\right]$$

$$+ \frac{1}{n(1-\rho_n^*)^3} \operatorname{tr} V' V \Sigma_n^{*-1}$$

$$\xrightarrow{p} -\frac{1}{2} \delta p \sum_{k>1}^{K} f_k \left[\frac{k-1}{1+(k-1)\rho_0}\right]^2 - \frac{1}{2}(1-\delta) p \frac{1}{(1-\rho_0)^2}$$

$$+ \delta p \sum_{k>1}^{K} f_k \left[\frac{k-1}{1+(k-1)\rho_0}\right]^2 + (1-\delta) p \frac{1}{(1-\rho_0)^2}$$

$$= \frac{1}{2} p \left\{\delta \sum_{k>1}^{K} f_k \left[\frac{k-1}{1+(k-1)\rho_0}\right]^2 + \frac{1-\delta}{(1-\rho_0)^2}\right\}$$

$$= \lim \mathrm{E}\ n^{-1}(\partial^2 \mathrm{L}_n / \partial \rho^2)_{\theta_0}.$$

Therefore, $n^{-1}(\partial^2 \mathrm{L}_n / \partial\theta\partial\theta')_{\theta_n^*}$ converges to $\mathrm{A}(\theta_0) = \lim \mathrm{E}\ n^{-1}(\partial^2 \mathrm{L}_n / \partial\theta\partial\theta')_{\theta_0}$ in probability for any sequence θ_n^* such that plim $\theta_n^* = \theta_0$.

$\mathrm{A}(\theta_0) =$

$$\frac{1}{2}\begin{bmatrix} \mathrm{K}_p^-(\Sigma_0^{-1} \otimes \Sigma_0^{-1})\mathrm{K}_p^{-\prime} & \left[\delta \sum_{k>1}^{K} f_k \frac{k-1}{1+(k-1)\rho_0} - \frac{1-\delta}{1-\rho_0}\right]\mathrm{K}_p^{-}\mathrm{Vec}\Sigma_0^{-1} \\ \text{Sym.} & \left[\delta \sum_{k>1}^{K} f_k \left(\frac{k-1}{1+(k-1)\rho_0}\right)^2 + \frac{1-\delta}{(1-\rho_0)^2}\right]p \end{bmatrix}.$$

The positive definiteness of A (θ_0) is to be shown next.

Let $\mathrm{A}(\theta_0) = 2^{-1}\begin{bmatrix} \mathrm{K}_p^- & 0 \\ 0 & 1 \end{bmatrix} \bar{\mathrm{A}}(\theta_0) \begin{bmatrix} \mathrm{K}_p^{-\prime} & 0 \\ 0 & 1 \end{bmatrix}$,

$$\bar{\mathrm{A}}(\theta_0) = \begin{bmatrix} \Sigma_0^{-1} \otimes \Sigma_0^{-1} & \left[\delta \sum_{k>1}^{K} f_k \frac{k-1}{1+(k-1)\rho_0} - \frac{1-\delta}{1-\rho_0}\right]\mathrm{Vec}\Sigma_0^{-1} \\ \text{Sym.} & \left[\delta \sum_{k>1}^{K} f_k \left(\frac{k-1}{1+(k-1)\rho_0}\right)^2 + \frac{1-\delta}{(1-\rho_0)^2}\right]p \end{bmatrix}$$

$$= \begin{bmatrix} \bar{\mathrm{A}}_{11} & \bar{\mathrm{A}}_{12} \\ \bar{\mathrm{A}}_{21} & \bar{\mathrm{A}}_{22} \end{bmatrix}.$$

Since $\overline{A}_{11} = \Sigma_0^{-1} \otimes \Sigma_0^{-1}$ is positive definite, $\left|\overline{A}(\theta_0)\right| = \left|\overline{A}_{11}\right| \; \left|\overline{A}_{22} - \overline{A}_{21} \overline{A}_{11}^{-1} \overline{A}_{12}\right|$ (see, e.g., Graybill, 1983: 184), and $\overline{A}(\theta_0)$ is positive definite if and only if $\left|\overline{A}(\theta_0)\right| > 0$ (see, e.g., Basilevsky, 1983: 135).

$$\left|\overline{A}(\theta_0)\right| = \left|\overline{A}_{11}\right| \; \left|\overline{A}_{22} - \overline{A}_{21} \overline{A}_{11}^{-1} \overline{A}_{12}\right|$$

$$= p\left|\Sigma_0\right|^{-2p} \left\{ \frac{\delta(1-\delta)}{(1-\rho_0)^2} + 2\frac{\delta(1-\delta)}{1-\rho_0} \sum_{k>1}^{K} f_k \frac{k-1}{1+(k-1)\rho_0} \right.$$

$$\left. + \delta\left[\sum_{k>1}^{K} f_k \left(\frac{k-1}{1+(k-1)\rho_0} \right)^2 - \delta\left(\sum_{k>1}^{K} f_k \frac{k-1}{1+(k-1)\rho_0} \right)^2 \right] \right\}$$

$$> \delta p\left|\Sigma_0\right|^{-2p} \left[\sum_{k>1}^{K} f_k \left(\frac{k-1}{1+(k-1)\rho_0} \right)^2 - \left(\sum_{k>1}^{K} f_k \frac{k-1}{1+(k-1)\rho_0} \right)^2 \right]$$

$$= \delta p\left|\Sigma_0\right|^{-2p} \sum_{k>1}^{K} f_k \left[\frac{k-1}{1+(k-1)\rho_0} - \sum_{k>1}^{K} f_k \frac{k-1}{1+(k-1)\rho_0} \right]^2 \geq 0.$$

So, $\left|\overline{A}(\theta_0)\right| > 0$ and $\overline{A}(\theta_0) > 0$. Note that $\begin{bmatrix} K_p^{-\prime} & 0 \\ 0 & 1 \end{bmatrix}$ is of full column rank $p^* + 1$. Therefore, $A(\theta_0)$ is positive definite.

□□

Verification of Condition (C)

The first derivative of L_n with respect to θ at θ_0 is rewritten before the asymptotic distribution of $n^{-1/2}(\partial L_n/\partial\theta)_{\theta_0}$ is shown.

$$\left(\frac{\partial L_n}{\partial\sigma}\right)_{\theta_0}=2^{-1}K_p^{-}\left\{n\mathrm{Vec}\Sigma_0^{-1}-(\Sigma_0^{-1}\otimes\Sigma_0^{-1})\mathrm{Vec}\sum_{k>1}^{K}\sum_{\{g:n_g=k\}}\frac{u_g u_g'}{1+(k-1)\rho_0}\right.$$

$$\left.-(\Sigma_0^{-1}\otimes\Sigma_0^{-1})(n-G)\mathrm{Vec}\left[\frac{V'V}{(n-G)(1-\rho_0)}\right]\right\}$$

$$=2^{-1}K_p^{-}\left\{n\,\mathrm{Vec}\,\Sigma_0^{-1}-G\,\mathrm{Vec}\,\Sigma_0^{-1}\right.$$

$$-(\Sigma_0^{-1}\otimes\Sigma_0^{-1})\left[\sum_{k>1}^{K}m_k\ \mathrm{Vec}\left(\frac{1}{m_k}\sum_{\{g:n_g=k\}}\frac{u_g u_g'}{1+(k-1)\rho_0}-\Sigma_0\right)\right]$$

$$\left.-(n-G)(\Sigma_0^{-1}\otimes\Sigma_0^{-1})\mathrm{Vec}\left[\frac{V'V}{(n-G)(1-\rho_0)}-\Sigma_0\right]-(n-G)\mathrm{Vec}\ \Sigma_0^{-1}\right\}$$

$$=-2^{-1}K_p^{-}(\Sigma_0^{-1}\otimes\Sigma_0^{-1})\left\{\sum_{k>1}^{K}m_k\ \mathrm{Vec}\left[\frac{1}{m_k}\sum_{\{g:n_g=k\}}\frac{u_g u_g'}{1+(k-1)\rho_0}-\Sigma_0\right]\right.$$

$$\left.+(n-G)\mathrm{Vec}\left[\frac{V'V}{(n-G)(1-\rho_0)}-\Sigma_0\right]\right\}.$$

$$\left(\frac{\partial L_n}{\partial\rho}\right)_{\theta_0}=2^{-1}\left\{p\sum_{g=1}^{G}\frac{n_g-1}{1+(n_g-1)\rho_0}-p(n-G)\frac{1}{1-\rho_0}\right.$$

$$-\sum_{k>1}^{K} m_k \frac{k-1}{1+(k-1)\rho_0} \operatorname{tr}\left[\left(\frac{1}{m_k} \sum_{\{g: n_g = k\}} \frac{u_g u_g'}{1+(k-1)\rho_0} - \Sigma_0\right)\Sigma_0^{-1}\right]$$

$$-p \sum_{k>1}^{K} m_k \frac{k-1}{1+(k-1)\rho_0}$$

$$\left. + \frac{n-G}{1-\rho_0} \operatorname{tr}\left[\left(\frac{V'V}{(n-G)(1-\rho_0)} - \Sigma_0\right)\Sigma_0^{-1}\right] + p(n-G)\frac{1}{1-\rho_0}\right\}$$

$$= -2^{-1} \operatorname{Vec}'\Sigma_0^{-1}\left\{\sum_{k>1}^{K} m_k \frac{k-1}{1+(k-1)\rho_0} \operatorname{Vec}\left[\frac{1}{m_k} \sum_{\{g:n_g=k\}} \frac{u_g u_g'}{1+(k-1)\rho_0} - \Sigma_0\right]\right.$$

$$\left. - \frac{n-G}{1-\rho_0} \operatorname{Vec}\left[\frac{V'V}{(n-G)(1-\rho_0)} - \Sigma_0\right]\right\}.$$

$$\frac{1}{\sqrt{n}}\left(\frac{\partial L_n}{\partial \theta}\right)_{\theta_0} = \frac{1}{\sqrt{n}} \begin{bmatrix} \dfrac{\partial L_n}{\partial \sigma} \\ \\ \dfrac{\partial L_n}{\partial \rho} \end{bmatrix}_{\theta_0} = -\frac{1}{2\sqrt{n}} \Delta \quad D, \quad \text{Where}$$

$\triangle =$

$$\begin{bmatrix} K_p^-(\Sigma_0^{-1}\otimes\Sigma_0^{-1})\left[(\sqrt{m_2}\ \sqrt{m_3}\ \ldots\sqrt{m_K}\ \sqrt{n-G})\otimes I_{p^2}\right] \\ \mathrm{Vec}'\Sigma_0^{-1}\left[\left(\sqrt{m_2}\dfrac{2-1}{1+\rho_0}\ \sqrt{m_3}\dfrac{3-1}{1+2\rho_0}\ \ldots\sqrt{m_K}\dfrac{K-1}{1+(K-1)\rho_0}\ \ -\dfrac{\sqrt{n-G}}{1-\rho_0}\right)\otimes I_{p^2}\right] \end{bmatrix},$$

and,

$$D=\begin{bmatrix} \sqrt{m_2}\ \mathrm{Vec}\left[\dfrac{1}{m_2}\ \underset{\{g:n_g=2\}}{\Sigma}\dfrac{u_g u_g'}{1+(2-1)\rho_0}-\Sigma_0\right] \\ \sqrt{m_3}\ \mathrm{Vec}\left[\dfrac{1}{m_3}\ \underset{\{g:n_g=3\}}{\Sigma}\dfrac{u_g u_g'}{1+(3-1)\rho_0}-\Sigma_0\right] \\ \cdot \\ \cdot \\ \cdot \\ \sqrt{m_K}\ \mathrm{Vec}\left[\dfrac{1}{m_K}\ \underset{\{g:n_g=K\}}{\Sigma}\dfrac{u_g u_g'}{1+(K-1)\rho_0}-\Sigma_0\right] \\ \sqrt{n-G}\ \mathrm{Vec}\left[\dfrac{V'V}{(n-G)(1-\rho_0)}-\Sigma_0\right] \end{bmatrix}.$$

All the u_g's and rows of V are mutually independent. For every k, $\sqrt{m_k}$ Vec $\left[\dfrac{1}{m_k}\ \underset{\{g:n_g=k\}}{\Sigma}\dfrac{u_g u_g'}{1+(k-1)\rho_0}-\Sigma_0\right]\to N[0,2M_p(\Sigma_0\otimes\Sigma_0)]$, where $M_p=K_pK_p^-$ is a $p^2\times p^2$ symmetric idempotent matrix (see, e.g., Browne, 1974). And

$\sqrt{n-G} \quad \text{Vec} \left[\frac{V'V}{(n-G)(1-\rho_0)} - \Sigma_0 \right]$ has the same asymptotic distribution.

The joint asymptotic distribution of D is $N\left[0, \quad I \otimes 2M_p(\Sigma_0 \otimes \Sigma_0)\right]$.

It can be shown through matrix algebra that $\frac{1}{\sqrt{n}}\left(\frac{\partial L_n}{\partial \theta}\right)_{\theta_0}$ is asymptotically distributed as N[0, B(θ_0)], and B(θ_0) is identical to A(θ_0).

□□

Verification of Condition (D)

Theorem 4.2.2 of Amemiya (1985) is used to show condition (D) on uniform convergence of the likelihood function.

Define $\tilde{\Theta}$, a compact subset of q-dimensional Euclidean space, as

$\tilde{\Theta} = \{\theta = (\Sigma, \rho): \lambda_2 I_p \le \Sigma \le \lambda_1 I_p, -\varepsilon \le \rho \le \gamma$, for $0 < \lambda_2 < \lambda_1$, and some $\varepsilon < K^{-1}$ and $\gamma < 1\}$. Note $\Theta \subset \tilde{\Theta}$. Let $h_i(y_i, \theta) = \ln f_i(y_i, \theta) - E_{\theta_0} \ln f_i(y_i, \theta)$, where $f_i(y_i, \theta)$ is the normal density function for y_i. Note $E_{\theta_0} h_i(y_i, \theta) = 0$. It is to be shown that for every i, there exists some $\delta > 0$

such that $E \sup_{\theta \in \tilde{\Theta}} \left|h_i(y_i, \theta)\right|^{1+\delta} < \infty$. Then, $n^{-1} \sum_{i=1}^{n} h_i(y_i, \theta)$ converges to 0 in probability uniformly in $\theta \in \tilde{\Theta}$ according to Theorem 4.2.2 of Amemiya (1985). So $n^{-1} L_n(\theta)$ converges to a nonstochastic function L(θ) in probability uniformly in θ in an open neighborhood of θ_0.

Let $n_0 = 0$.

For $i = \sum_{k=0}^{g-1} n_k + 1$, where g = 1, 2, ..., G,

$$h_i(y_i,\theta)=\frac{1}{1+(n_g-1)\rho}\operatorname{tr}\{y_iy_i'-[1+(n_g-1)\rho_0]\Sigma_0\}\Sigma^{-1}.$$

Otherwise,

$$h_i(y_i,\theta)=\frac{1}{1-\rho}\operatorname{tr}\{y_iy_i'-(1-\rho_0)\Sigma_0\}\Sigma^{-1}.$$

Let us first discuss the case where $i=\sum_{k=0}^{g-1} n_k+1$, with g = 1, 2, ..., G.

Note $0<\dfrac{1}{1+(n_g-1)\rho}<K$.

$$\begin{aligned}\left|h_i(y_i,\theta)\right| &< K\left|\operatorname{tr}\{y_iy_i'-[1+(n_g-1)\rho_0]\Sigma_0\}\Sigma^{-1}\right|\\ &=K\left|y_i'\Sigma^{-1}y_i-[1+(n_g-1)\rho_0]\operatorname{tr}(\Sigma_0\Sigma^{-1})\right|.\end{aligned}$$

If $y_i'\Sigma^{-1}y_i\geq[1+(n_g-1)\rho_0]\operatorname{tr}(\Sigma_0\Sigma^{-1})$,

$$\begin{aligned}\left|h_i(y_i,\theta)\right| &< Ky_i'\Sigma^{-1}y_i\\ &=K(\Sigma_0^{-1/2}y_i)'\Sigma_0^{1/2}\Sigma^{-1}\Sigma_0^{1/2}(\Sigma_0^{-1/2}y_i).\end{aligned}$$

Let $z_i=\Sigma_0^{-1/2}y_i, z_i\sim N(0,[1+(n_g-1)\rho_0]I_p)$.

Because $\lambda_2 I_p\leq\Sigma\leq\lambda_1 I_p$, we have $\lambda(\Sigma_0^{1/2}\Sigma^{-1}\Sigma_0^{1/2})\leq\lambda_2^{-1}\lambda_{\max}(\Sigma_0)=M$, say.

So, $\Sigma_0^{1/2}\Sigma^{-1}\Sigma_0^{1/2}\leq MI_p$.

Therefore, $\left|h_i(y_i,\theta)\right|<Kz_i'(MI_p)z_i$

$$=KM(z_i'z_i).$$

$$\sup_{\theta\in\tilde{\Theta}}\left|h_i(y_i,\theta)\right|^2<K^2M^2(z_i'z_i)^2.$$

$$E \sup_{\theta \in \tilde{\Theta}} \left| h_i(y_i, \theta) \right|^2 < K^2 M^2 E(z_i' z_i)^2$$

$$\leq 3p^2 K^2 M^2 [1 + (n_g - 1)\rho_0]^2 < \infty \text{ for every g.}$$

If $y_i' \Sigma^{-1} y_i < [1 + (n_g - 1)\rho_0] \mathrm{tr}(\Sigma_0 \Sigma^{-1})$,

$$\left| h_i(y_i, \theta) \right| < K[1 + (n_g - 1)\rho_0] \mathrm{tr}(\Sigma_0 \Sigma^{-1})$$

$$\leq pKM[1 + (K - 1)\rho_0] < \infty \text{ for every g.}$$

Therefore, when $i = \sum_{k=0}^{g-1} n_k + 1$, with g = 1, 2, ...,G, there exists some $\delta > 0$

such that $E \sup_{\theta \in \tilde{\Theta}} \left| h_i(y_i, \theta) \right|^{1+\delta} < \infty$.

A similar relationship can be derived for other $h_i(y_i, \theta)$'s, noting

$\frac{K}{K+1} < \frac{1}{1-\rho} \leq \frac{1}{1-\gamma} < \infty$. Details are given in Weng (1990). Therefore, for

every i, there exists sone δ >0 such that $E \sup_{\theta \in \tilde{\Theta}} \left| h_i(y_i, \theta) \right|^{1+\delta} < \infty$. This

completes the proof. □□

Verification of Condition (E)

$$\operatorname{plim} \frac{1}{n} \frac{\partial^2 L_n}{\partial \sigma \partial \sigma'} = \operatorname{plim} \frac{1}{2n} K_p^- \left\{ -n(\Sigma^{-1} \otimes \Sigma^{-1}) \right.$$

$$+\sum_{k>1}^{K}\frac{1}{1+(k-1)\rho}\left[\Sigma^{-1}\otimes\Sigma^{-1}\left(\sum_{\{g:n_g=k\}}u_g u_g'\right)\Sigma^{-1}+\Sigma^{-1}\left(\sum_{\{g:n_g=k\}}u_g u_g'\right)\Sigma^{-1}\otimes\Sigma^{-1}\right]$$

$$\left.+\frac{1}{1-\rho}(\Sigma^{-1}\otimes\Sigma^{-1}V'V\Sigma^{-1}+\Sigma^{-1}V'V\Sigma^{-1}\otimes\Sigma^{-1})\right\}K_p^{-'}$$

$$=2^{-1}K_p^{-}\left\{-(\Sigma^{-1}\otimes\Sigma^{-1})\right.$$

$$+\delta\sum_{k>1}^{K}f_k\frac{1+(k-1)\rho_0}{1+(k-1)\rho}(\Sigma^{-1}\otimes\ \Sigma^{-1}\Sigma_0\Sigma^{-1}+\Sigma^{-1}\Sigma_0\Sigma^{-1}\otimes\ \Sigma^{-1})$$

$$\left.+(1-\delta)\frac{1-\rho_0}{1-\rho}(\Sigma^{-1}\otimes\ \Sigma^{-1}\Sigma_0\Sigma^{-1}+\Sigma^{-1}\Sigma_0\Sigma^{-1}\otimes\ \Sigma^{-1})\right\}K_p^{-'}.$$

$$\mathrm{p}\lim\frac{1}{n}\frac{\partial^2 L_n}{\partial\sigma\partial\rho}=\mathrm{plim}\frac{1}{2n}K_p^{-}\left\{\sum_{k>1}^{K}\frac{k-1}{[1+(k-1)\rho]^2}\mathrm{Vec}\left[\Sigma^{-1}\left(\sum_{\{g:n_g=k\}}u_g u_g'\right)\Sigma^{-1}\right]\right.$$

$$\left.-\frac{1}{(1-\rho)^2}\mathrm{Vec}(\Sigma^{-1}V'V\Sigma^{-1})\right\}$$

$$= 2^{-1}\left\{ \delta \sum_{k>1}^{K} f_k \frac{(k-1)[1+(k-1)\rho_0]}{[1+(k-1)\rho]^2} - (1-\delta)\frac{1-\rho_0}{(1-\rho)^2} \right\} K_p^{-} \mathrm{Vec}(\Sigma^{-1}\Sigma_0\Sigma^{-1}).$$

$$\operatorname{plim} \frac{1}{n}\frac{\partial^2 L_n}{\partial \rho^2} = \operatorname{plim} \frac{1}{2n}\left\{ -p \sum_{k>1}^{K} m_k \left[\frac{k-1}{1+(k-1)\rho}\right]^2 - p(n-G)\frac{1}{(1-\rho)^2} \right.$$

$$+2 \sum_{k>1}^{K} \frac{(k-1)^2}{[1+(k-1)\rho]^3} \operatorname{tr}\left(\sum_{\{g:n_g=k\}} u_g u_g' \right) \Sigma^{-1}$$

$$\left. +2\frac{1}{(1-\rho)^3} \operatorname{tr} \quad V'V\Sigma^{-1} \right\}$$

$$= 2^{-1}\left\{ -\delta \quad p \quad \sum_{k>1}^{K} f_k \left[\frac{k-1}{1+(k-1)\rho}\right]^2 - (1-\delta)p\frac{1}{(1-\rho)^2} \right\}$$

$$+\left\{ \delta \sum_{k>1}^{K} f_k \frac{(k-1)^2[1+(k-1)\rho_0]}{[1+(k-1)\rho]^3} + (1-\delta)\frac{1-\rho_0}{(1-\rho)^3} \right\} \operatorname{tr} \quad \Sigma_0\Sigma^{-1}.$$

Since plim $n^{-1}\partial^2 L_n / \partial\theta\partial\theta'$ exists and is continuous for every $\theta \in \Theta$, Condition (E) is satisfied. □□