Additional Information and Precision of Estimators in Multivariate Structural Models

YUTAKA KANO, PETER M. BENTLER, and AB MOOIJAART

Department of Mathematical Sciences, College of Engineering, University of Osaka Prefecture, Sakai, Osaka 593, JAPAN,
Department of Psychology, University of California, Los Angeles, CA 90024-1563, U.S.A., and
Department of Psychometrics and Research, Faculty of Social Sciences, University of Leiden, Leiden, THE NETHERLANDS

Abstract. This paper investigates the effect of additional information upon parameter estimation in multivariate structural models. It is shown that the asymptotic covariances of estimators based on a model with additional variables are smaller than those based on a model with no additional variables, where the estimation methods employed are the methods of maximum likelihood and minimum chi-square. Some applications to moment structure models are provided.

Key words: Additional observation, asymptotic covariance matrix, Fisher information matrix, method of maximum likelihood, method of minimum chi-square, nuisance parameters.

1. INTRODUCTION

In multivariate analysis it is important to investigate effects of additional variables as well as the latent structure of given observed variables. In principal component analysis, discriminant analysis, and canonical correlation analysis, this issue has been studied in some detail and the results can be applied to variable selection (see e.g., Fujikoshi [1]; Fujikoshi, Krishnaiah, and Schmidhammer [2]; Wijsman [3]). In principal component analysis, for example, the largest eigenvalues of the sample covariance matrix of a set of variables are usually smaller than the corresponding ones based on the covariance matrix of original variables augmented by additional variables. By statistically observing the increase in the magnitude of the eigenvalues, one can investigate the effect of the additional variables, e.g., whether the added variables significantly contribute to the principal components. Rao [4] investigated the effect of additional variables upon the power of tests in multivariate regression.

In this article, we study a similar but alternative type of effects of additional variables in multivariate structural analysis. We define a small model and a large model in which the variables in the small model are part of those in the large model, and then estimators for a common parameter θ based on the two models are compared. A more specific setup is the following.

Let $\mathbf{y}_1$ and $\mathbf{y}_2$ be random p_1- and p_2-vectors. Assume that the distribution of $\mathbf{y}_1$ involves an unknown parameter vector θ; the distribution of $\mathbf{y}_2$ relates not only to the parameter θ but to a nuisance parameter ϕ as well. The joint distribution of $[\mathbf{y}_1', \mathbf{y}_2']'$ then depends on $[\theta', \phi']'$. We will call $\mathbf{y}_1$ and $[\mathbf{y}_1', \mathbf{y}_2']'$ a small model and a large model, respectively. The subject of this note is to investigate whether inference regarding θ is better based on the small model or based on the large model. The (asymptotic) covariance matrices of estimators for θ based on the two models are compared. The large model may have more information about θ than the small one, but the nuisance parameter ϕ involved in the large model may disturb estimation of θ.

The above situation appears in many models in multivariate structural analysis as described in Section 3. Here let us take a multiple regression model as an example. Let

$$\mathrm{E}[\mathbf{y}_1] = X_{11}\beta_1, \tag{1.1}$$

where X_{11} is a design matrix of order $p_1 \times k_1$ and β_1 is a k_1-vector of regression coefficients. Assume that the covariance matrix of $\mathbf{y}_1$ is $\sigma^2 I_{p_1}$. This is a small model. Then the best linear unbiased estimator (BLUE) for β_1 is given as $\widehat{\beta}_1 = (X_{11}'X_{11})^{-1}X_{11}'\mathbf{y}_1$, and $\mathrm{Var}(\widehat{\beta}_1) = \sigma^2(X_{11}'X_{11})^{-1}$.

A large model is constructed by adding extra observations $\mathbf{y}_2$ to $\mathbf{y}_1$. The covariance matrix of $[\mathbf{y}_1', \mathbf{y}_2']'$ is assumed to be $\sigma^2 I_{p_1+p_2}$. When the independent variables are the same as in the small model, the large model can be expressed in the form:

$$\mathrm{E}\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}X_{11}\\ X_{21}\end{bmatrix}\beta_1, \tag{1.2}$$

where X_{21} is of $p_2 \times k_1$. The covariance matrix of the BLUE for β_1 in the large model is $\sigma^2(X_{11}'X_{11} + X_{21}'X_{21})^{-1}$, which is obviously smaller than $\mathrm{Var}(\widehat{\beta}_1)$ in the small model. This is because the large model has more observations (or information) and no nuisance parameters to be estimated.

It is often argued that more observations need to introduce more independent variables. In such a case, the large model can be described as

$$\mathrm{E}\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}X_{11} & 0\\ X_{21} & X_{22}\end{bmatrix}\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}, \tag{1.3}$$

where X_{22} is of $p_2 \times k_2$ and β_2 is a k_2-vector. Note that X_{12}=0 since the small model holds. In this circumstance, $\theta = [\beta_1', \sigma^2]'$ that is a parameter vector of interest, while $\phi = \beta_2$. Let $(\widetilde{\beta}_1, \widetilde{\beta}_2)$ denote the BLUE in the regression model (1.3). It will be shown that

$$\mathrm{Var}(\widetilde{\beta}_1) = \sigma^2\left(\begin{bmatrix}X_{11} & 0\\ X_{21} & X_{22}\end{bmatrix}'\begin{bmatrix}X_{11} & 0\\ X_{21} & X_{22}\end{bmatrix}\right)^{11} \le \sigma^2(X_{11}'X_{11})^{-1} = \mathrm{Var}(\widehat{\beta}_1), \tag{1.4}$$

where A^{11} denotes the (1,1) block of the inverse matrix A^{-1}. Consequently the large model permits better inference than the small one.

In this article, we shall show that a result similar to (1.4) holds for a wide variety of statistical models. The estimation methods treated here are the methods of maximum likelihood and minimum chi-square due to Ferguson [5]. The results are applied to moment structure models, including a factor analysis model in Section 3. In Section 4, we discuss the relationship of our results to well-known inequalities.

2. MAIN RESULTS

First we consider the method of maximum likelihood. Let the distribution of $\mathbf{y}_1$ permit the density function $f(y_1|\theta)$; let $(\mathbf{y}_1, \mathbf{y}_2)$ permit $g(y_1, y_2|\theta, \phi)$. The first one is a small model; the latter is a large model. Assume that the model parameters are identified in both models and that these density functions satisfy the usual regularity conditions to derive the following asymptotic results. The purpose here is to show that the asymptotic covariance matrix of the maximum likelihood estimator (MLE) for θ based on the large model is smaller than that based on the small model.

The Fisher information matrices in the small and large models will be written as

$$i_{\theta\theta} = \mathrm{E}\left[\left(\frac{f_\theta}{f}\right)\left(\frac{f_\theta}{f}\right)'\right],$$

$$I = \begin{bmatrix} I_{\theta\theta} & I_{\theta\phi} \\ I_{\phi\theta} & I_{\phi\phi} \end{bmatrix} = \mathrm{E}\left(\begin{bmatrix} \frac{g_\theta}{g} \\ \frac{g_\phi}{g} \end{bmatrix}\begin{bmatrix} \frac{g_\theta}{g} \\ \frac{g_\phi}{g} \end{bmatrix}'\right).$$

Here we have written $f_\theta = \frac{\partial}{\partial\theta} f(\mathbf{y}_1|\theta)$, $g_\theta = \frac{\partial}{\partial\theta} g(\mathbf{y}_1, \mathbf{y}_2|\theta, \phi)$, and $g_\phi = \frac{\partial}{\partial\phi} g(\mathbf{y}_1, \mathbf{y}_2|\theta, \phi)$. As is well-known, the asymptotic covariance matrix of the MLE is given by the inverse of the Fisher information matrix. As a consequence, the inequality to be shown is

$$i_{\theta\theta}^{-1} \geq I^{11} = \left(I_{\theta\theta} - I_{\theta\phi} I_{\phi\phi}^{-1} I_{\phi\theta}\right)^{-1}. \tag{2.1}$$

A result (b) on page 331 in Rao [6] mentions that the information matrix based on the observed vector $(\mathbf{y}_1, \mathbf{y}_2)$ is not less than that based on any function of $(\mathbf{y}_1, \mathbf{y}_2)$. This result implies that the difference

$$\begin{bmatrix} I_{\theta\theta} & I_{\theta\phi} \\ I_{\phi\theta} & I_{\phi\phi} \end{bmatrix} - \begin{bmatrix} i_{\theta\theta} & 0 \\ 0 & 0 \end{bmatrix}$$

is non-negative definite, which further implies that

$$(I_{\theta\theta} - i_{\theta\theta}) - I_{\theta\phi} I_{\phi\phi}^{-1} I_{\phi\theta}$$

is also non-negative. Thus, (2.1) follows.

A direct proof that does not use Rao's result will be given in the APPENDIX.

Next we shall consider the method of minimum chi-square (Ferguson [5]) to estimate parameters. Let $\mathbf{y}_1$ and $\mathbf{y}_2$ be statistics depending on a sample size n such that

$$\sqrt{n}\begin{bmatrix} \mathbf{y}_1 - m_1(\theta) \\ \mathbf{y}_2 - m_2(\theta, \phi) \end{bmatrix} \xrightarrow{L} N(0, V),$$

where $m_1(\theta)$ and $m_2(\theta, \phi)$ are the asymptotic expectations of $\mathbf{y}_1$ and $\mathbf{y}_2$, and $V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$ is the asymptotic covariance matrix of $[\mathbf{y}_1', \mathbf{y}_2']'$. Notice that some elements of θ may be unrelated to $m_2(\theta, \phi)$.

According to Ferguson [5], the estimator $\widehat{\theta}$ based on the small model $\mathbf{y}_1$ is defined by a solution to

$$\min_\theta (\mathbf{y}_1 - m_1(\theta))' \widehat{V}_{11}^{-1} (\mathbf{y}_1 - m_1(\theta))$$

with $\widehat{V}_{11}$ a consistent estimator for V_{11}. Let $\Delta_{11} = \dfrac{\partial m_1(\theta)}{\partial \theta'}$. The asymptotic covariance matrix of $\widehat{\theta}$ is then represented as $(\Delta_{11}' V_{11}^{-1} \Delta_{11})^{-1}$. In the same manner, the estimator $(\widetilde{\theta}, \widetilde{\phi})$ based on the large model is defined through

$$\min_{(\theta,\phi)} \begin{bmatrix} \mathbf{y}_1 - m_1(\theta) \\ \mathbf{y}_2 - m_2(\theta,\phi) \end{bmatrix}' \widehat{V}^{-1} \begin{bmatrix} \mathbf{y}_1 - m_1(\theta) \\ \mathbf{y}_2 - m_2(\theta,\phi) \end{bmatrix}, \tag{2.2}$$

where $\widehat{V}$ is also a consistent estimator for V. Let $\Delta_{21} = \dfrac{\partial m_2(\theta,\phi)}{\partial \theta'}$ and $\Delta_{22} = \dfrac{\partial m_2(\theta,\phi)}{\partial \phi'}$. The asymptotic covariance matrix of $\widetilde{\theta}$ is given by

$$\left(\begin{bmatrix} \Delta_{11} & 0 \\ \Delta_{21} & \Delta_{22} \end{bmatrix}' V^{-1} \begin{bmatrix} \Delta_{11} & 0 \\ \Delta_{21} & \Delta_{22} \end{bmatrix} \right)^{11}. \tag{2.3}$$

Thus, what to prove here is

$$(\Delta_{11}' V_{11}^{-1} \Delta_{11})^{-1} \geq \left(\begin{bmatrix} \Delta_{11} & 0 \\ \Delta_{21} & \Delta_{22} \end{bmatrix}' V^{-1} \begin{bmatrix} \Delta_{11} & 0 \\ \Delta_{21} & \Delta_{22} \end{bmatrix} \right)^{11}. \tag{2.4}$$

The proof is very simple. Since V^{-1} can be written as $\begin{bmatrix} V_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}$ plus a nonnegative definite matrix, substitution of this into (2.3) leads to

$$\left(\begin{bmatrix} \Delta_{11}' V_{11}^{-1} \Delta_{11} & 0 \\ 0 & 0 \end{bmatrix} + M \right)^{11}$$

for some nonnegative definite matrix $M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$. Thus, we have

$$\begin{aligned} \begin{bmatrix} \Delta_{11}' V_{11}^{-1} \Delta_{11} + M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}^{11} &= (\Delta_{11}' V_{11}^{-1} \Delta_{11} + M_{11} - M_{12} M_{22}^{-1} M_{21})^{-1} \\ &\leq (\Delta_{11}' V_{11}^{-1} \Delta_{11})^{-1}, \end{aligned}$$

since $M_{11} - M_{12} M_{22}^{-1} M_{21}$ is nonnegative definite. This proves (2.4).

Consequently, it is seen that the larger model always makes better inference concerning θ as long as the above two estimation methods are employed. These results trivially extend to estimators based on a wider class of discrepancy functions in view of the equivalence of minimum chi-square and other minimum discrepancy functions (Shapiro [7]) and the asymptotic equivalence of MLE and other generalized least squares estimators in the context of covariance structure analysis (Browne [8]).

The inequality in (1.4) is obtained as a corollary of the result in (2.4) when $\Delta_{ij} = X_{ij}$ and $V = \sigma^2 I_{p_1+p_2}$.

It is important and interesting to investigate when the strict inequalities hold in (2.1) and (2.4). It would be possible to establish a corresponding result on the problem of testing statistical hypotheses. These are left as the future work.

3. EXAMPLES

This work was motivated by Bentler and Chou's [9] paper, which showed numerically that the asymptotic covariances of estimates for the factor loadings in a mean and covariance structure model are substantially smaller than those in the usual factor analysis model which is a covariance structure model.

Let S be the unbiased sample covariance matrix based on a sample of size $N = n+1$. In a covariance structure model, the population covariance matrix is represented by $\Sigma(\theta)$, that is

$$\mathrm{E}[S] = \Sigma(\theta). \tag{3.1}$$

One makes inference concerning θ based only on S. The method of maximum likelihood under the normality minimizes

$$F_{WL}(S, \Sigma(\theta)) = \log|\Sigma(\theta)| - \log|S| + \mathrm{tr}[\Sigma(\theta)^{-1}S] - p.$$

In covariance structure analysis, the method of minimum chi-square due to Ferguson is called the asymptotically distribution-free (ADF) method, which was introduced by Browne ([10], [11]) and developed by Bentler and Dijkstra [12]. The ADF method uses the fact that

$$\sqrt{n}(\mathrm{v}(S) - \mathrm{v}(\Sigma(\theta))) \xrightarrow{L} N(0,\, V_{11}),$$

where $\mathrm{v}(S)$ denotes a p^*-vector consisting of the distinct elements of a symmetric matrix S with $p^* = p(p+1)/2$ and V_{11} is a $p^* \times p^*$ matrix involving the fourth-order moments of observed variables.

Let $\bar{\mathbf{x}}$ be the sample mean vector. A mean and covariance structure model assumes

$$\begin{aligned} \mathrm{E}[S] &= \Sigma(\theta) \\ \mathrm{E}[\bar{\mathbf{x}}] &= \mu(\theta, \phi), \end{aligned} \tag{3.2}$$

where θ is a parameter of interest. Thus S is a small model; $(S, \bar{\mathbf{x}})$ is a large model. Under the normality assumption, the density functions of S and $(S, \bar{\mathbf{x}})$ are, respectively, $W(s \mid \Sigma(\theta)/n, n)$ and $W(s \mid \Sigma(\theta)/n, n) \cdot N(\bar{x} \mid \mu(\theta, \phi), \Sigma(\theta)/n)$. The method of maximum likelihood is equivalent to minimizing

$$F_{WL}(S, \Sigma(\theta)) + (\bar{\mathbf{x}} - \mu(\theta, \phi))'\Sigma(\theta)^{-1}(\bar{\mathbf{x}} - \mu(\theta, \phi)).$$

See Bentler ([13] page 226).

In the general case, we know that

$$\sqrt{n}\left(\begin{bmatrix} \mathrm{v}(S) \\ \bar{\mathbf{x}} \end{bmatrix} - \begin{bmatrix} \mathrm{v}(\Sigma(\theta)) \\ \mu(\theta, \phi) \end{bmatrix}\right) \xrightarrow{L} N(0,\, V),$$

where $V_{22} = \Sigma(\theta)$ and V_{12} is a matrix of the third-order moments. The estimator for (θ, ϕ) is obtained via (2.2). Ferguson's method for this problem was first implemented by Muthén [14]. The results obtained in the previous section show that the estimators for θ in the mean and covariance structure model (3.2) are better than those in the covariance structure model (3.1).

A mean and covariance structure model is naturally introduced from a usual factor analysis model in which

$$\mathbf{x} = \Lambda(\lambda)\mathbf{f} + \mathbf{u}, \tag{3.3}$$

where $\mathbf{f}$ and $\mathbf{u}$ are common and unique factors satisfying

$$\begin{aligned} \mathrm{E}[\mathbf{f}] = 0, \quad \mathrm{E}[\mathbf{u}] = 0, \quad \mathrm{Cov}[\mathbf{f}, \mathbf{u}] = 0, \\ \mathrm{Var}[\mathbf{f}] = \Phi, \quad \text{and} \quad \mathrm{Var}[\mathbf{u}] = \Psi. \end{aligned} \tag{3.4}$$

Then we have

$$\mathrm{Var}[\mathbf{x}] = \Lambda(\lambda)\Phi\Lambda(\lambda)' + \Psi \ \ (= \Sigma(\theta), \ \text{say}). \tag{3.5}$$

The common factor $\mathbf{f}$ sometimes has a nonzero mean vector μ_f. In such a case, the mean of the observable vector is

$$\mathrm{E}[\mathbf{x}] = \Lambda(\lambda)\mu_f. \tag{3.6}$$

The model defined in (3.5) and (3.6) is a mean and covariance structure model (see also Browne [15]). The results obtained previously show that the mean and covariance structure model makes better inference than the covariance structure model (3.5) as far as estimation of $\theta = (\lambda, \Phi, \Psi)$ is concerned. What Bentler and Chou [9] demonstrated numerically is that the standard errors of the $\widehat{\lambda}$ in the mean and covariance structure model are substantially smaller than those in the covariance structure model in the ADF setting. Thus, the result obtained here provides a theoretical basis for their observation.

Muthén [14] discussed the ADF estimation in a multiple-group mean and covariance structure model, where in the g-th population the observed random p-vector is of the form:

$$\mathbf{x}^{(g)} = \mu^{(g)} + \Lambda^{(g)}\mathbf{f}^{(g)} + \mathbf{u}^{(g)}$$

and the structural equation is

$$\mathbf{f}^{(g)} = \alpha^{(g)} + B^{(g)}\mathbf{f}^{(g)} + \zeta^{(g)}$$

for $g = 1, \cdots, G$. Here $\mu^{(g)}$ is a p-vector of location parameters, $\alpha^{(g)}$ is a k-vector of intercepts, $B^{(g)}$ is a $k \times k$ matrix of slopes, $\zeta^{(g)}$ is a k-vector of residuals (for more details, see Muthén [14]). In such a case, let us write

$$\begin{aligned} \mathrm{E}[\mathbf{x}^{(g)}] &= \mu^{(g)}(\theta, \phi^{(g)}), \\ \mathrm{Var}[\mathbf{x}^{(g)}] &= \Sigma^{(g)}(\theta, \phi^{(g)}) \end{aligned}$$

for $g = 1, \cdots, G$, where θ is a common parameter vector across the G populations and $\phi^{(g)}$ appears only in the g-th population.

Consider the following two estimation methods: One is to estimate $(\theta, \phi^{(1)}, \cdots, \phi^{(G)})$ simultaneously based on the pooled samples from the G populations (this is a large model); the other is to estimate $(\theta, \phi^{(g)})$ by using the only sample from the g-th population (this is a small model) for each g. For the second case, we get G estimators for θ. Then, the results obtained here prove that the asymptotic covariance matrix of $\widehat{\theta}$ in the first method is smaller than that of each of the $\widehat{\theta}$'s in the second one. Furthermore, the asymptotic covariance matrix of $\widehat{\phi}^{(g)}$ in the first method is also smaller than that in the second one.

The next example is concerned with an estimation method in a factor analysis model, which was developed by Mooijaart [16], especially for nonnormal observations. He made use of the third-order moments as well as the second-order moments, and applied Ferguson's method to estimate parameters. In the factor analysis model described in (3.3) and (3.4), we further assume that $\mathbf{f}$ and $\mathbf{u}$ are independently distributed. In such a case, the third-order moments of $\mathbf{x}$ can be expressed as

$$\mathrm{E}[\mathbf{x} \otimes \mathbf{x} \otimes \mathbf{x}] = (\Lambda(\lambda) \otimes \Lambda(\lambda) \otimes \Lambda(\lambda))\, \Phi_3 + \Psi_3 \tag{3.7}$$

where $\Phi_3 = \mathrm{E}[\mathbf{f} \otimes \mathbf{f} \otimes \mathbf{f}]$ and $\Psi_3 = \mathrm{E}[\mathbf{u} \otimes \mathbf{u} \otimes \mathbf{u}]$. Let $m_3(\theta, \phi)$ be a column vector consisting of nonduplicated elements of (3.7), where $\phi = (\Phi_3, \Psi_3)$, and let s_3 be a vector of the sample third-order moments corresponding to $m_3(\theta, \phi)$. The estimation method of Mooijaart [16] is to minimize

$$\begin{bmatrix} \mathrm{v}(S) - \mathrm{v}(\Sigma(\theta)) \\ s_3 - m_3(\theta, \phi) \end{bmatrix}' \widehat{V}^{-1} \begin{bmatrix} \mathrm{v}(S) - \mathrm{v}(\Sigma(\theta)) \\ s_3 - m_3(\theta, \phi) \end{bmatrix}. \tag{3.8}$$

Here V is a weight matrix involving up to the 6th-order moments. For more details, see Mooijaart's paper.

In this example, again, it follows that the standard error of $\widehat{\theta}$ based on the large model (3.5) and (3.7) is smaller than that based on the small model (3.5) above.

The final example is two factor analysis models with the same number of factors in which observable variates in one factor model are part of those in the other model, that is, the large model, $[\mathbf{y}_1', \mathbf{y}_2']'$, has the following covariance structure:

$$\mathrm{Var}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}\right) = \begin{bmatrix} \Lambda_1(\lambda) \\ \Lambda_2(\lambda, \gamma) \end{bmatrix} \Phi \begin{bmatrix} \Lambda_1(\lambda) \\ \Lambda_2(\lambda, \gamma) \end{bmatrix}' + \begin{bmatrix} \Psi_1 & 0 \\ 0 & \Psi_2 \end{bmatrix}.$$

The small model, $\mathbf{y}_1$, is then

$$\mathrm{Var}(\mathbf{y}_1) = \Lambda_1(\lambda)\Phi\Lambda_1(\lambda)' + \Psi_1.$$

Several authors have investigated effects of the augmentation of the number of variables in factor analysis, including Bentler and Kano [17], Kano [18], [19], [20], Schneeweiss [21], Williams [22] and others. Kano and Shapiro [23] showed that the asymptotic covariances of the MLE for Ψ_1 decrease as observable variates increase in number. The results obtained here apply to this situation, so that it follows that the asymptotic covariances of the estimators for (λ, Φ, Ψ_1) in the large model are smaller than those in the small model. Therefore, this result includes Kano and Shapiro's as a special case.

4. DISCUSSION

We should distinguish the situation developed in this article from the one which has been often discussed, in which case one model is obtained by replacing with the true value some of the parameters in the other model, that is, one compares between

$$g(y|\theta, \phi_0) \tag{4.1}$$

and

$$g(y|\theta, \phi), \tag{4.2}$$

where ϕ_0 is the true value of ϕ. In the model (4.1), ϕ does not have to be estimated. In such a situation, the asymptotic covariances of the MLE for θ based on the model (4.1) have been shown to be smaller than those on (4.2) by using a well-known inequality:

$$I_{\theta\theta}^{-1} \leq \begin{bmatrix} I_{\theta\theta} & I_{\theta\phi} \\ I_{\phi\theta} & I_{\phi\phi} \end{bmatrix}^{11} = \left(I_{\theta\theta} - I_{\theta\phi} I_{\phi\phi}^{-1} I_{\phi\theta}\right)^{-1}. \tag{4.3}$$

More generally, Altham [24] compared between the MLEs for (θ, ϕ) based on the two models $g(y|\theta, \phi)$ and $g(y|\theta(u), \phi(u))$, where u is a more basic parameter vector and the structures $\theta(u)$ and $\phi(u)$ are known, and showed that the model using the structure permits better inference. In the analysis of covariance structures, this fact was noted by Bentler and Mooijaart [25].

We should notice that the situation treated in this article is different from the above one because in our setting the numbers of observable variates are different between two models to be compared whereas Altham has assumed that they are the same.

The result in Rao [6] that was referred to in Section 2 is closely related to the current work. Suppose that the small model $\mathbf{y}_1$ is dependent on ϕ as well as θ, and let $i = \begin{bmatrix} i_{\theta\theta} & i_{\theta\phi} \\ i_{\phi\theta} & i_{\phi\phi} \end{bmatrix}$ be the Fisher information matrix based on $\mathbf{y}_1$. Then the result in Rao guarantees that $i \leq I$, which implies that $I^{11} \leq i^{11}$, provided that i is nonsingular. The large model yields better estimation. In our setting, unfortunately, the distribution of $\mathbf{y}_1$ is *unrelated* to ϕ, so that the information matrix i is singular. Thus, our results are not a straightforward corollary of Rao's, while the theoretical development here may be minimal. Some practical implications of our results (2.1) and (2.4) are interesting and useful. As shown in Section 3, there are many examples to which the result can be applied. For instance, even when one is not interested in the mean vector but is interested in the covariance structure only, the result obtained here suggests that use of the sample mean could make the inference on the covariance structure more accurate. For another thing, computational output for the standard errors of $\widehat{\theta}$ in a large model should probably be smaller than that in a small model (this is not always true, however); therefore, if the relation does not hold, one should doubt whether the analysis has been made properly. Thus, it is worthwhile noting explicitly the superiority of large models for estimation of θ.

Consider the following inequalities

$$I_{\theta\theta}^{-1} \leq I^{11} \leq i_{\theta\theta}^{-1}.$$

The first inequality is already known as described in (4.3), and the second one is the one we have developed here. These inequalities show that the MLE for θ using the model $g(y_1, y_2|\theta, \phi)$ is better than the MLE based on $f(y_1|\theta)$, and the MLE based on $g(y_1, y_2|\theta, \phi_0)$ is the best.

The corresponding inequalities involving estimators by the method of minimum chi-square are

$$\begin{aligned} \left(\begin{bmatrix} \Delta_{11} \\ \Delta_{21} \end{bmatrix}' V^{-1} \begin{bmatrix} \Delta_{11} \\ \Delta_{21} \end{bmatrix}\right)^{-1} &\leq \left(\begin{bmatrix} \Delta_{11} & 0 \\ \Delta_{21} & \Delta_{22} \end{bmatrix}' V^{-1} \begin{bmatrix} \Delta_{11} & 0 \\ \Delta_{21} & \Delta_{22} \end{bmatrix}\right)^{11} \\ &\leq (\Delta_{11}' V_{11}^{-1} \Delta_{11})^{-1}. \end{aligned} \tag{4.4}$$

Note that the matrix in the far left side in (4.4) represents the asymptotic covariance matrix of the estimator based on (2.2) with a known ϕ_0. It turns out that the best is the estimator using $(m_1(\theta), m_2(\theta, \phi_0))$, the second is the one based on $(m_1(\theta), m_2(\theta, \phi))$, and the worst is based on $m_1(\theta)$.

Recall the multiple regression model described in Section 1. The far left quantity in (4.4) corresponds to the model (1.2), whereas the middle and the right side in (4.4) correspond to the models (1.3) and (1.1) as already noted in Section 2.

Finally, we have assumed the models are correctly specified and the sample is large enough. Generally speaking, a larger model will require more samples than a smaller model in order that the asymptotic results are relevant. Also, in practical situations, it is computationally more difficult to fit a larger model. Nevertheless, these results are quite informative.

ACKNOWLEDGMENT

This work is supported in part by USPHS grants DA00017 and DA01070. The authors would like to thank the referee for pointing out Rao's result which helps to simplify the proof of (2.1).

APPENDIX

We give a simple and direct alternative proof of (2.1).

Since $\mathbf{y}_1$ is the vector of marginal variates of $(\mathbf{y}_1, \mathbf{y}_2)$, we note that

$$f(y_1|\theta) = \int g(y_1, y_2|\theta, \phi) dy_2,$$

from which it follows that

$$\int g_\phi(y_1, y_2|\theta, \phi) dy_2 = \frac{\partial}{\partial \phi} f(y_1|\theta) = 0.$$

From these relations, we can express the conditional expectations of $\frac{g_\phi}{g}$ and $\frac{f_\phi}{f}$ given $\mathbf{y}_1 = y_1$ as follows:

$$\mathrm{E}\left[\frac{g_\phi}{g}\bigg|\mathbf{y}_1 = y_1\right] = \int \frac{g_\phi}{g}\frac{g}{f} dy_2 = \frac{1}{f}\int g_\phi dy_2 = 0, \tag{A1}$$

$$\mathrm{E}\left[\frac{g_\theta}{g}\bigg|\mathbf{y}_1 = y_1\right] = \frac{1}{f}\int g_\theta dy_2 = \frac{1}{f}\frac{\partial}{\partial\theta}\int g dy_2 = \frac{f_\theta}{f}. \tag{A2}$$

By Schwartz' inequality, we can evaluate the following conditional expectation as

$$\begin{aligned}\mathrm{E}&\left[\left(\frac{g_\theta}{g} - I_{\theta\phi}I_{\phi\phi}^{-1}\frac{g_\phi}{g}\right)\left(\frac{g_\theta}{g} - I_{\theta\phi}I_{\phi\phi}^{-1}\frac{g_\phi}{g}\right)'\bigg|\mathbf{y}_1\right] \\ &\geq \mathrm{E}\left[\frac{g_\theta}{g} - I_{\theta\phi}I_{\phi\phi}^{-1}\frac{g_\phi}{g}\bigg|\mathbf{y}_1\right]\mathrm{E}\left[\frac{g_\theta}{g} - I_{\theta\phi}I_{\phi\phi}^{-1}\frac{g_\phi}{g}\bigg|\mathbf{y}_1\right]' \\ &= \left(\frac{f_\theta}{f}\right)\left(\frac{f_\theta}{f}\right)'.\end{aligned} \tag{A3}$$

The last equality holds in view of (A1) and (A2). Note that

$$I_{\theta\theta} - I_{\theta\phi}I_{\phi\phi}^{-1}I_{\phi\theta} = \mathrm{E}\left[\left(\frac{g_\theta}{g} - I_{\theta\phi}I_{\phi\phi}^{-1}\frac{g_\phi}{g}\right)\left(\frac{g_\theta}{g} - I_{\theta\phi}I_{\phi\phi}^{-1}\frac{g_\phi}{g}\right)'\right].$$

Taking expectation of both sides in (A3) in terms of $\mathbf{y}_1$ and using the above equality lead to (2.1). *Q.E.D.*

REFERENCES

1. Y. Fujikoshi, *Ann. Inst. Statist. Math.* **34**, 523–530 (1982).
2. Y. Fujikoshi, P. R. Krishnaiah, and J. Schmidhammer, in: *Advances in Multivariate Statistical Analysis*, A. K. Gupta (Ed.), pp. 45–61. D. Reidel Publishing Company (1987).
3. R. A. Wijsman, *J. Multi. Anal.* **18**, 169–177 (1986).
4. C. R. Rao, in: *Multivariate Analysis*, P. R. Krishnaiah (Ed.), pp. 87–103. Academic Press, New York (1966).
5. T. Ferguson, *Ann. Math. Statist.* **29**, 1046–1062 (1958).
6. C. R. Rao, *Linear Statistical Inference and Its Applications.* 2nd ed. Wiley, New York (1973).
7. A. Shapiro, *South African Statist. J.* **19**, 73–81 (1985).
8. M. W. Browne, *South African Statist. J.* **8**, 1–24 (1974).
9. P. M. Bentler and C. -P. Chou, Paper presented at Psychometric meetings, Los Angeles, CA (1989, July).
10. M. W. Browne, in: *Topics in Applied Multivariate Analysis*, D. M. Hawkins (Ed.), pp. 72–141. Cambridge University Press, Cambridge, England (1982).
11. M. W. Browne, *British J. Math. Statist. Psychol.* **37**, 62–83 (1984).
12. P. M. Bentler and T. Dijkstra, in: *Multivariate Analysis* **VI**, P. R. Krishnaiah (Ed.), pp. 9–42, North Holland, Amsterdam, The Netherlands (1985).
13. P. M. Bentler, *EQS Structural Equations Program Manual.* BMDP Statistical Software, Los Angeles (1989).
14. B. Muthén, *British J. Math. Statist. Psychol.* **42**, 55–62 (1989).
15. M. W. Browne, in: *Statistical Analysis of Measurement Error Models and Applications (Contemporary Mathematics, 112)*, P. J. Brown and W. A. Fuller (Eds.), pp. 211–225. American Mathematical Society, Providence, RI (1990).
16. A. Mooijaart, *Psychometrika* **50**, 323–342 (1985).
17. P. M. Bentler and Y. Kano, *Multi. Behav. Res.* **25**, 67–74 (1990).
18. Y. Kano, *Statist. Probab. Lett.* **2**, 241–244 (1984).
19. Y. Kano, *Ann. Inst. Statist. Math.* **39**, 57–68 (1986).
20. Y. Kano, *British J. Math. Statist. Psychol.* **39**, 221–227 (1986).
21. H. Schneeweiss, in: *Statistical Analysis of Measurement Error Models and Applications (Contemporary Mathematics, 112)*, P. J. Brown and W. A. Fuller (Eds.), pp. 33–40. American Mathematical Society, Providence, RI (1990).
22. J. S. Williams, *Psychometrika* **43**, 293–306 (1978).
23. Y. Kano and A. Shapiro, *South African Statist. J.* **21**, 131–139 (1987).
24. P. M. E. Altham, *J. Roy. Statist. Soc., Ser. B* **46**, 118-119 (1984).
25. P. M. Bentler and A. Mooijaart, *Psychol. Bull.* **106**, 315–317 (1989).

October 5, 1992

To: Peter and Ab,

Our joint paper "Additional Information $\cdots$" has been accepted for publication in the Proceedings Book of the PASC conference held in Japan last year. Thank you very much for your cooperation. Enclosed herewith is a copy of the final version of our paper.

Formally they have not noticed the title of the book, publisher, year, etc. So if you refer to the paper, in reference section, you may write

The Conference Proceedings of the 3rd Pacific Area Statistical Conference, Makuhari, Japan

following the title of the paper. I will inform you of the book title, etc. as soon as I know.

Yutaka