JOURNAL OF MULTIVARIATE ANALYSIS 49, 110–131 (1994)

Asymptotic Properties of the Estimators for Multivariate Components of Variance

SELLEM REMADI AND YASUO AMEMIYA

Iowa State University

Estimation of the covariance matrices in the multivariate balanced one-way random effect model is discussed. The rank of the between-group covariance matrix plays a large role in model building as well as in assessing asymptotic properties of the estimated covariance matrices. The restricted (residual) maximum likelihood estimators derived under a rank condition are considered. Asymptotic properties of the estimators are derived for a possibly incorrectly specified rank and under either the number of groups, the number of replicates, or both, tending to infinity. A higher order expansion covering various cases leads to a common approximate inference procedure which can be used in a wide range of practical situations. A simulation study is also presented. © 1994 Academic Press, Inc.

1. INTRODUCTION

Suppose that a $p \times 1$ observation vector $\mathbf{Y}_{ij}$ taken on the jth individual in the ith group satisfies

$$\mathbf{Y}_{ij} = \boldsymbol{\mu} + \mathbf{b}_i + \mathbf{w}_{ij}, \qquad i = 1, ..., n, \; j = 1, ..., r, \tag{1}$$

where $\boldsymbol{\mu}$ is a $p \times 1$ vector of unknown parameters, the $p \times 1$ $\mathbf{b}_i$ represents the ith group effect, and $\mathbf{w}_{ij}$ is the $p \times 1$ error term assumed to have $N_p(0, \boldsymbol{\Sigma}_{ww})$ distribution. When n groups are assumed to be taken from some population, we often assume that the between-group effects $\mathbf{b}_i$'s are independent $N_p(\mathbf{0}, \boldsymbol{\Sigma}_{bb})$ random vectors distributed independently from the $\mathbf{w}_{ij}$'s. Assume that $\boldsymbol{\Sigma}_{bb}$ is nonnegative definite and $\boldsymbol{\Sigma}_{ww}$ is positive definite, and that $n > 1$ and $r > 1$. This is the multivariate balanced one-way components of variance model.

The univariate components of variance model has been used and discussed extensively in the literature. For reviews, see, e.g., Harville [12], Robinson [21], and Searle *et al.* [23]. A multivariate model such as (1)

Received August 3, 1992; revised May 17, 1993

AMS 1980 subject classification: primary 62H12, 62F12, secondary 62J99.

Key words and phrases: random effect model, estimation under rank constraint, restricted maximum likelihood estimator, covariance matrix rank, common asymptotic expansion.

0047-259X/94 $6.00

containing unknown covariance matrices can be transformed to the general univariate form by stacking the $p \times 1$ response vectors. However, such rewriting may not solve some of the statistical problems for the multivariate case, because of the parameter space restriction and the possibility of singular covariance matrices. Thus, development of statistical procedures for multivariate models often requires approaches slightly different from those used for the univariate case. The literature on the multivariate model has been rather limited. The ordinary or residual maximum likelihood estimation for Model (1) with no restricton on Σ_{bb} was discussed by Klotz and Putter [14], Bock and Petersen [9], and Amemiya [1]. Hill and Thompson [13] and Bhargava and Disch [8] discussed the problem of a possible nonnegative definite estimate of Σ_{bb}. Mathew *et al.* [15] derived estimators of Σ_{bb} which have uniformly smaller values of a certain risk function than the usual unbiased statistic. Estimation under the rank condition and testing for the rank was treated in Anderson [5, 6], Amemiya and Fuller [2], Schott and Saw [22], Anderson *et al.* [4], Amemiya *et al.* [3], Anderson and Amemiya [7], and Remadi and Amemiya [20]. Thompson [24] and Meyer [16] discuss algorithms for computing the restricted (residual) maximum likelihood estimators (REML) of covariance components in the multivariate mixed effect model. For the balanced multivariate random effect model, Calvin and Dykstra [10] proposed a computational algorithm which is guaranteed to converge to the REML. Properties of estimators or inference procedures for functions of covariance components have received virtually no treatment in the literature.

Here we consider estimation of covariance components Σ_{bb} and Σ_{ww} in Model (1). Although Model (1) is the simplest multivariate components of variance model, properties of the estimators have been largely unknown. Consideration of Model (1) highlights some of the common problems for multivariate models and suggests possible extensions to more general multivariate models. In model (1), the between-group effect $\mathbf{b}_i$ is $p \times 1$; i.e., each of the p response variables has one corresponding group effect variable in $\mathbf{b}_i$. But the actual between-group variability can be concentrated in a space of dimension less than p. For example, some of the p variables or some linear combinations may have no between-group differences. Thus, a random effect in the multivariate model can exist with a singular covariance matrix with various values of rank, while a variance component in the univariate model is either zero or positive. As shown later, the true rank of a covariance component also affects properties of an estimated covariance component. In this paper, we consider, for Model (1), properties of estimators of Σ_{bb} and Σ_{ww} obtained under the assumption that rank $\Sigma_{bb} \leqslant m$. In practice, such a rank condition is imposed based on subject matter consideration or based on statistical tests of rank. See, e.g., Amemiya *et al.* [3] and Anderson and Amemiya [7]. Note that any estimator should take values in (the closure of) the parameter space (with probability one), i.e., an estimator of $\boldsymbol{\Sigma}_{bb}$ should be a symmetric nonnegative definite matrix of rank at most m, and an estimator of $\boldsymbol{\Sigma}_{ww}$ should be symmetric positive definite, with probability one. A set of such estimators is the restricted (residual) maximum likelihood estimators derived under the assumption of rank $\boldsymbol{\Sigma}_{bb} \leqslant m$. To present the estimators, let the between-group and within-group mean-square matrices be defined to be

$$\mathbf{m}_{bb} = \frac{r}{n-1} \sum_{i=1}^{n} (\bar{\mathbf{Y}}_{i\cdot} - \bar{\mathbf{Y}}_{\cdot\cdot})(\bar{\mathbf{Y}}_{i\cdot} - \bar{\mathbf{Y}}_{\cdot\cdot})',$$

$$\mathbf{m}_{ww} = \frac{1}{n(r-1)} \sum_{i=1}^{n} \sum_{j=1}^{r} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{i\cdot})(\bar{\mathbf{Y}}_{ij} - \bar{\mathbf{Y}}_{i\cdot})',$$

where $\mathbf{Y}_{i\cdot} = (1/r) \sum_{j=1}^{r} \mathbf{Y}_{ij}$ and $\mathbf{Y}_{\cdot\cdot} = (1/nr) \sum_{i=1}^{n} \sum_{j=1}^{r} \mathbf{Y}_{ij}$. Note that the statistic $1/r(\mathbf{m}_{bb} - \mathbf{m}_{ww})$ is unbiased for $\boldsymbol{\Sigma}_{bb}$, but does not always take values in the parameter space. Let a $p \times p$ orthogonal $\hat{\mathbf{Q}}$ and $p \times p$ diagonal $\hat{\boldsymbol{\Lambda}} = \operatorname{diag}\{\hat{\lambda}_1, ..., \hat{\lambda}_p\}$ be such that

$$\begin{gathered} \mathbf{m}_{ww}^{-1/2} \mathbf{m}_{bb} \mathbf{m}_{ww}^{-1/2} = \hat{\mathbf{Q}} \hat{\boldsymbol{\Lambda}} \hat{\mathbf{Q}}', \\ \hat{\lambda}_1 \geqslant \cdots \geqslant \hat{\lambda}_p. \end{gathered} \tag{2}$$

Define $\hat{k} = \min\{m, \text{number of } \hat{\lambda}_i\text{'s} > 1\}$. This $\hat{k}$ is the rank of the REML estimator $\hat{\boldsymbol{\Sigma}}_{bb}$. We write

$$\hat{\mathbf{P}} = (\hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2) = \mathbf{m}_{ww}^{1/2} \hat{\mathbf{Q}},$$

$$\hat{\boldsymbol{\Lambda}} = \begin{pmatrix} \hat{\boldsymbol{\Lambda}}_1 & 0 \\ 0 & \hat{\boldsymbol{\Lambda}}_2 \end{pmatrix},$$

where $\hat{\mathbf{P}}_1$ is $p \times \hat{k}$, and $\hat{k} \times \hat{k}$ $\hat{\boldsymbol{\Lambda}}_1$ consists of the $\hat{k}$ largest roots $\hat{\lambda}_i$'s. Then the REML estimators of $\boldsymbol{\Sigma}_{bb}$ and $\boldsymbol{\Sigma}_{ww}$ under rank $\boldsymbol{\Sigma}_{bb} \leqslant m$ are, respectively,

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{bb}(m) &= \mathbf{0}, \qquad \text{if } \hat{k} = 0, \\ &= \frac{1}{r} \hat{\mathbf{P}}_1 (\hat{\boldsymbol{\Lambda}}_1 - \mathbf{I}_{\hat{k}}) \hat{\mathbf{P}}_1', \qquad \text{if } \hat{k} > 0, \end{aligned} \tag{3}$$

$$\hat{\boldsymbol{\Sigma}}_{ww}(m) = \frac{1}{nr-1} \{(n-1)[\mathbf{m}_{bb} - r\, \hat{\boldsymbol{\Sigma}}_{bb}] + n(r-1)\, \mathbf{m}_{ww}\}.$$

See, e.g., Anderson [5] and Amemiya and Fuller [2]. Note that $\hat{\boldsymbol{\Sigma}}_{bb}(m)$ is a symmetric nonnegative definite matrix of rank $\hat{k} \leqslant m$ and that $\hat{\boldsymbol{\Sigma}}_{ww}(m)$ is a weighted average of $\mathbf{m}_{ww}$ and a part of $\mathbf{m}_{bb}$ not used for estimating $\boldsymbol{\Sigma}_{bb}$. An alternative form of the REML estimators is

$$\hat{\boldsymbol{\Sigma}}_{bb}(m) = \frac{1}{r}(\mathbf{m}_{bb} - \mathbf{m}_{ww}) - \frac{1}{r}\hat{\boldsymbol{\Omega}}_m,$$
$$\hat{\boldsymbol{\Sigma}}_{ww}(m) = \mathbf{m}_{ww} + \frac{n-1}{nr-1}\hat{\boldsymbol{\Omega}}_m, \tag{4}$$

where

$$\hat{\boldsymbol{\Omega}}_m = \hat{\mathbf{P}}_2(\hat{\boldsymbol{\Lambda}}_2 - \mathbf{I}_{p-\hat{k}})\,\hat{\mathbf{P}}_2'.$$

In this form we see that $\hat{\boldsymbol{\Sigma}}_{bb}(m)$ and $\hat{\boldsymbol{\Sigma}}_{ww}(m)$ are obtained by adjusting the unbiased statistics $(1/r)(\mathbf{m}_{bb} - \mathbf{m}_{ww})$ and $\mathbf{m}_{ww}$ using terms involving $\hat{\boldsymbol{\Omega}}_m$, and that a partition of the total sum of squares holds:

$$\begin{aligned}(n-1)\,r\hat{\boldsymbol{\Sigma}}_{bb}(m) + (nr-1)\,\hat{\boldsymbol{\Sigma}}_{ww}(m) &= (n-1)\,\mathbf{m}_{bb} + n(r-1)\,\mathbf{m}_{ww}\\ &= \sum_{i=1}^{n}\sum_{j=1}^{r}(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{..})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{..})'.\end{aligned}$$

In this paper, we derive asymptotic properties of $\hat{\boldsymbol{\Sigma}}_{bb}(m)$ and $\hat{\boldsymbol{\Sigma}}_{ww}(m)$. Development of asymptotic results for the random effect models requires some special care. See, e.g., Miller [17]. Model (1) contains two indices n and r, i.e., the numbers of groups and replicates. A practical situation may have large n, large r, or large n and r. We develop an asymptotic theory covering any of these cases. Throughout this paper, we refer our different assumptions for asymptotics as

Case I. $n \to \infty$ and r is fixed.

Case II. $n \to \infty$ and $r \to \infty$.

Case III. n is fixed and $r \to \infty$.

Our eventual goal is to develop approximate inference procedures for $\boldsymbol{\Sigma}_{bb}$ and $\boldsymbol{\Sigma}_{ww}$ (or functions of $\boldsymbol{\Sigma}_{bb}$ and $\boldsymbol{\Sigma}_{ww}$) which can be applied in a wide range of practical situations. After developing specific results for each of these cases, we suggest approximate inference procedure which can be used in a situation corresponding to any one of Cases I, II, and III. Another problem associated with developing asymptotic properties of $\hat{\boldsymbol{\Sigma}}_{bb}(m)$ and $\hat{\boldsymbol{\Sigma}}_{ww}(m)$ is their dependency on the true rank m_0 of $\boldsymbol{\Sigma}_{bb}$. Although the estimators are obtained under the assumption that rank $\boldsymbol{\Sigma}_{bb} \leqslant m$, the true rank m_0 is generally unknown. We investigate the effect of not knowing m_0 on the properties of the estimators.

2. CONSISTENCY

To discuss asymptotic properties of $\hat{\boldsymbol{\Sigma}}_{bb}(m)$ and $\hat{\boldsymbol{\Sigma}}_{ww}(m)$, we need to note that for Case III with fixed n, a consistent estimator of $\boldsymbol{\Sigma}_{bb}$ does not exist. Thus, for $\hat{\boldsymbol{\Sigma}}_{bb}(m)$, we discuss the consistency by checking whether or not

$$\hat{\boldsymbol{\Sigma}}_{bb}(m) - \mathbf{S}_{bb} \xrightarrow{P} 0, \tag{5}$$

where

$$\mathbf{S}_{bb} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{b}_i - \bar{\mathbf{b}})(\mathbf{b}_i - \bar{\mathbf{b}})',$$

$$\bar{\mathbf{b}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{b}_i.$$

This is equivalent to the ordinary consistency for Cases I and II with $n \to \infty$. Note that $\mathbf{S}_{bb}$ is unobservable. First, we consider the case with $m \geqslant m_0$; i.e., the case where the maximum allowable rank is larger or equal to the true rank of $\boldsymbol{\Sigma}_{bb}$.

THEOREM 1. *If $m \geqslant m_0$, then for all cases I, II, and III*

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{bb}(m) &= \frac{1}{r}(\mathbf{m}_{bb} - \mathbf{m}_{ww}) + O_p\left(\frac{1}{r\sqrt{n}}\right) \\
&= \mathbf{S}_{bb} + O_p\left(\frac{1}{\sqrt{rn}}\right), \\
\hat{\boldsymbol{\Sigma}}_{ww}(m) &= \mathbf{m}_{ww} + O_p\left(\frac{1}{r\sqrt{n}}\right) \\
&= \boldsymbol{\Sigma}_{ww} + O_p\left(\frac{1}{\sqrt{rn}}\right),
\end{aligned}$$

where $\mathbf{S}_{bb}$ is defined in (5), r is constant for Case I, and n is constant for Case III.

Proof. Note that for all three cases,

$$\begin{aligned}
\frac{1}{r}\mathbf{m}_{bb} - \mathbf{S}_{bb} - \frac{1}{r}\boldsymbol{\Sigma}_{ww} &= O_p\left(\frac{1}{\sqrt{nr}}\right), \\
\mathbf{m}_{ww} - \boldsymbol{\Sigma}_{ww} &= O_p\left(\frac{1}{\sqrt{nr}}\right), \\
\hat{\mathbf{P}} = \mathbf{m}_{ww}^{1/2}\hat{\mathbf{Q}} &= \boldsymbol{\Sigma}_{ww}^{1/2}\hat{\mathbf{Q}} + O_p\left(\frac{1}{\sqrt{nr}}\right) \\
&= O_p(1),
\end{aligned}$$

where we have used the fact that the elements of $\hat{\mathbf{Q}}$ are bounded by one in absolute value. By the result on the limiting distribution of the roots $\hat{\lambda}_i$'s,

$$\sqrt{n}\,(\hat{\lambda}_i - 1) = O_p(1), \qquad i = m_0 + 1, ..., p. \tag{7}$$

See, e.g., Remadi and Amemiya [19]. Note that $\hat{\mathbf{\Omega}}_m$ in (4) is a function of $\hat{\lambda}_i$, $i = m+1, ..., p$, with $m \geqslant m_0$. It follows from (6) and (7) that for all three cases

$$\hat{\mathbf{\Omega}}_m = O_p\left(\frac{1}{\sqrt{n}}\right).$$

Thus the result follows from (6), (9), and (11). ∎

Hence, with the understanding of the consistency of $\hat{\mathbf{\Sigma}}_{bb}(m)$ as given in (5) for Case III, $\hat{\mathbf{\Sigma}}_{bb}(m)$ and $\hat{\mathbf{\Sigma}}_{ww}(m)$ are consistent for $\mathbf{\Sigma}_{bb}$ and $\mathbf{\Sigma}_{ww}$, provided that $m \geqslant m_0$. Thus, for example, $\hat{\mathbf{\Sigma}}_{bb}(p)$ and $\hat{\mathbf{\Sigma}}_{ww}(p)$ obtained under no rank condition of $\mathbf{\Sigma}_{bb}$ are always consistent.

To discuss the case $m < m_0$, define the $p \times p$ matrix

$$\mathbf{\Psi} = \begin{cases} \mathbf{\Sigma}_{ww}^{-1/2} \mathbf{\Sigma}_{bb} \mathbf{\Sigma}_{ww}^{-1/2}, & \text{for Cases I and II,} \\ \mathbf{\Sigma}_{ww}^{-1/2} \mathbf{S}_{bb} \mathbf{\Sigma}_{ww}^{-1/2}, & \text{for Case III.} \end{cases}$$

Note that $\mathbf{S}_{bb}$ is a random matrix defined in (5). Let $v_1 \geqslant \cdots \geqslant v_p$ be the eigenvalues of $\mathbf{\Psi}$, and let $\mathbf{Q}_0$ be the $p \times (m_0 - m)$ matrix of a set of eigenvectors corresponding to $v_{m+1}, ..., v_{m_0}$. For Case III, v_i's and $\mathbf{Q}_0$ are also random. By Okamoto [18], for Case III, v_i's are distinct with probability one. For simplicity, we assume for Cases I and II that $v_m > v_{m+1}$. Now we present the following result on the consistency when $m < m_0$.

THEOREM 2. *If $m < m_0$, then for all cases I, II, and III,*

$$\hat{\mathbf{\Sigma}}_{bb}(m) - \mathbf{S}_{bb} \xrightarrow{P} -\mathbf{B}$$

$$\hat{\mathbf{\Sigma}}_{ww}(m) - \mathbf{\Sigma}_{ww} \xrightarrow{P} \begin{cases} \mathbf{B}, & \textit{for Cases I and II,} \\ \dfrac{n-1}{n}\mathbf{B}, & \textit{for Case III,} \end{cases}$$

where

$$\mathbf{B} = \mathbf{\Sigma}_{ww}^{1/2} \mathbf{Q}_0 \operatorname{diag}\{v_{m+1}, ..., v_{m_0}\} \mathbf{Q}_0' \mathbf{\Sigma}_{ww}^{1/2}.$$

Proof. It follows from (6) that for all cases I, II, and III,

$$\frac{1}{r} \mathbf{m}_{ww}^{-1/2} \mathbf{m}_{bb} \mathbf{m}_{ww}^{-1/2} \xrightarrow{\text{a.s.}} \mathbf{\Psi}_0, \tag{8}$$

where

$$\boldsymbol{\Psi}_0 = \begin{cases} \boldsymbol{\Psi} + \frac{1}{r}\mathbf{I}_p, & \text{for Case I,} \\ \boldsymbol{\Psi}, & \text{for Cases II and III.} \end{cases}$$

Thus, by the continuity of the eigenvalues,

$$\frac{1}{r}\hat{\lambda}_i \xrightarrow{\text{a.s.}} v_i^0, \qquad i = 1, 2, ..., p, \tag{9}$$

where $v_i^0 = v_i + 1/r$ for Case I and $v_i^0 = v_i$ for Cases II and III. Let $\hat{\mathbf{Q}}_0$ be the $(m+1)$-st through m_0-th columns of $\hat{\mathbf{Q}}$. Since the elements of $\hat{\mathbf{Q}}_0$ are bounded, with probability one, every subsequence has a converging subsubsequence. By (8) and (9) over such a converging subsubsequence $\hat{\mathbf{Q}}_l$ with a limit $\mathbf{Q}_0^*$

$$\mathbf{0} = \frac{1}{r}\mathbf{m}_{ww}^{-1/2}\mathbf{m}_{bb}\mathbf{m}_{ww}^{-1/2}\hat{\mathbf{Q}}_l - \hat{\mathbf{Q}}_l \frac{1}{r}\operatorname{diag}\{\hat{\lambda}_{m+1}, ..., \hat{\lambda}_{m_0}\}$$

$$\rightarrow \boldsymbol{\Psi}_0\mathbf{Q}_0^* - \mathbf{Q}_0^* \operatorname{diag}\{v_{m+1}^0, ..., v_{m_0}^0\}.$$

Because $v_m^0 > v_{m+1}^0$ and $v_{m_0}^0 > v_{m_0+1}^0$ for Cases I and II, and for Case III with probability one, $\mathbf{Q}_0^*$ is unique up to orthogonal rotation of each eigenspace of $\boldsymbol{\Psi}_0$ corresponding to $v_{m+1}^0, ..., v_{m_0}^0$. Since $\mathbf{Q}_0^*\mathbf{Q}_0^{*\prime}$ and $\mathbf{Q}_0^* \operatorname{diag}\{v_{m+1}^0, ..., v_{m_0}^0\}\,\mathbf{Q}_0^{*\prime}$ are unique under such orthogonal rotations, and equal to $\mathbf{Q}_0\mathbf{Q}_0'$ and $\mathbf{Q}_0 \operatorname{diag}\{v_{m+1}^0, ..., v_{m_0}^0\}\,\mathbf{Q}_0'$, it follows that for all three cases

$$\mathbf{R}_m = \hat{\mathbf{Q}}_0 \frac{1}{r}\operatorname{diag}\{\hat{\lambda}_{m+1} - 1, ..., \hat{\lambda}_{m_0} - 1\}\,\hat{\mathbf{Q}}_0' \xrightarrow{\text{a.s.}} \boldsymbol{\Sigma}_{ww}^{-1/2}\mathbf{B}\boldsymbol{\Sigma}_{ww}^{-1/2}. \tag{10}$$

Since $v_{m_0}^0 > 1$ for Case I, and since $r \rightarrow \infty$ for Cases II and III, (9) implies that for all three cases

$$P\{\hat{\lambda}_{m_0} > 1\} \rightarrow 1. \tag{11}$$

For $m < m_0$

$$P\{\hat{k} = m\} \rightarrow 1.$$

Hence, using the form (4), we can write with probability approaching one,

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{bb}(m) &= \hat{\boldsymbol{\Sigma}}_{bb}^0(m_0) - \mathbf{m}_{ww}^{1/2}\mathbf{R}_m\mathbf{m}_{ww}^{1/2}, \\ \hat{\boldsymbol{\Sigma}}_{ww}(m) &= \hat{\boldsymbol{\Sigma}}_{ww}^0(m_0) + \frac{(n-1)r}{nr-1}\mathbf{m}_{ww}^{1/2}\mathbf{R}_m\mathbf{m}_{ww}^{1/2}, \end{aligned} \tag{12}$$

where $\hat{\Sigma}^0_{bb}(m_0)$ and $\hat{\Sigma}^0_{ww}(m_0)$ are $\hat{\Sigma}_{bb}(m_0)$ and $\hat{\Sigma}_{ww}(m_0)$ with $\hat{k}=m_0$, and $\mathbf{R}_m$ is defined in (10). Thus, the result follows from (10), (12), and Theorem 1. ∎

Hence, $\hat{\Sigma}_{bb}(m)$ and $\hat{\Sigma}_{ww}(m)$ are not consistent when the specified maximum rank m for Σ_{bb} is smaller than the true rank m_0. Since the matrix $\mathbf{B}$ in Theorem 2 is nonnegative definite, $\hat{\Sigma}_{bb}(m)$ "underestimates" Σ_{bb} and $\hat{\Sigma}_{ww}(m)$ "overestimates" Σ_{ww}. Thus, it is important not to underspecify the rank of Σ_{bb} in estimation of Σ_{bb} and Σ_{ww}.

3. Limiting Distribution

By Theorem 2, if $m<m_0$, $\hat{\Sigma}_{bb}(m)$ and $\hat{\Sigma}_{ww}(m)$ are inconsistent in the sense given in the theorem. For such a case, a limiting distribution result useful for asymptotic inferences cannot be found easily. Hence, we consider only the case with $m \geqslant m_0$, i.e., where the true rank is less than or equal to the assumed maximum rank. For a $p \times p$ symmetric matrix $\mathbf{A}$, we use the notation vech $\mathbf{A}$, a $p(p+1)/2 \times 1$ vector containing the elements on and below the diagonal of $\mathbf{A}$ starting with the first column. For a $p \times p$ symmetric matrix $\mathbf{A}$, there is a unique $p^2 \times p(p+1)/2$ matrix $\mathbf{K}_p$ such that $\text{vec}\,\mathbf{A} = \mathbf{K}_p\,\text{vech}\,\mathbf{A}$, where vec $\mathbf{A}$ is the $p^2 \times 1$ vector listing the elements of the columns of $\mathbf{A}$ starting with the first. For any such $\mathbf{A}$ we write

$$\boldsymbol{\Gamma}(\mathbf{A}) = 2\mathbf{K}_p^{+}(\mathbf{A} \otimes \mathbf{A})\,\mathbf{K}_p^{+\prime}, \tag{13}$$

where $\mathbf{K}_p^{+} = (\mathbf{K}_p'\mathbf{K}_p)^{-1}\,\mathbf{K}_p$ and $\otimes$ is the Kronecker product. Note that for $\mathbf{A}=(a_{ij})$, a typical element of $\boldsymbol{\Gamma}(\mathbf{A})$ is $a_{ik}a_{jl}+a_{il}a_{jk}$.

If $m \geqslant m_0$ the limiting distributions for Cases II and III are relatively simple and are given in the following theorem.

Theorem 3. *Suppose that $m \geqslant m_0$. For Case II,*

$$\left\{ \begin{matrix} \sqrt{n}\,\text{vech}(\hat{\Sigma}_{bb}(m) - \Sigma_{bb}) \\ \sqrt{n(r-1)}\,\text{vech}(\hat{\Sigma}_{ww}(m) - \Sigma_{ww}) \end{matrix} \right\} \xrightarrow{L} N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Gamma}(\Sigma_{bb}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}(\Sigma_{ww}) \end{pmatrix} \right\}.$$

For Case III,

$$\left\{ \begin{matrix} \text{vech}\,\hat{\Sigma}_{bb}(m) \\ \sqrt{n(r-1)}\,\text{vech}(\hat{\Sigma}_{ww}(m) - \Sigma_{ww}) \end{matrix} \right\} \xrightarrow{L} \left\{ \begin{matrix} \text{vech}\,\mathbf{S}_{bb} \\ \mathbf{W} \end{matrix} \right\},$$

where $\mathbf{S}_{bb}$ *and* $\mathbf{W}$ *are independent,* $(r-1)\mathbf{S}_{bb} \sim \mathbf{W}_p(\boldsymbol{\Sigma}_{bb}, n-1)$, *and* $\mathbf{W} \sim N(\mathbf{0}, \boldsymbol{\Gamma}(\boldsymbol{\Sigma}_{ww}))$.

Proof. The results follow immediately from Theorem 1. ∎

Note that the limiting distribution of $\hat{\boldsymbol{\Sigma}}_{bb}(m)$ and $\hat{\boldsymbol{\Sigma}}_{ww}(m)$ in Theorem 3 is that of $\mathbf{S}_{bb}$ and m_{ww}. Thus, for Cases II and III with $r \to \infty$, the limiting distributions are simple in that the rather complex nature of the rank restriction and interrelationship disappears in the limit. Also, the form of these limiting distributions is the same for all $m \geqslant m_0$. In this sense, the limiting distribution in Theorem 3 may be considered too optimistic in practice. In Section 4, a higher order asymptotic expansion is considered.

For Case I, we need to distinguish two cases, $m = m_0$ and $m > m_0$. For rank $\boldsymbol{\Sigma}_{bb} = m_0 < p$, let C be a $p \times (p - m_0)$ matrix of rank $(p - m_0)$ such that $\mathbf{C}'\boldsymbol{\Sigma}_{bb}\mathbf{C} = \mathbf{0}$. Define

$$\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_{ww}\mathbf{C}(\mathbf{C}'\boldsymbol{\Sigma}_{ww}\mathbf{C})^{-1}\mathbf{C}'\boldsymbol{\Sigma}_{ww}. \tag{14}$$

Note that $\mathbf{C}$ is not unique but $\boldsymbol{\Sigma}_0$ is free of the choice of $\mathbf{C}$. If $m_0 = p$, $\boldsymbol{\Sigma}_0$ is understood to be zero.

THEOREM 4. *Consider Case* I. *If* $m = m_0$, *then*

$$\left\{ \begin{matrix} \sqrt{n}\,\operatorname{vech}(\hat{\boldsymbol{\Sigma}}_{bb}(m_0) - \boldsymbol{\Sigma}_{bb}) \\ \sqrt{n(r-1)}\,\operatorname{vech}(\hat{\boldsymbol{\Sigma}}_{ww}(m_0) - \boldsymbol{\Sigma}_{ww}) \end{matrix} \right\} \xrightarrow{L} N\left\{ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V}^{\mathrm{I}}_{bb} & \mathbf{V}^{\mathrm{I}}_{bw} \\ \mathbf{V}^{\mathrm{I}}_{wb} & \mathbf{V}^{\mathrm{I}}_{ww} \end{pmatrix} \right\},$$

where

$$\mathbf{V}^{\mathrm{I}}_{bb} = \boldsymbol{\Gamma}\left(\boldsymbol{\Sigma}_{bb} + \frac{1}{r}\boldsymbol{\Sigma}_{ww}\right) + \frac{1}{r-1}\boldsymbol{\Gamma}\left(\frac{1}{r}\boldsymbol{\Sigma}_{ww}\right) - \frac{1}{r(r-1)}\boldsymbol{\Gamma}(\boldsymbol{\Sigma}_0),$$

$$\mathbf{V}^{\mathrm{I}}_{bw} = \frac{1}{r\sqrt{r-1}}\left[\boldsymbol{\Gamma}(\boldsymbol{\Sigma}_0) - \boldsymbol{\Gamma}(\boldsymbol{\Sigma}_{ww})\right],$$

$$\mathbf{V}^{\mathrm{I}}_{ww} = \boldsymbol{\Gamma}(\boldsymbol{\Sigma}_{ww}) - \frac{1}{r}\boldsymbol{\Gamma}(\boldsymbol{\Sigma}_0).$$

If $m > m_0$, *the limiting distribution of*

$$\left\{ \begin{matrix} \sqrt{n}\,\operatorname{vech}(\hat{\boldsymbol{\Sigma}}_{bb}(m) - \boldsymbol{\Sigma}_{bb}) \\ \sqrt{n(r-1)}\,\operatorname{vech}(\hat{\boldsymbol{\Sigma}}_{ww}(m) - \boldsymbol{\Sigma}_{ww}) \end{matrix} \right\}$$

does not exist if $m < p$ *and is not normal if* $m = p$.

Proof. The result for $m = m_0$ is a consequence of the expansion given in Theorem 6 of the next section. If $m > m_0$, for example, for $\hat{\Sigma}_{bb}$,

$$\sqrt{n}\,(\hat{\Sigma}_{bb}(m) - \Sigma_{bb}) = \sqrt{n}\,(\hat{\Sigma}^0_{bb}(m_0) - \Sigma_{bb}) + \frac{1}{r}\,\mathbf{m}^{1/2}_{ww}\sqrt{n}\,\mathbf{R}\mathbf{m}^{1/2}_{ww}, \quad (15)$$

where $\hat{\Sigma}^0_{bb}(m_0)$is $\hat{\Sigma}_{bb}(m_0)$ with $\hat{k} = m_0$,

$$\mathbf{R} = \sum_{i=m_0+1}^{m} (\hat{\lambda}_i - 1)\, I(\hat{\lambda}_i > 1)\, \hat{\mathbf{q}}_i \hat{\mathbf{q}}_i',$$

and $\hat{\mathbf{q}}_i$ is the ith column of $\hat{\mathbf{Q}}$. Since $\hat{\Sigma}^0_{bb}(m_0)$ is a function of $\hat{\lambda}_i$, $i = 1, 2, ..., m_0$, and since $\mathbf{m}_{ww} \xrightarrow{P} \Sigma_{ww}$, the two terms in (15) are independent in the limit. By the first part of this theorem the first term converges to a normal distribution. Thus for the sum to have a limiting normal distribution, the second term must have a limiting normal distribution. See, e.g., Feller [11, p. 525, Cramer–Levy theorem]. Note that

$$\sqrt{n}\,(\hat{\lambda}_i - 1)\, I(\sqrt{n}\,(\hat{\lambda}_i - 1) > 0), \qquad i = m_0 + 1, ..., m,$$

converge to nonnegative random variables. Thus, if $\sqrt{n}\,\mathbf{R}$ has a nondegenerate limiting distribution, then it is not normal. For $m < p$, following the argument used in the proof of Theorem 2, the limit of a converging subsequence of $\sqrt{n}\,\mathbf{R}$ depends on the subsequence, because the limiting $\hat{\mathbf{q}}_i$'s, $i = m_0 + 1, ..., p$, span the eigenspace of dimension $(p - m_0)$ corresponding to the unit root of $\Sigma^{-1/2}_{ww}(r\Sigma_{bb} + \Sigma_{ww})\,\Sigma^{-1/2}_{ww}$. Hence, for $m < p$, $\sqrt{n}\,(\hat{\Sigma}_{bb}(m) - \Sigma_{bb})$ does not have a limiting distribution. ∎

Thus, for Case I, with $m > m_0$, the estimators are consistent but do not have a limiting normal distribution. The discrepancies among the limiting distributions for Cases I, II, and III as given in Theorems 3 and 4 show that the use of the limiting result for asymptotic inferences requires some special care in practice. It may not be apparent which of the three cases is most appropriate for a given situation.

4. Asymptotic Approximation

A possible approach to developing approximate inference procedure useful for all cases I, II, and III is to consider a common asymptotic expansion. To this end, we assume that $m = m_0$, and first derive an expansion for each of the three cases which is of higher order than that given in Theorem 1. We recall the definition of the $p \times (p - m_0)$ matrix $\mathbf{C}$ in (14).

Let $\mathbf{D}$ be a $p \times m_0$ matrix of rank m_0 satisfying that $(\mathbf{C}, \mathbf{D})$ is a $p \times p$ nonsingular matrix. Note that $\mathbf{D}$ is unique only up to multiplication of a $m_0 \times m_0$ nonsingular matrix from the right, and that $\mathbf{D}'\mathbf{\Sigma}_{bb}\mathbf{D}$ is nonsingular. Define

$$\begin{aligned}
\mathbf{S}_{xx} &= \mathbf{D}'\mathbf{S}_{bb}\,\mathbf{D},\\
\mathbf{S}_{vv} &= \mathbf{C}'\frac{r}{n-1}\sum_{i=1}^{n}(\bar{\mathbf{w}}_{i\cdot}-\bar{\mathbf{w}}_{\cdot\cdot})(\bar{\mathbf{w}}_{i\cdot}-\bar{\mathbf{w}}_{\cdot\cdot})'\,\mathbf{C},\\
\mathbf{S}_{vx} &= \mathbf{C}'\frac{\sqrt{r}}{n-1}\sum_{i=1}^{n}(\bar{\mathbf{w}}_{i\cdot}-\bar{\mathbf{w}}_{\cdot\cdot})(\mathbf{b}_i-\bar{\mathbf{b}}_{\cdot})'\,\mathbf{D},\\
\mathbf{S}_{ee} &= \mathbf{C}'\frac{1}{n(r-1)}\sum_{i=1}^{n}\sum_{j=1}^{r}(\mathbf{w}_{ij}-\bar{\mathbf{w}}_{i\cdot})(\bar{\mathbf{w}}_{ij}-\bar{\mathbf{w}}_{i\cdot})'\,\mathbf{C},
\end{aligned} \tag{16}$$

where

$$\bar{\mathbf{w}}_{i\cdot} = \frac{1}{r}\sum_{i=1}^{r}\mathbf{w}_{ij},\ \bar{\mathbf{w}}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{w}}_{i\cdot},\ \text{and}\ \bar{\mathbf{b}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{b}_i.$$

Let

$$\mathbf{\Omega}_{\mathrm{I}} = \mathbf{\Sigma}_{ww}\mathbf{C}(\mathbf{C}'\mathbf{\Sigma}_{ww}\mathbf{C})^{-1}(\mathbf{S}_{vv}-\mathbf{S}_{ee})(\mathbf{C}'\mathbf{\Sigma}_{ww}\mathbf{C})^{-1}\mathbf{C}'\mathbf{\Sigma}_{ww},$$

$$\mathbf{\Omega}_{\mathrm{II}} = \mathbf{\Sigma}_{ww}\mathbf{C}(\mathbf{C}'\mathbf{\Sigma}_{ww}\mathbf{C})^{-1}(\mathbf{S}_{vv\cdot x}-\mathbf{S}_{ee})(\mathbf{C}'\mathbf{\Sigma}_{ww}\mathbf{C})^{-1}\mathbf{C}'\mathbf{\Sigma}_{ww},$$

where

$$\mathbf{S}_{vv\cdot x} = \mathbf{S}_{vv} - \mathbf{S}_{vx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{vx}'.$$

Note that $\mathbf{\Omega}_{\mathrm{I}}$ and $\mathbf{\Omega}_{\mathrm{II}}$ are invariant for different choices of $\mathbf{C}$ and $\mathbf{D}$.

THEOREM 5. *Suppose that $m = m_0 < p$. Then, for Case I,*

$$\hat{\mathbf{\Sigma}}_{bb}(m_0) - \mathbf{\Sigma}_{bb} = \frac{1}{r}[\mathbf{m}_{bb} - (r\mathbf{\Sigma}_{bb} + \mathbf{\Sigma}_{ww})] - \frac{1}{r}(\mathbf{m}_{ww} - \mathbf{\Sigma}_{ww}) - \frac{1}{r}\mathbf{\Omega}_{\mathrm{I}} + O_p\left(\frac{1}{n}\right),$$

$$\hat{\mathbf{\Sigma}}_{ww}(m_0) - \mathbf{\Sigma}_{ww} = \mathbf{m}_{ww} - \mathbf{\Sigma}_{ww} + \frac{n-1}{nr-1}\mathbf{\Omega}_{\mathrm{I}} + O_p\left(\frac{1}{n}\right).$$

For Case II,

$$\hat{\boldsymbol{\Sigma}}_{bb}(m_0)-\boldsymbol{\Sigma}_{bb}=\frac{1}{r}\left[\mathbf{m}_{bb}-(r\boldsymbol{\Sigma}_{bb}+\boldsymbol{\Sigma}_{ww})\right]$$

$$-\frac{1}{r}(\mathbf{m}_{ww}-\boldsymbol{\Sigma}_{ww})-\frac{1}{r}\boldsymbol{\Omega}_{\text{II}}+O_p\left(\frac{1}{nr\sqrt{r}}\right),$$

$$\hat{\boldsymbol{\Sigma}}_{ww}(m_0)-\boldsymbol{\Sigma}_{ww}=\mathbf{m}_{ww}-\boldsymbol{\Sigma}_{ww}+\frac{n-1}{nr-1}\boldsymbol{\Omega}_{\text{II}}+O_p\left(\frac{1}{nr\sqrt{r}}\right).$$

For Case III,

$$\hat{\boldsymbol{\Sigma}}_{bb}(m_0)-\boldsymbol{\Sigma}_{bb}=\frac{1}{r}\left[\mathbf{m}_{bb}-(r\boldsymbol{\Sigma}_{bb}+\boldsymbol{\Sigma}_{ww})\right]-\frac{1}{r}\boldsymbol{\Omega}_{\text{II}}+O_p\left(\frac{1}{r\sqrt{r}}\right),$$

$$\hat{\boldsymbol{\Sigma}}_{ww}(m_0)-\boldsymbol{\Sigma}_{ww}=\mathbf{m}_{ww}-\boldsymbol{\Sigma}_{ww}+\frac{n-1}{nr-1}\boldsymbol{\Omega}_{\text{II}}+O_p\left(\frac{1}{r\sqrt{r}}\right).$$

The proof of Theorem 5 is given in the Appendix. These three expansions are meaningful in the sense that every term explicitly given is of order larger than the remainder. All explicit terms for Case I are $O_p(1/\sqrt{n})$. But, for Cases II and III, terms have different order, representing higher order expansions. Since the expansions for the three cases are similar, an expression valid for all cases can be derived.

THEOREM 6. *If $m=m_0$, then for all Cases I, II,* and *III,*

$$\hat{\boldsymbol{\Sigma}}_{bb}(m_0)-\boldsymbol{\Sigma}_{bb}=\frac{1}{r}\left[\mathbf{m}_{bb}-(r\boldsymbol{\Sigma}_{bb}+\boldsymbol{\Sigma}_{ww})\right]$$

$$-\frac{1}{r}(\mathbf{m}_{ww}-\boldsymbol{\Sigma}_{ww})-\frac{1}{r}\boldsymbol{\Omega}_{\text{II}}+O_p\left(\frac{1}{nr\sqrt{r}}\right),$$

$$\hat{\boldsymbol{\Sigma}}_{ww}(m_0)-\boldsymbol{\Sigma}_{ww}=\mathbf{m}_{ww}-\boldsymbol{\Sigma}_{ww}+\frac{n-1}{nr-1}\boldsymbol{\Omega}_{\text{II}}+O_p\left(\frac{1}{nr\sqrt{r}}\right),$$

where it is understood that n is constant for Case III, r is constant for Case I, and that $\boldsymbol{\Omega}_{\text{II}}=0$ for $m_0=p$.

Proof. Note that for Case I,

$$\mathbf{S}_{vv}-\mathbf{S}_{vv\cdot x}=\mathbf{S}_{vx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{vx}'=O_p\left(\frac{1}{n}\right),$$

and thus $\mathbf{\Omega}_{\mathrm{I}} - \mathbf{\Omega}_{\mathrm{II}} = O_p(1/n)$. For Case III,

$$\frac{1}{r}(\mathbf{m}_{ww} - \mathbf{\Sigma}_{ww}) = O_p\left(\frac{1}{r\sqrt{r}}\right).$$

Thus, the result follows. ▮

The common expansion given in Theorem 6 is in fact the one for Case II given in Theorem 5. For Cases I and III, this common expansion simply adds extra terms of the same order as the remainder. This expansion also highlights some characteristics of the estimators derived under the rank condition much better than the expansion in Theorem 1. The adjustment or improvement made to the naive estimators $(1/r)(\mathbf{m}_{bb} - \mathbf{m}_{ww})$ and $\mathbf{m}_{ww}$ is given in terms of $\mathbf{\Omega}_{\mathrm{II}}$. The term $\mathbf{\Omega}_{\mathrm{II}}$ can be characterized to be a part of $(\mathbf{m}_{bb} - \mathbf{m}_{ww})$ estimating the error variability, not the between-group variability. This term is subtracted from $(1/r)(\mathbf{m}_{bb} - \mathbf{m}_{ww})$ for an efficient estimator $\hat{\mathbf{\Sigma}}_{bb}(m)$, and is pooled with $\mathbf{m}_{ww}$ for an improved estimator of $\mathbf{\Sigma}_{ww}$. This expansion provides a means for obtaining an approximate inference procedure which works for a wide range of practical situations.

5. Approximate Inference Procedures

We develop approximate inference procedures for functions of the elements of $\mathbf{\Sigma}_{bb}$ or functions of the elements of $\mathbf{\Sigma}_{bb}$ and $\mathbf{\Sigma}_{ww}$. Typical examples are a linear combination of elements of $\mathbf{\Sigma}_{bb}$, a between-group correlation (a correlation computed from $\mathbf{\Sigma}_{bb}$), and an intra-group correlation (a diagonal element of $\mathbf{\Sigma}_{bb}$ divided by the sum of diagonal elements of $\mathbf{\Sigma}_{bb}$ and $\mathbf{\Sigma}_{ww}$). In practice, it may be difficult to decide which of the three cases I, II, and III is most appropriate for a particular situation. Thus, our goal here is to develop procedures useful for various situations. From the results in Sections 2 and 3, we note that without some knowledge of the rank of $\mathbf{\Sigma}_{bb}$ inference procedures can be incorrect, especially for Case I. Thus, if the rank is unknown we suggest performing some statistical inference for the rank. See Anderson [6], Amemiya *et al.* [3], Anderson and Amemiya [7], and Remadi and Amemiya [20]. Here we assume that some idea about m_0, the true rank of $\mathbf{\Sigma}_{bb}$, is obtained so that $\hat{\mathbf{\Sigma}}_{bb}(m_0)$ and $\hat{\mathbf{\Sigma}}_{ww}(m_0)$ can be used at least with large enough probability. First we consider the covariance matrix of the terms in the common expansion in Theorem 6 as a common approximate covariance matrix. Taking the covariance matrix of the expansion terms and ignoring the remainder, the approximate covariance matrix of

$$\begin{Bmatrix} \operatorname{vech} \hat{\boldsymbol{\Sigma}}_{bb}(m_0) \\ \operatorname{vech} \hat{\boldsymbol{\Sigma}}_{ww}(m_0) \end{Bmatrix}$$

is

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{bb} & \mathbf{V}_{bw} \\ \mathbf{V}_{wb} & \mathbf{V}_{ww} \end{pmatrix},$$

where

$$\begin{aligned}
\mathbf{V}_{bb} &= \frac{1}{n-1} \boldsymbol{\Gamma}\left(\boldsymbol{\Sigma}_{bb} + \frac{1}{r}\boldsymbol{\Sigma}_{ww}\right) + \frac{1}{n(r-1)} \boldsymbol{\Gamma}\left(\frac{1}{r}\boldsymbol{\Sigma}_{ww}\right) \\
&\quad - \frac{1}{r^2}\left(\frac{1}{n-1} + \frac{1}{n(r-1)} + \frac{m_0}{(n-1)^2}\right) \boldsymbol{\Gamma}(\boldsymbol{\Sigma}_0), \\
\mathbf{V}_{bw} &= \frac{1}{r}\left(\frac{1}{n(r-1)} + \frac{m_0}{(n-1)(nr-1)}\right) \boldsymbol{\Gamma}(\boldsymbol{\Sigma}_0) - \frac{1}{rn(r-1)} \boldsymbol{\Gamma}(\boldsymbol{\Sigma}_{ww}), \\
\mathbf{V}_{ww} &= \frac{1}{n(r-1)} \boldsymbol{\Gamma}(\boldsymbol{\Sigma}_{ww}) - \left(\frac{n-1}{n(r-1)(nr-1)} + \frac{m_0}{(nr-1)^2}\right) \boldsymbol{\Gamma}(\boldsymbol{\Sigma}_0),
\end{aligned}$$

and the $\boldsymbol{\Gamma}(\cdot)$ function and $\boldsymbol{\Sigma}_0$ are defined in (13) and (14), respectively. The unknown matrices $\boldsymbol{\Sigma}_{bb}$ and $\boldsymbol{\Sigma}_{ww}$ in $\mathbf{V}$ can be estimated by $\hat{\boldsymbol{\Sigma}}_{bb}(m_0)$ and $\hat{\boldsymbol{\Sigma}}_{bb}(m_0)$. To estimate $\boldsymbol{\Sigma}_0$, we consider any $p \times (p-\hat{k})\, \hat{\mathbf{C}}$ such that $\hat{\mathbf{C}}'\hat{\boldsymbol{\Sigma}}_{bb}(m_0)\,\hat{\mathbf{C}} = 0$. Recall that rank $\hat{\boldsymbol{\Sigma}}_{bb}(m_0) = \hat{k}$. By (17) in the Appendix, $\mathbf{P}\{\hat{k} = m_0\} \to 1$ for all three cases. Thus, for all three cases,

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_0 &= \hat{\boldsymbol{\Sigma}}_{ww}(m_0)\,\hat{\mathbf{C}}(\hat{\mathbf{C}}'\hat{\boldsymbol{\Sigma}}_{ww}(m_0)\,\hat{\mathbf{C}})^{-1}\,\hat{\mathbf{C}}'\hat{\boldsymbol{\Sigma}}_{ww}(m_0) \\
&\xrightarrow{\mathbf{P}} \boldsymbol{\Sigma}_0.
\end{aligned}$$

Hence, we have an estimated covariance matrix $\hat{\mathbf{V}}$ obtained by evaluating $\mathbf{V}$ at $\hat{\boldsymbol{\Sigma}}_{bb}(m_0)$, $\hat{\boldsymbol{\Sigma}}_{ww}(m_0)$, and $\hat{\boldsymbol{\Sigma}}_0$. We suggest the use of $\hat{\boldsymbol{\Sigma}}_{bb}(m)$, $\boldsymbol{\Sigma}_{ww}(m)$, $\hat{\mathbf{V}}$, the standard normal cut-off points, and possibly the delta method in approximate inference for functions of the elements of $\boldsymbol{\Sigma}_{bb}$ and $\boldsymbol{\Sigma}_{ww}$. As can be seen from Theorems 3, 4, and 6, this procedure is asymptotically justified for Cases I and II, if some knowledge of m_0 is available. For Case III, the normal distribution based inference is not exactly appropriate. But, for many functions, Wishart-based inference is difficult. By taking into account the higher order terms in the common expansion of Theorem 6, our normal approximation is expected to be practically adequate even for relatively small n.

6. Simulation

A simulation study was conducted to assess finite sample properties of the asymptotic inference procedures developed in Sections 2–5. We considered Model (1) with $p=4$ and normally distributed $\mathbf{b}_i$ and $\mathbf{w}_{ij}$. We set $\boldsymbol{\mu}=\mathbf{0}$, but $\boldsymbol{\mu}$ was estimated. For the sample configuration (n, r), we chose three sets, (50, 5), (50, 50), and (5, 50), loosely corresponding to Cases I, II, and III used to develop the asymptotic theory. Note that in many applications $n=50$ is not necessarily considered very large and $r=50$ is unusually large. For the covariance components $\boldsymbol{\Sigma}_{bb}$ and $\boldsymbol{\Sigma}_{ww}$, we considered two parameterizations both of which have rank $\boldsymbol{\Sigma}_{bb}=m_0=2$.

$$\text{(i)}\quad \boldsymbol{\Sigma}_{bb}^{(i)}=\begin{pmatrix}1 & 1 & 0.5 & 0.5\\ 1 & 1 & 0.5 & 0.5\\ 0.5 & 0.5 & 1 & 1\\ 0.5 & 0.5 & 1 & 1\end{pmatrix},\qquad \boldsymbol{\Sigma}_{ww}^{(i)}=\begin{pmatrix}1 & 0.25 & 0.25 & 0.25\\ 0.25 & 1 & 1 & 0.25\\ 0.25 & 0.25 & 1 & 0.25\\ 0.25 & 0.25 & 0.25 & 1\end{pmatrix},$$

$$\text{(ii)}\quad \boldsymbol{\Sigma}_{bb}^{(ii)}=\boldsymbol{\Sigma}_{bb}^{(i)},\qquad \boldsymbol{\Sigma}_{ww}^{(ii)}=30\boldsymbol{\Sigma}_{ww}^{(i)}.$$

One way to characterize the parameter sets is to consider the roots of $|\boldsymbol{\Sigma}_{bb}-\gamma\boldsymbol{\Sigma}_{ww}|=0$. For (i), $\gamma=1.714$, 1.333, 0, 0, and for (ii), $\gamma=0.057$, 0.044, 0, 0. Thus (ii) can be considered to be a case where $\boldsymbol{\Sigma}_{bb}$ of rank 2 is relatively close to a rank 1 matrix while the rank of $\boldsymbol{\Sigma}_{bb}$ in (i) can easily be detected to be 2. We first looked at the sample roots $\hat{\lambda}_i$'s as defined in (2) to see differences among the sample configurations and parameterizations. Table I reports the empirical frequency of the number of $\hat{\lambda}_i$'s larger than one. The true rank of $\boldsymbol{\Sigma}_{bb}$ being 2 implies that we expect exactly two roots to be larger than 1. We note that this number is the rank of $\hat{\boldsymbol{\Sigma}}_{bb}$ in (3) with $m=4$, the estimate with no information on the rank. For parameterization (i), all the samples produced at least two roots larger than 1, but can

TABLE I

Frequency of the Number of Roots $\hat{\lambda}_i$ Greater than One

		Parameterization					
		(i)			(ii)		
number	(n, r)	(50, 5)	(50, 50)	(5, 50)	(50, 5)	(50, 50)	(5, 50)
0		0	0	0	0	0	1
1		0	0	0	0	60	126
2		212	262	677	599	269	722
3		703	688	321	331	670	151
4		85	70	2	10	61	0

produce three or more roots larger than 1 quite often regardless of the sample configuration. For parameterization (ii), the number of roots larger than 1 tends to be smaller than parameterization (i).

Even though the true rank of Σ_{bb} is 2, we considered four different situations where a statistician believes the rank is at most m where $m = 1, 2, 3, 4$. The case with $m = 4$ corresponds to that with no information on the rank. As a summary, we report only on inferences for two parametric functions σ_{bb11}, the (1, 1) element of Σ_{bb} and $\tau_1 = \sigma_{bb11}/(\sigma_{bb11} + \sigma_{ww11})$, the intraclass correlation for the first variable. The true values are

$$(\sigma_{bb11}, \tau_1) = (1, 0.500), \quad \text{for (i)},$$
$$= (1, 0.032), \quad \text{for (ii)}.$$

Table II presents the relative bias (bias divided by the true value) of each of the four estimators corresponding to $m = 1, 2, 3, 4$. For the estimator with $m = 1$, the relative bias is large for parametrization (i), and in general the bias does not necessarily decrease with larger n or r. This result is consistent with the fact the estimator with $m = 1$ is not a consistent estimator in the sense of Section 2. The relative biases of the estimators with $m > 1$ are very similar. For these estimators, the bias is not a serious problem except for parameterization (ii) with either small n or small r. Large relative biases for (ii) with small r are due to a combination of the difficult parameter structure and the small true value of a positive parameter. For parameterization (ii), the estimator with $m = 2$ has a smaller bias than that

TABLE II

The Relative Biases (Bias/True Value) of the Estimators with $m = 1, 2, 3, 4$

	Parameterization					
	(i)			(ii)		
(n, r):	(50, 5)	(50, 50)	(5, 50)	(50, 5)	(50, 50)	(5, 50)
σ_{bb11} estimator						
$m = 1$	−0.294	−0.292	−0.166	0.184	−0.274	0.007
$m = 2$	0.012	0.006	0.019	0.591	0.022	0.200
$m = 3$	0.019	0.007	0.020	0.637	0.038	0.208
$m = 4$	0.019	0.007	0.020	0.638	0.039	0.208
τ_1 estimator						
$m = 1$	−0.322	−0.317	−0.291	0.171	−0.279	−0.022
$m = 2$	−0.007	−0.007	−0.097	0.577	0.019	0.171
$m = 3$	0.	−0.007	−0.096	0.620	0.034	0.176
$m = 4$	0.	−0.007	−0.096	0.623	0.034	0.176

with $m=3$ or $m=4$. Table III gives the mean square errors of the estimators. The estimator with $m=1$ has a mean square error which is either large or irregular (with relative to n or r). Once again, the estimators with $m=2, 3, 4$ exhibit similar behavior. For $m=3$ or $m=4$, the estimator is different from that with $m=2$ only if $\hat{\lambda}_3$ (or $\hat{\lambda}_4$) given in Table I is larger than one. Even for such a case, $\hat{\lambda}_3$ contributes to the estimator only through $\hat{\lambda}_3 - 1$; i.e., the part of $\hat{\lambda}_3$ larger than 1. Thus, when this part is small relative to $\hat{\lambda}_1 - 1$ and $\hat{\lambda}_2 - 1$, the differences are small. For the difficult case for estimation, i.e., for (ii) with small n or r, the estimator with $m=2$ seems to have a smaller mean square error than that with $m=3$ or 4. Hence, the under-specification of the rank seems to lead to a poor performing estimator, while the over-specification does not seriously hamper point estimation except possibly for the difficult cases with small random effect.

To assess the usefulness of the aproximate inference procedure suggested in Section 5, confidence intervals with nominal 95% of coverage were computed. These are based on each estimate, the corresponding standard error using $\hat{\mathbf{V}}$ in Section 5, the standard normal percentile (1.96), and the delta method for τ_1. Recall that this procedure based on a higher order expansion and normal approximation is asymptotically valid for Cases I and II and its use for Case III does not have valid justification. For Case III, and asymptotically valid procedure is possible for some parametric functions, using Wishart limiting distribution given in Theorem 3. For σ_{bb11}, such a procedure is the chi-square confidence interval with $n-1$

TABLE III

The Mean Square Errors ($\times$ 100 for τ_i) of the estimators with $m = 1, 2, 3, 4$

	Parameterization					
	(i)			(ii)		
(n, r):	(50, 5)	(50, 50)	(5, 50)	(50, 5)	(50, 50)	(5, 50)
σ_{bb11} estimator						
$m=1$	0.20273	0.24504	0.62424	1.45636	0.29814	1.17939
$m=2$	0.05809	0.04340	0.51607	1.75602	0.10434	1.18803
$m=3$	0.05788	0.04340	0.51701	1.80604	0.10435	1.18970
$m=4$	0.05789	0.04340	0.51601	1.80779	0.10436	1.18970
τ_1 estimator						
$m=1$	5.93730	5.53200	6.94470	0.14366	0.03189	0.10642
$m=2$	0.45376	0.26559	2.99960	0.17147	0.01007	0.10487
$m=3$	0.44039	0.26505	2.99310	0.17610	0.00998	0.10494
$m=4$	0.43971	0.26503	2.99310	0.17622	0.00998	0.10494

degrees of freedom. We also computed such an interval using each of the estimators with $m = 1, 2, 3, 4$ for all (n, r) pairs. For each of these different nominal 95% confidence intervals (8 for σ_{bb11} and 4 for τ_1), the percentage of containing the true value over 1000 replications was obtained. Table IV gives such results. As expected, the intervals based on the estimation under the rank at most 1 have very poor coverage. The differences among those based on $m = 2, 3, 4$ are small for either $\hat{\mathbf{V}}$-based or χ^2_{n-1}-based procedures. Comparing these two approaches with $m = 2, 3, 4$, the $\hat{\mathbf{V}}$-based procedures have larger (sometimes much larger) coverage than the χ^2-based, except for parameterization (i) with $(n, r) = (5, 50)$. The χ^2-based procedure is justifiable for large r cases, but does not seem to perform very well for the difficult cases with small random effect (relative to error) and for small r cases. The $\hat{\mathbf{V}}$-based normal interval, originally suggested as a possible procedure regardless of the sampling configuration, in fact provides good coverage properties over different cases, except for parameterization (i) with small n. For small n, the use of t cut-off points with $n-1$ or $n-1-m$ degrees of freedom would improve the coverage. For parameterization (ii), the over-specified rank ($m = 4$) tends to decrease the coverage. Once the maximum possible rank of a covariance component is reliably established

TABLE IV

Percentages of the Nominal 95% Confidence Intervals Containing the True Values (Methods Based on $\hat{\mathbf{V}}$ and χ^2_{n-1}).

		Parameterization					
		(i)			(ii)		
Method	(n, r):	(50, 5)	(50, 50)	(5, 50)	(50, 5)	(50, 50)	(5, 50)
	σ_{bb11} estimator						
$\hat{\mathbf{V}}$	$m = 1$	59.4	54.0	66.2	87.1	64.6	85.1
	$m = 2$	93.7	94.1	82.5	95.4	94.2	92.9
	$m = 3$	94.2	94.2	82.5	96.7	95.3	94.5
	$m = 4$	94.3	94.2	82.5	95.2	95.5	89.7
χ^2_{n-1}	$m = 1$	54.3	55.9	80.7	19.4	47.1	74.8
	$m = 2$	89.7	94.0	93.8	27.0	79.0	84.1
	$m = 3$	90.0	94.0	93.7	26.9	79.5	84.2
	$m = 4$	89.9	94.0	93.7	26.9	79.5	84.2
	τ_1 interval						
$\hat{\mathbf{V}}$	$m = 1$	53.2	45.3	67.2	85.9	62.4	84.9
	$m = 2$	95.1	94.6	86.9	94.6	93.3	92.6
	$m = 3$	94.7	94.6	87.0	95.6	94.6	93.9
	$m = 4$	94.7	94.6	87.0	94.3	94.6	89.2

(e.g., using a test procedure in Anderson and Amemiya [7], and Remadi and Amemiya [20]), we recommend estimation under the rank condition and approximate inference based on $\hat{\mathbf{V}}$ and normal or t cur-off points.

APPENDIX

Proof of Theorem 5. We first note that (16) in the proof of Theorem 2 and the restriction $m = m_0$ imply

$$P\{\hat{k} = m_0\} \rightarrow 1 \tag{17}$$

for all three cases. Thus, in deriving an asymptotic expansion, we can consider $\hat{\boldsymbol{\Sigma}}_{bb}(m_0)$ and $\hat{\boldsymbol{\Sigma}}_{ww}(m_0)$ with $\hat{k} = m_0$. It turns out to be easier to derive first the common expansion given in Theorem 6.

Note that for given $\mathbf{C}$ and $\mathbf{D}$ in (16), with probability one, $\mathbf{C}\mathbf{b}_i = 0$ for all i and $\mathbf{D}\mathbf{b}_i$ has a nonsingular covarinace matrix. Since the result is free of the choice of $\mathbf{C}$ and $\mathbf{D}$, we use, without loss of generality and with possible reordering of variables, $\mathbf{C}$ and $\mathbf{D}$ given by

$$\mathbf{C} = (\mathbf{I}_{p-m_0}, -\beta')', \qquad \mathbf{D} = (\mathbf{0}, \mathbf{I}_{m_0})' - \mathbf{C}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vu},$$

where β is a $m_0 \times (p - m_0)$ nonzero matrix, $\boldsymbol{\Sigma}_{vv} = \mathbf{C}'\boldsymbol{\Sigma}_{ww}\mathbf{C}$, and $\boldsymbol{\Sigma}_{vu} = \mathbf{C}'\boldsymbol{\Sigma}_{ww}(\mathbf{0}, \mathbf{I}_{m_0})'$. Correspondingly, we assume that $\boldsymbol{\Sigma}_{bb}$ has the form

$$\boldsymbol{\Sigma}_{bb} = \begin{pmatrix} \beta' \\ \mathbf{I}_{m_0} \end{pmatrix} \boldsymbol{\Sigma}_{xx}(\beta, \mathbf{I}_{m_0}),$$

where $\boldsymbol{\Sigma}_{xx}$ is a $m_0 \times m_0$ symmetric positive definite matrix. Note that $\mathbf{C}'\boldsymbol{\Sigma}_{bb}\mathbf{C} = \mathbf{0}$, rank $\mathbf{D} = m_0$, and $(\mathbf{C}, \mathbf{D})$ is a $p \times p$ nonsingular matrix. Also, $\mathbf{D}\mathbf{b}_i$ has the nonsingular covariance matrix $\boldsymbol{\Sigma}_{xx}$.

Using the expansion in Theorem 1 with $\hat{k} = m_0$ and the fact that $\mathbf{m}_{ww} = \boldsymbol{\Sigma}_{ww} + O_p(1/\sqrt{nr})$, we can write

$$\hat{\boldsymbol{\Sigma}}_{bb}(m_0) - \boldsymbol{\Sigma}_{bb} = \frac{1}{r}\left[\mathbf{m}_{bb} - (r\boldsymbol{\Sigma}_{bb} + \boldsymbol{\Sigma}_{ww})\right] + O_p\left(\frac{1}{r\sqrt{n}}\right),$$

$$\hat{\boldsymbol{\Sigma}}_{ww}(m_0) - \boldsymbol{\Sigma}_{ww} = \mathbf{m}_{ww} - \boldsymbol{\Sigma}_{ww} + O_p\left(\frac{1}{r\sqrt{n}}\right).$$

Let $\hat{\mathbf{P}} = \mathbf{m}_{ww}^{1/2}\hat{\mathbf{Q}}$, $\hat{\mathbf{T}} = \mathbf{m}_{ww}^{-1/2}\hat{\mathbf{Q}}$, and partition these matrices as

$$\hat{\mathbf{P}} = \begin{pmatrix} \hat{\mathbf{P}}_{11} & \hat{\mathbf{P}}_{12} \\ \hat{\mathbf{P}}_{21} & \hat{\mathbf{P}}_{22} \end{pmatrix}, \qquad \hat{\mathbf{T}} = \begin{pmatrix} \hat{\mathbf{T}}_{11} & \hat{\mathbf{T}}_{12} \\ \hat{\mathbf{T}}_{21} & \hat{\mathbf{T}}_{22} \end{pmatrix},$$

where $\hat{\mathbf{P}}_{11}$ and $\hat{\mathbf{T}}_{11}$ are $(p-m_0)\times m_0$ and $\hat{\mathbf{Q}}$ is defined in (2). Define

$$\hat{\mathbf{\Sigma}}_{xx}=\frac{1}{r}\hat{\mathbf{P}}_{21}(\hat{\mathbf{\Lambda}}_1-\mathbf{I}_{m_0})\hat{\mathbf{P}}_{21}',$$

$$\hat{\beta}=(\hat{\mathbf{P}}_{11}\hat{\mathbf{P}}_{21}^{-1})'=-\hat{\mathbf{T}}_{22}\hat{\mathbf{T}}_{12}^{-1}.$$

Furthermore, if we let $\hat{\mathbf{C}}=(\mathbf{I}_{p-m_0}, -\hat{\beta}')'$ and $\hat{\mathbf{A}}=\hat{\mathbf{T}}_{12}\hat{\mathbf{T}}_{12}'$, then

$$\begin{aligned}\hat{\mathbf{A}}&=(\hat{\mathbf{C}}'\mathbf{m}_{ww}\hat{\mathbf{C}})^{-1}\\ \mathbf{\Omega}_{m_0}&=\mathbf{m}_{ww}\hat{\mathbf{C}}\hat{\mathbf{A}}\hat{\mathbf{C}}'(\mathbf{m}_{bb}-\mathbf{m}_{ww})\hat{\mathbf{C}}\hat{\mathbf{A}}\hat{\mathbf{C}}'\mathbf{m}_{ww},\end{aligned}\tag{19}$$

where $\mathbf{\Omega}_{m_0}$ is defined in (4) with $\hat{k}=m_0$. See Amemiya and Fuller [2, p. 449]. Let $\mathbf{S}_{xx}$, $\mathbf{S}_{vx}$, $\mathbf{S}_{vv}$, $\mathbf{S}_{ee}$, and $\mathbf{S}_{vv\cdot x}$ be as defined in (16) with this particular choice of $\mathbf{C}$ and $\mathbf{D}$.

Multiplying the first equation in (18) by $(\mathbf{0}, \mathbf{I}_{m_0})$ on the left and by $(\mathbf{0}, \mathbf{I}_{m_0})'$ on the right, we get for all tree cases

$$\hat{\mathbf{\Sigma}}_{xx}=\mathbf{S}_{xx}+O_p\left(\frac{1}{\sqrt{nr}}\right).$$

Multiplying the first equation in (18) by $\mathbf{C}'$ on the left and by $(\mathbf{0}, \mathbf{I}_{m_0})'$ on the right, we get for all three cases

$$\hat{\beta}-\beta=\frac{1}{\sqrt{r}}\mathbf{S}_{xx}^{-1}\mathbf{S}_{vx}'+O_p\left(\frac{1}{r\sqrt{n}}\right).\tag{20}$$

Since $1/\sqrt{r}\,\mathbf{S}_{vx}=O_p(1/\sqrt{nr})$, (20) implies that

$$\hat{\mathbf{C}}-\mathbf{C}=O_p\left(\frac{1}{\sqrt{nr}}\right).\tag{21}$$

Using the form $\hat{\mathbf{C}}=\mathbf{C}-\mathbf{E}$ with $\mathbf{E}=[0, (\hat{\beta}-\beta)']'$, we have for all cases

$$\hat{\mathbf{C}}'(\mathbf{m}_{bb}-\mathbf{m}_{ww})\hat{\mathbf{C}}=\mathbf{S}_{vv\cdot x}-\mathbf{S}_{ee}+O_p\left(\frac{1}{n\sqrt{r}}\right).\tag{22}$$

Also, from (19), (21), and the fact that $\mathbf{m}_{ww}=\mathbf{\Sigma}_{ww}+O_p(1/\sqrt{nr})$, we have for all three cases

$$\hat{\mathbf{A}}=\mathbf{\Sigma}_{vv}^{-1}+O_p\left(\frac{1}{\sqrt{nr}}\right).\tag{23}$$

Now, using the second equation in (19), (21), (22), and (23), and the fact that $\mathbf{m}_{ww} = \mathbf{\Sigma}_{ww} + O_p(1/\sqrt{nr})$, it follows that

$$\mathbf{\Omega}_{m0} = \mathbf{\Sigma}_{ww}\mathbf{C}\mathbf{\Sigma}_{vv}^{-1}(\mathbf{S}_{vv\cdot x} - \mathbf{S}_{ee})\,\mathbf{\Sigma}_{vv}^{-1}\,\mathbf{C}'\mathbf{\Sigma}_{ww} + O_p\left(\frac{1}{n\sqrt{r}}\right). \qquad (24)$$

The result for Case II follows from (21) and (24). The approximation for Cases I and III follow from the fact that $\mathbf{\Omega}_{\rm II} = \mathbf{\Omega}_{\rm I} + O_p(1/n)$ for Case I and $(1/r)(\mathbf{m}_{ww} - \mathbf{\Sigma}_{ww}) = O_p(1/r\sqrt{r})$ for Case III.

References

1. Amemiya, Y. (1985). What should be done when an estimated between-group covariance matrix is not nonnegative definite. *Amer. Statist.* **39** 112–117.
2. Amemiya, Y. and Fuller, W. A. (1984). Estimation for the multivariate errors-in-variables model with estimated error covariance matrix. *Ann. Statist.* **12** 497–509.
3. Amemiya, Y., Anderson, T. W., and Lewis, P. A. W. (1990). Percentage points for a test of rank in multivariate components of variance. *Biometrika* **77** 637–641.
4. Anderson, B., Anderson, T. W., and Olkin, I. (1986). Maximum likelihood estimators and likelihood ratio criterion in multivariate components of variance. *Ann. Statist.* **14** 405–417.
5. Anderson, T. W. (1984). Estimating linear statistical relationships. *Ann. Statist.* **12** 1–45.
6. Anderson, T. W. (1989). The asymptotic distribution of the likelihood ratio criterion for testing the rank in multivariate components of variance. *J. Multivariate Anal.* **30** 72–79.
7. Anderson, T. W., and Amemiya, Y. (1991). Testing dimensionality in the multivarite analysis of variance. *Statist. Probab. Lett.* **12** 445–463.
8. Bhargava, A. K., and Disch, D. (1982). Exact probabilities of obtaining estimated non-positive definite between-group covariance matrices. *J. Statist. Comput. Simulation* **15** 27–32.
9. Bock, R. D., and Petersen, A. C. (1975). A multivariate correction for attenuation. *Biometrika* **62** 673–678.
10. Calvin, J. A., and Dykstra, R. L. (1991). Maximum likelihood estimation of a set of covariance matrices under lower order restrictions with applications to balanced multivariate variance components models. *Ann. Statist.* **19** 850–869.
11. Feller, W. (1970). *An Introduction to Probability and Its Applications*, Vol. II, 2nd ed. Wiley, New York.
12. Harville, D. A. (1977). Maximum likelihood approaches to variance components estimation and to related problems. *J. Amer. Statist. Assoc.* **72** 320–338.
13. Hill, W. G., and Thompson, R. (1978). Probabilities of nonpositive definite between-group or genetic covariance matrices. *Biometrics* **34**, 429–439.
14. Klotz, J. H., and Putter, J. (1969). Maximum likelihood estimation of multivariate covariance components for the balanced one-way lay-out. *Ann. Math. Statist.* **40**, 1100–1105.
15. Mathew, T., Niyogi, A. and Sinha, B. K. (1992). *Improved Nonnegative Estimation of Variance Components in Balanced Multivariate Mixed Models*, Technical Report. Department of Mathematics and Statistics, University of Maryland, Baltimore County.
16. Meyer, K. (1985). Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics* **41** 153–165.

17. MILLER, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.* **5** 746–762.
18. OKAMOTO, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *Ann. Statist.* **1** 763–765.
19. REMADI, S., AND AMEMIYA, Y. (1992). Limiting distribution of roots with differential rates of convergence and its applications. *Statist. Brobab. Lett.*, to appear.
20. REMADI, S., AND AMEMIYA, Y. (1992). *Test of Rank for Covariance Components*, Preprints 92–13. Department of Statistics, Iowa State University.
21. ROBINSON, G. K. (1991). That a BLUP is a good thing: The estimation of random effects. *Statist. Sci.* **6** 15–51.
22. SCHOTT, J. R., AND SAW, J. G. (1984). A multivariate one-way classification model with random effects. *J. Multivariate Anal.* **15** 1–12.
23. SEARLE, S. R., CASELLA, G., AND MCCULLOCH, C. E. (1992). *Variance Components.* Wiley, New York.
24. THOMPSON, R. (1973). The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics* **22** 527–550.