JOURNAL OF MULTIVARIATE ANALYSIS **51**, 83–101 (1994)

# Improved Nonnegative Estimation of Variance Components in Balanced Multivariate Mixed Models

THOMAS MATHEW,* ANINDITA NIYOGI, AND BIMAL KUMAR SINHA

*University of Maryland–Baltimore County*

Consider the independent Wishart matrices $S_1 \sim W(\Sigma + \lambda\Theta, q_1)$ and $S_2 \sim W(\Sigma, q_2)$, where $\Sigma$ is an unknown positive definite (p.d.) matrix, $\Theta$ is an unknown nonnegative definite (n.n.d.) matrix, and $\lambda$ is a known positive scalar. For the estimation of $\Theta$, a class of estimators of the form $\hat{\Theta}_{(c,\varepsilon)} = (c/\lambda)\{S_1/q_1 - \varepsilon(S_2/q_2)\}$ $(c \geq 0,\ \varepsilon \leq 1)$, uniformly better than the unbiased estimator $\hat{\Theta}_U = (1/\lambda)\{S_1/q_1 - S_2/q_2\}$, is derived (for the squared error loss function). Necessary and sufficient conditions are obtained for the existence of an n.n.d. estimator of the form $\hat{\Theta}_{(c,\varepsilon)}$ uniformly better than $\hat{\Theta}_U$. It turns out that such an n.n.d. estimator exists only under restrictive conditions. However, for a suitable choice of $c > 0$, $\varepsilon > 0$, the estimator obtained by taking the positive part of $\hat{\Theta}_{(c,\varepsilon)}$ results in an n.n.d. estimator, say $\hat{\Theta}_{(c,\varepsilon)+}$, that is uniformly better than $\hat{\Theta}_U$. Numerical results indicate that in terms of mean squared error, $\hat{\Theta}_{(c,\varepsilon)+}$ performs much better than both $\hat{\Theta}_U$ and the restricted maximum likelihood estimator $\hat{\Theta}_{\mathrm{REML}}$ of $\Theta$. Similar results are also obtained for the nonnegative estimation of $\operatorname{tr}\Theta$ and $\mathbf{a}'\Theta\mathbf{a}$, where $\mathbf{a}$ is an arbitrary nonzero vector. For estimating $\Sigma$, we have derived estimators that are claimed to be uniformly better than the unbiased estimator $\hat{\Sigma}_U = S_2/q_2$ under the squared error loss function and the entropy loss function. We have been able to establish the claim only in the bivariate case. Numerical results are reported showing the risk improvement of our proposed estimators of $\Sigma$. 

## 1. INTRODUCTION AND SUMMARY

In the context of univariate mixed effects models, a problem that has received considerable attention is the nonnegative estimation of variance components corresponding to the random effects. Several procedures, which include the maximum likelihood (ML) and restricted maximum likelihood (REML) approaches, have been suggested to arrive at nonnegative estimators of the variance components in the univariate set up.

Received September 24, 1992; revised August 24, 1993.

AMS 1980 subject classifications: primary 62H12; secondary 62J10.

Key words and phrases: balanced models, entropy loss, multivariate components of variance, restricted maximum likelihood estimator, squared error loss.

* Research sponsored by Grant AFOSR 89-0237.



0047-259X/94 $6.00

However, the problem has not received the same amount of attention in the multivariate context, even though multivariate models with mixed effects can frequently arise in applications. In a recent application considered in Calvin and Dykstra (1991a, b) and Calvin and Sedransk (1991) dealing with the quality of care received by cancer patients at hospitals in the United States, the data was analysed using the following two-way nested random effects model:

$$\mathbf{y}_{ijk} = \mu + \mathbf{a}_i + \mathbf{b}_{ij} + \mathbf{e}_{ijk}. \tag{1.1}$$

Here $\mathbf{y}_{ijk}$ is a bivariate response vector, $\mu$ is a general mean, $\mathbf{a}_i$, $\mathbf{b}_{ij}$, and $\mathbf{e}_{ijk}$ are independent bivariate random vectors distributed as $\mathbf{a}_i \sim N(0, \Sigma_a)$, $\mathbf{b}_{ij} \sim N(0, \Sigma_b)$, and $\mathbf{e}_{ijk} \sim N(0, \Sigma_e)$. When experimental data can be modelled using a multivariate mixed effects model, an important problem is the inference concerning the multivariate components of variance, for example, estimation of the parameter matrices $\Sigma_a$, $\Sigma_b$, and $\Sigma_e$ in the model (1.1). In the present paper, we shall treat this problem in multivariate balanced models involving exactly one random effect and hence two variance components, namely the variance component corresponding to the random effect and that corresponding to the experimental error. The problem of estimating a variance component in such models reduces to the consideration of two independent random $p \times p$ matrices, say $S_1$ and $S_2$, following the central Wishart distributions

$$S_1 \sim W_p(\Sigma + \lambda\Theta, q_1) \quad \text{and} \quad S_2 \sim W_p(\Sigma, q_2), \tag{1.2}$$

where $q_1$ and $q_2$ satisfy $q_1 \geqslant p$, $q_2 \geqslant p$, $\lambda$ is a positive scalar, $\Sigma$ is the covariance matrix of the experimental errors in the model, and $\Theta$ is the covariance matrix corresponding to the random effect. Here $\Sigma$ is assumed to be a positive definite (p.d.) matrix and $\Theta$ is assumed to be nonnegative definite (n.n.d.). Unbiased estimators of $\Theta$ and $\Sigma$, which are obviously given by

$$\hat{\Theta}_U = \frac{1}{\lambda}\left\{\frac{S_1}{q_1} - \frac{S_2}{q_2}\right\} \tag{1.3}$$

and

$$\hat{\Sigma}_U = \frac{S_2}{q_2}, \tag{1.4}$$

are obtained using the multivariate analysis of variance procedure applied to the mixed model involving $\Theta$ and $\Sigma$ as the components of variance. As is well known, $\hat{\Sigma}_U$ is p.d. but $\hat{\Theta}_U$ is not always n.n.d.

Amemiya (1985) has suggested a procedure to modify $\hat{\Theta}_U$ to make it n.n.d. It turns out that the resulting estimator is the REML estimator of $\Theta$. The procedure can be described as follows. First obtain a nonsingular matrix $T$ satisfying $S_1 = T'\Lambda T$ and $S_2 = T'T$, where $\Lambda = \text{diag}(\lambda_1, ..., \lambda_p)$ is a diagonal matrix. Then $\hat{\Theta}_U = (1/\lambda)\, T'(\Lambda/q_1 - I/q_2)\, T$. In the diagonal matrix $\Delta = \Lambda/q_1 - I/q_2$, replace any negative diagonal element by zero and call the resulting matrix $\Delta_*$. Then n.n.d. estimator suggested by Amemiya (1985) is

$$\hat{\Theta}_{\text{REML}} = \frac{1}{\lambda}\, T'\, \Delta_*\, T, \tag{1.5}$$

where we use the notation $\hat{\Theta}_{\text{REML}}$ to emphasize that the estimator in (1.5) is also the REML estimator of $\Theta$. In fact, for a general balanced multivariate mixed effects model, Calvin and Dykstra (1991a) have proposed an algorithm to derive the REML estimators of the multivariate components of variance. In another paper, Calvin and Dykstra (1991b), a least squares type approach is used for estimating multivariate components of variance. For the model (1.2), which involves only two variance components, the REML estimators of $\Theta$ and $\Sigma$ have explicit expressions as seen from (1.5) and (3.1) However, for a general multivariate balanced mixed effects model, the REML estimators do not have analytic expressions and have to be computed by an iterative algorithm.

The estimator suggested by Amemiya (1985) and those derived by Calvin and Dykstra (1991a, b) are motivated only by the requirement that the estimated covariance matrix $\hat{\Theta}$ be n.n.d. It was not required that the estimators possess good frequentist properties in the sense of having a smaller risk, say, compared to the unbiased estimator $\hat{\Theta}_U$. [If $\hat{\Theta}$ is an estimator of $\Theta$ in (1.2), the loss function we shall use is $\text{tr}(\hat{\Theta} - \Theta)^2$]. In fact, unlike in the univariate case, risk comparisons of various n.n.d. estimators of multivariate components of variance is not available in the literature. A major goal of this article is to derive n.n.d. estimators of $\Theta$ having good performance in terms of risk; in particular, n.n.d. estimators that are uniformly better than $\hat{\Theta}_U$ in (1.3). Furthermore, the estimators we shall derive have explicit analytic expressions and hence are easily computable. Our research in this context is motivated by similar results in the univariate case recently obtained by Mathew *et al.* (1992a) and Kelly and Mathew (1993). In these papers, the authors have obtained satisfactory nonnegative estimators of the variance components for univariate balanced mixed models. In Section 2, we have achieved the same for the model (1.2). Numerical results regarding the performance of the proposed estimators as well as those derived by Amemiya (1985) and Calvin and Dykstra (1991a, b) are also reported. It turns out that the n.n.d. estimators that we have proposed have considerable advantage over the others in the sense of having a smaller risk. We would also like to point out that our results in Section 2, derived in the context of a balanced multivariate model involving only one random effect (namely, the model (1.2)) is also applicable to some models involving more than one random effect, including, for example, the model (1.1). Even though the model (1.1) involves three variance components, namely, $\Sigma_a$, $\Sigma_b$, and $\Sigma_e$, for estimating the random effects variance components $\Sigma_a$ and $\Sigma_b$, the problem actually reduces to the same in a model that involves only two variance components. This is so because for the model (1.1) the sum of squares and sum of products matrices, say $S_a$, $S_b$, and $S_e$ are independently distributed as

$$S_a \sim W_2(\Sigma_e + 2\Sigma_b + 4\Sigma_a, 6), \qquad S_b \sim W_2(\Sigma_e + 2\Sigma_b, 7),$$

and

$$S_e \sim W_2(\Sigma_e, 14). \tag{1.6}$$

Thus, for estimating $\Sigma_a$, it is enough to consider the matrices $S_a$ and $S_b$ so that the model involves only the two variance components $(\Sigma_e + 2\Sigma_b)$ and $\Sigma_a$. This is obviously true for estimating $\Sigma_b$ as well. (However, a similar conclusion is not always true for any multivariate balanced mixed model; for some counter examples in the univariante case, see Mathew *et al.* (1992a)).

In applications, it may be of interest to estimate certain scalar valued functions of a multivariate component of variance, for example, tr $\Theta$ or the sum of all the elements of the variance component matrix $\Theta$ or just its diagonal elements $\Theta_{ii}$. In the context of the model (1.1), it may be of some interest to estimate tr $\Sigma_b$ or $\mathbf{1}'\Sigma_b\mathbf{1}$ or $\Sigma_{b(ii)}$, where $\mathbf{1}$ denotes a vector of ones of appropriate dimension and $\Sigma_{b(ii)}$ denotes the $i$th diagonal element of $\Sigma_b$. tr $\Sigma_b$ is clearly the sum of the variances of the components of the random effect $\mathbf{b}_{ij}$ and $\mathbf{1}'\Sigma_b\mathbf{1}$ is the variance of $\mathbf{1}'\mathbf{b}_{ij}$, i.e., the variance of the sum of the random variables in $\mathbf{b}_{ij}$. For the model (1.2), the nonnegative estimation of tr $\Theta$ and $\mathbf{a}'\Theta\mathbf{a}$, $\mathbf{a} \neq \mathbf{0}$, is also addressed in Section 2. We have once again numerically compared the various competing estimators of these scalar valued functions of $\Theta$.

Section 3 deals with the estimation of the error variance component $\Sigma$ in the model (1.2). The unbiased estimator $\hat{\Sigma}_U$ in (1.4) is obviously p.d. and is a function of only $S_2$. In Section 3 we claim that this estimator can be uniformly improved using an estimator that is a function of both $S_1$ and $S_2$ under two loss functions, namely, the entropy loss function $L_1(\hat{\Sigma}, \Sigma) = \operatorname{tr} \hat{\Sigma}\Sigma^{-1} - \ln |\hat{\Sigma}\Sigma^{-1}| - p$ and the invariant squared error loss function $L_2(\hat{\Sigma}, \Sigma) = \operatorname{tr}(\hat{\Sigma}\Sigma^{-1} - I)^2$. However, we have been able to prove this claim only when $p = 2$. Numerical results are reported regarding the performance of the new estimators. Our results in this regard generalize the corresponding univariate results given in Mathew *et al.* (1992b).

## 2. Nonnegative Estimation of $\Theta$

The problem we shall address in this section is that of obtaining some n.n.d. estimators of $\Theta$ that are uniformly better than $\hat{\Theta}_U$ given in (1.3). We first explore the possibility of improving $\hat{\Theta}_U$ using estimators that are linear combinations of $S_1$ and $S_2$ given in (1.2). Following Kelly and Mathew (1993), we express such an estimator as

$$\hat{\Theta}_{(c,\varepsilon)}=\frac{c}{\lambda}\left[\frac{1}{q_1}S_1-\frac{\varepsilon}{q_2}S_2\right], \tag{2.1}$$

where $c$ and $\varepsilon$ are real numbers. Clearly, $\hat{\Theta}_{(c,\varepsilon)}$ is n.n.d. if $c\geqslant 0$ and $\varepsilon\leqslant 0$.

We shall first characterize $c$ and $\varepsilon$ such that $\hat{\Theta}_{(c,\varepsilon)}$ has a uniformly smaller risk compared to $\hat{\Theta}_U$, where, for an estimator $\hat{\Theta}$ of $\Theta$, the risk function is defined as

$$R(\hat{\Theta},\Theta)=\operatorname{tr} E(\hat{\Theta}-\Theta)^2. \tag{2.2}$$

We shall consider only $c$ and $\varepsilon$ satisfying $c\geqslant 0$ and $\varepsilon\leqslant 1$ since, as will be seen shortly, it is enough to consider only such values to achieve risk improvement over $\hat{\Theta}_U$. Also note that if $c<0$ or if $\varepsilon>1$, $\hat{\Theta}_{(c,\varepsilon)}$ is not always n.n.d. and furthermore when $c>0$, if $\varepsilon>1$ then $\hat{\Theta}_{(c,\varepsilon)}$ is more likely to be not n.n.d. compared to $\hat{\Theta}_U$. We now compute the risk of $\hat{\Theta}_{(c,\varepsilon)}$. For this we use the fact that if $A=((a_{ij}))\sim W_p(\Sigma,m)$, then $E(A)=m\Sigma$ and $\operatorname{Cov}(a_{ij},a_{kl})=m(\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk})$, where $\sigma_{ij}$ is the $(ij)$th element of $\Sigma$ (see Muirhead, 1982, p. 90). Direct computations then yield

$$\begin{aligned}R(\hat{\Theta}_{(c,\varepsilon)},\Theta)=&\frac{c^2}{\lambda^2 q_1}\left[\operatorname{tr}(\Sigma+\lambda\Theta)^2+\{\operatorname{tr}(\Sigma+\lambda\Theta)\}^2\right]\\&+\frac{c^2\varepsilon^2}{\lambda^2 q_2}\left[\operatorname{tr}(\Sigma^2)+(\operatorname{tr}\Sigma)^2\right]+\frac{c^2}{\lambda^2}(1-\varepsilon)^2\operatorname{tr}(\Sigma^2)\\&+(c-1)^2\operatorname{tr}(\Theta^2)+2\frac{c}{\lambda}(c-1)(1-\varepsilon)\operatorname{tr}(\Sigma\Theta).\end{aligned} \tag{2.3}$$

Equation (2.3) obviously reduces to the risk of $\hat{\Theta}_U$ when $c=1$, $\varepsilon=1$. Letting $\Sigma\to 0$ and $\Theta\to 0$ respectively in the expression for the risk difference $R(\hat{\Theta}_{(c,\varepsilon)},\Theta)-R(\hat{\Theta}_U,\Theta)$, we see that the inequalities (2.4) and (2.5) below are necessary for $R(\hat{\Theta}_{(c,\varepsilon)},\Theta)\leqslant R(\hat{\Theta}_U,\Theta)$;

$$\frac{1}{q_1}(c^2-1)\{\operatorname{tr}(\Theta^2)+(\operatorname{tr}\Theta)^2\}+(c-1)^2\operatorname{tr}(\Theta^2)\leqslant 0 \tag{2.4}$$

for all n.n.d. $\Theta$, and

$$\left\{c^2\left(\frac{1}{q_1}+\frac{\varepsilon^2}{q_2}\right)-\left(\frac{1}{q_1}+\frac{1}{q_2}\right)\right\}\{\operatorname{tr}(\Sigma^2)+(\operatorname{tr}\Sigma)^2\}$$
$$+c^2(1-\varepsilon)^2\operatorname{tr}(\Sigma^2)\leqslant 0 \tag{2.5}$$

for all p.d. $\Sigma$. Observe that for (2.4) to hold, the condition $c^2-1\leqslant 0$ is necessary. Arguing as in the proof of Lemma 2.1 in Mathew *et al.* (1992a), it can be shown that (2.4) and (2.5) are sufficient as well for $R(\hat{\Theta}_{(c,\varepsilon)},\Theta)\leqslant R(\hat{\Theta}_U,\Theta)$.

We now derive the conditions under which (2.4) and (2.5) will hold uniformly in $\Theta$ and $\Sigma$ respectively. Observe that it is enough to derive conditions under which (2.4) will hold for all n.n.d. $\Theta$ satisfying $\operatorname{tr}(\Theta^2)=1$. Since $c^2-1\leqslant 0$, the LHS of (2.4) is a maximum when $\operatorname{tr}\Theta$ is a minimum and subject to $\operatorname{tr}\Theta^2=1$, the minimum value of $\operatorname{tr}\Theta$ is 1. Consequently, (2.4) holds for all n.n.d. $\Theta$ if and only if

$$\frac{2(c^2-1)}{q_1}+(c-1)^2\leqslant 0. \tag{2.6}$$

Using a similar argument, we see that (2.5) holds for all p.d. $\Sigma$ if and only if

$$2\left\{c^2\left(\frac{1}{q_1}+\frac{\varepsilon^2}{q_2}\right)-\left(\frac{1}{q_1}+\frac{1}{q_2}\right)\right\}+c^2(1-\varepsilon)^2\leqslant 0. \tag{2.7}$$

Thus $R(\hat{\Theta}_{(c,\varepsilon)},\Theta)\leqslant R(\hat{\Theta}_U,\Theta)$ uniformly in $\Sigma$ and $\Theta$ if and only if $c$ and $\varepsilon$ satisfy (2.6) and (2.7). Finally, we observe that (2.6) is equivalent to

$$c_0\leqslant c\leqslant 1, \tag{2.8}$$

where

$$c_0=\max\left\{0,\frac{q_1-2}{q_1+2}\right\}. \tag{2.9}$$

Furthermore, for any $c$ satisfying (2.8), an interval for $\varepsilon$ can be specified from (2.7) as

$$\varepsilon_{0c}\leqslant\varepsilon\leqslant\varepsilon_{1c}, \tag{2.10}$$

where $\varepsilon_{0c}$ and $\varepsilon_{1c}$ are the lower and upper bounds (depending on $c$) for $\varepsilon$, obtained from (2.7).

It is interesting to observe that (2.6) and (2.7) involve only the d.f.'s of $S_1$ and $S_2$, and not their dimension, implying thereby that these two inequalities coincide with similar conditions derived by Kelly and Mathew (1993) in the univariate case. Consequently, for improving over $\hat{\Theta}_U$ using an estimator of the form $\hat{\Theta}_{(c,\varepsilon)}$, all the results in the univariate case can be adopted in the multivariate set up as well. Thus, there exists an n.n.d. estimator $\hat{\Theta}_{(c,\varepsilon)}$ having uniformly smaller risk compared to $\hat{\Theta}_U$ if and only if (Mathew *et al.*, 1992a)

$$q_1 \geqslant 3 \quad \text{and} \quad (q_1-2)^2 q_2 \leqslant 2(q_1+q_2)(q_1+2), \quad \text{or} \quad q_1 = 1 \text{ or } 2. \tag{2.11}$$

We summarize the above observations in the following theorem.

THEOREM 2.1. *Consider the model* (1.2) *where* $q_1$ *and* $q_2$ *satisfy* $q_1 \geqslant p$, $q_2 \geqslant p$. *Let* $c_0, \varepsilon_{0c}$ *and* $\varepsilon_{1c}$ *be given by* (2.9) *and* (2.10). *Then*

(i) *if* $q_1 = 1$ *or* 2, *the estimtor* $\hat{\Theta}_{(c,\varepsilon)}$ *has a uniformly smaller risk than* $\hat{\Theta}_U$ *if* $c=0$, *or if* $\varepsilon \leqslant \min\{1, \varepsilon_{1c}\}$ *for any* $c$ *satisfying* $0 < c \leqslant 1$;

(ii) *if* $q_1 \geqslant 3$, *the estimator* $\hat{\Theta}_{(c,\varepsilon)}$ *has a uniformly smaller risk than* $\hat{\Theta}_U$ *if* $\varepsilon_{0c} \leqslant \varepsilon \leqslant \min\{1, \varepsilon_{1c}\}$ *for any* $c$ *satisfying* $c_0 \leqslant c \leqslant 1$; *and*

(iii) *there exists a nonnegative definite estimator of the form* $\hat{\Theta}_{(c,\varepsilon)}$ *having a uniformly smaller risk than* $\hat{\Theta}_U$ *if and only if* (2.11) *holds.*

In situations where (2.11) holds, we have an n.n.d. $\hat{\Theta}_{(c,\varepsilon)}$ providing uniform risk improvement over $\hat{\Theta}_U$. However, if (2.11) does not hold, such an n.n.d. $\hat{\Theta}_{(c,\varepsilon)}$ does not exist and in this case if risk improvement over $\hat{\Theta}_U$ is desired along with nonnegativity, our recommendation is as follows: choose $c$ and $\varepsilon$ so that $R(\hat{\Theta}_{(c,\varepsilon)}, \Theta) \leqslant R(\hat{\Theta}_U, \Theta)$ (one possible choice is $c_1$ and $\varepsilon_1$ given later in (2.12)). Now "truncate" the estimator $\hat{\Theta}_{(c,\varepsilon)}$ at zero, i.e., take the spectral decomposition of $\hat{\Theta}_{(c,\varepsilon)}$ and replace any negative eigenvalue by zero. We shall denote the n.n.d. estimator so obtained by $\hat{\Theta}_{(c,\varepsilon)+}$. The following lemma, which is an obvious extension of the corresponding univariate result, shows that $\hat{\Theta}_{(c,\varepsilon)+}$ has a uniformly smaller risk compared to $\hat{\Theta}_U$.

LEMMA 2.1. *Let* $\tilde{\Theta}$ *be any estimator of* $\Theta$ *satisfying* $R(\tilde{\Theta}, \Theta) \leqslant R(\hat{\Theta}_U, \Theta)$ *and let* $\tilde{\Theta}_+$ *denote the estimator obtained by replacing the negative eigenvalues by zero in the spectral decomposition of* $\tilde{\Theta}$. *Then* $\operatorname{tr}(\tilde{\Theta}_+ - \Theta)^2 \leqslant \operatorname{tr}(\tilde{\Theta} - \Theta)^2$ *and hence,* $R(\tilde{\Theta}_+, \Theta) \leqslant R(\hat{\Theta}_U, \Theta)$ *uniformly.*

*Proof.* Consider the spectral decomposition $\tilde{\Theta} = PDP'$, where $P$ is an orthogonal matrix and $D = \operatorname{diag}(d_1, ..., d_r, d_{r+1}, ... d_p)$ is the diagonal matrix consisting of the eigenvalues of $\tilde{\Theta}$ with $d_i \geqslant 0$ for $i = 1, 2, ..., r$ and $d_i < 0$ for $i = r+1, ... p$. Then $\tilde{\Theta}_+ = P \operatorname{diag}(d_1, ..., d_r, 0, ..., 0) P'$. Defining $P\Theta P' = \Theta^* = ((\theta^*_{ij}))$, we get

$$\begin{aligned}
\operatorname{tr}(\tilde{\Theta}-\Theta)^2 &= \operatorname{tr}(D-\Theta^*)^2 \\
&= \sum_{i=1}^{p} (d_i-\theta_{ii}^*)^2 + \sum_{i\neq j} \theta_{ij}^{*2} \\
&\geqslant \sum_{i=1}^{r} (d_i-\theta_{ii}^*)^2 + \sum_{i=r+1}^{p} \theta_{ii}^{*2} + \sum_{i\neq j} \theta_{ij}^{*2} \\
&\qquad (\text{since } d_i<0 \text{ for } i=r+1, ..., p) \\
&= \operatorname{tr}(\tilde{\Theta}_+ - \Theta)^2,
\end{aligned}$$

which proves the lemma.

A question of practical importance is the choice of $c$ and $\varepsilon$. A choice that we recommend is obtained by minimizing the LHS of (2.6) with respect to $c$ and the LHS of (2.7) with respect to $\varepsilon$. The resulting values, say $c_1$ and $\varepsilon_1$, are given by

$$c_1=\frac{q_1}{q_1+2} \qquad \text{and} \qquad \varepsilon_1=\frac{q_2}{q_2+2}. \tag{2.12}$$

The above choice is motivated by the fact that the LHSs of (2.6) and (2.7) are respectively the coefficients of $\theta_{ii}^2$ and $\sigma_{ii}^2$ in the risk of $\hat{\Theta}_{(c,\varepsilon)}$. In other words, $c_1$ and $\varepsilon_1$ minimize the leading terms in the risk of $\hat{\Theta}_{(c,\varepsilon)}$. However, $\hat{\Theta}_{(c_1,\varepsilon_1)}$ is not always n.n.d. We thus recommend $\hat{\Theta}_{(c_1,\varepsilon_1)+}$ as an estimator of $\Theta$. The estimator $\hat{\Theta}_{(c_1,\varepsilon_1)+}$ is n.n.d., it is uniformly better than $\hat{\Theta}_U$ (from Lemma 2.1) and the simulation results reported below show that its risk performance is much better than the REML estimator $\hat{\Theta}_{\mathrm{REML}}$.

It should be noted that the above modification of $\hat{\Theta}_{(c_1,\varepsilon_1)}$ to arrive at $\hat{\Theta}_{(c_1,\varepsilon_1)+}$ is quite different from the modification of $\hat{\Theta}_U$ to arrive at the REML estimator given in (1.5). It is clear that the same modification can also be applied to $\hat{\Theta}_{(c_1,\varepsilon_1)}$ to arrive at the n.n.d. estimator, say $\hat{\Theta}_{(c_1,\varepsilon_1)*}$ and the spectral decomposition approach can in turn be applied to $\hat{\Theta}_U$ resulting in an n.n.d. estimator, say $\hat{\Theta}_{U+}$. While it is clear (from Lemma 2.1) that $\hat{\Theta}_{U+}$ is uniformly better than $\hat{\Theta}_U$, it is not clear if $\hat{\Theta}_{(c_1,\varepsilon_1)*}$ is uniformly better than $\hat{\Theta}_U$. However, our numerical computations indicate that the difference between the MSEs (and also the biases) of $\hat{\Theta}_{(c_1,\varepsilon_1)+}$ and $\hat{\Theta}_{(c_1,\varepsilon_1)*}$ (and also between those of $\hat{\Theta}_{U+}$ and $\hat{\Theta}_{\mathrm{REML}}$) is not significant.

In Table I, we have reported the MSEs and squared biases of several competing estimators of $\Theta$ in the model (1.2) for $p=2$, $q_1=5$ and $q_2=5$. $\lambda$ appearing in (1.2) was taken to be one. Note that if $\hat{\Theta}$ is an estimator of $\Theta$, the squared bias is defined as $\operatorname{tr}(E(\hat{\Theta})-\Theta)^2$. The estimators considered are $\hat{\Theta}_U$, $\hat{\Theta}_{\mathrm{REML}}$, $\hat{\Theta}_{U+}$, $\hat{\Theta}_{(c_1,\varepsilon_1)+}$, and $\hat{\Theta}_{(c_1,\varepsilon_1)*}$, where $c_1$ and $\varepsilon_1$ are given by (2.12) and $\hat{\Theta}_{(c_1,\varepsilon_1)+}$ and $\hat{\Theta}_{(c_1,\varepsilon_1)*}$ are defined above. The MSEs and biases

TABLE I

MSEs and Squared Biases (Based on 50,000 Simulations) of Different Estimators of $\Theta$ in the Model (1.2) for $p=2$; $q_1=5$, $q_2=5$; $\Sigma=I_2$; $\lambda=1$; $\Theta=\delta_1\left(\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right)$, and $\Theta=\delta_2 I_2$ for $\delta_1, \delta_2=0.5$, 1, and 4 (the Squared Biases Are Given in Parentheses)

| | $\delta_1$ | | | $\delta_2$ | | |
|---|---|---|---|---|---|---|
| | 0.5 | 1 | 4.0 | 0.5 | 1 | 4.0 |
| $\hat{\Theta}_U$ | 3.9884 | 6.3645 | 37.3221 | 3.9021 | 5.9980 | 31.1513 |
| $\hat{\Theta}_{\text{REML}}$ | 2.6711 | 5.0286 | 36.0903 | 2.4962 | 4.4812 | 29.7944 |
| | (0.2645) | (0.2296) | (0.1966) | (0.2158) | (0.1290) | (0.0143) |
| $\hat{\Theta}_{U+}$ | 2.6194 | 5.0037 | 36.1843 | 2.4283 | 4.4129 | 29.7504 |
| | (0.2489) | (0.2212) | (0.1986) | (0.1970) | (0.1152) | (0.0121) |
| $\hat{\Theta}_{(c_1,\varepsilon_1)+}$ | 1.4473 | 2.7812 | 22.8639 | 1.3098 | 2.3257 | 16.9524 |
| | (0.1438) | (0.2469) | (4.4644) | (0.0632) | (0) | (1.6733) |
| $\hat{\Theta}_{(c_1,\varepsilon_1)*}$ | 1.4572 | 2.7697 | 22.7403 | 1.3292 | 2.3399 | 16.9528 |
| | (0.1436) | (0.2316) | (4.3620) | (0.0682) | (0.0002) | (1.6654) |

were computed based on 50,000 simulations and the parameter values used are $\Sigma=I_2$ and $\Theta=\delta_1\mathbf{1}_2\mathbf{1}_2'$ and $\Theta=\delta_2 I_2$, for $\delta_1$, $\delta_2=0.5$, 1, and 4, where $\mathbf{1}_2$ denotes a $2\times 1$ vector of ones. In Table I, the squared biases are given in parentheses. From Table I, it is clear that some of our proposed nonnegative estimators provide substantial reduction in MSE over both $\hat{\Theta}_U$ and $\hat{\Theta}_{\text{REML}}$. The estimators that we recommend are $\hat{\Theta}_{(c_1,\varepsilon_1)+}$ and $\hat{\Theta}_{(c_1,\varepsilon_1)*}$. The difference between the MSEs (and the biases) of these two estimators is rather negligible.

*Nonnegative Estimation of* tr $\Theta$

We shall first characterize $c\geqslant 0$ and $\varepsilon\leqslant 1$ so that tr $\hat{\Theta}_{(c,\varepsilon)}$ has a uniformly smaller MSE than tr $\hat{\Theta}_U$. If the resultant estimator tr $\hat{\Theta}_{(c,\varepsilon)}$ is always nonnegative, nothing needs to be done. Otherwise, the estimator tr $\hat{\Theta}_{(c,\varepsilon)}$ can be truncated at zero to yield a satisfactory nonnegative estimator of tr $\Theta$. Towards this, we first compute the MSE of tr $\hat{\Theta}_{(c,\varepsilon)}$, which is given by

$$\begin{aligned}\text{MSE}(\operatorname{tr}\hat{\Theta}_{(c,\varepsilon)}) &= E(\operatorname{tr}\hat{\Theta}_{(c,\varepsilon)}-\operatorname{tr}\Theta)^2=\frac{2c^2}{\lambda^2 q_1}\operatorname{tr}(\Sigma+\lambda\Theta)^2+\frac{2c^2\varepsilon^2}{\lambda^2 q_2}\operatorname{tr}(\Sigma^2)\\ &\quad+\frac{c^2}{\lambda^2}(1-\varepsilon)^2(\operatorname{tr}\Sigma)^2+(c-1)^2(\operatorname{tr}\Theta)^2\\ &\quad+2\frac{c}{\lambda}(c-1)(1-\varepsilon)(\operatorname{tr}\Sigma)(\operatorname{tr}\Theta). \qquad (2.13)\end{aligned}$$

Using arguments similar to those that lead to Theorem 2.1, we see that the inequalities (2.14) and (2.15) below are necessary and sufficient for $\text{MSE}(\text{tr}\, \hat{\Theta}_{(c,\varepsilon)}) \leqslant \text{MSE}(\text{tr}\, \hat{\Theta}_U)$ uniformly in $\Sigma$ and $\Theta$:

$$\frac{2}{q_1}(c^2-1)+(c-1)^2\, p \leqslant 0 \tag{2.14}$$

$$2\left\{c^2\left(\frac{1}{q_1}+\frac{\varepsilon^2}{q_2}\right)-\left(\frac{1}{q_1}+\frac{1}{q_2}\right)\right\}+c^2(1-\varepsilon)^2\, p \leqslant 0. \tag{2.15}$$

We note that the conditions (2.14) and (2.15) are stronger compared to (2.6) and (2.7), since $p$ occurs above with a positive coefficient in both the inequalities. Equation (2.14) is obviously equivalent to

$$\max\left\{0, \frac{q_1 p-2}{q_1 p+2}\right\} \leqslant c \leqslant 1. \tag{2.16}$$

For any $c$ satisfying (2.16), an interval for $\varepsilon$ can be obtained from (2.15), Arguing as in the derivation of (2.11), we also conclude that there exists a nonnegative estimator of the form $\text{tr}\, \hat{\Theta}_{(c,\varepsilon)}$ having a uniformly smaller MSE compared to $\text{tr}\, \hat{\Theta}_U$ if and only if

$$q_1 p \geqslant 3,$$

and (2.17)

$$(q_1 p-2)^2\, q_2 \leqslant 2(q_1+q_2)(q_1 p+2) \qquad \text{or} \qquad q_1 p = 1 \text{ or } 2.$$

Values of $c$ and $\varepsilon$ that we recommend for practical use are obtained by minimizing the LHS of (2.14) with respect to $c$ and the LHS of (2.15) with respect to $\varepsilon$. These values, say $c_2$ and $\varepsilon_2$, are given by

$$c_2 = \frac{q_1 p}{q_1 p+2} \qquad \text{and} \qquad \varepsilon_2 = \frac{q_2 p}{q_2 p+2}. \tag{2.18}$$

The above choice of $c$ and $\varepsilon$ can be justified by noting that the LHSs of (2.14) and (2.15) are respectively the maximum values of the leading terms involving $\Theta$ and $\Sigma$ in the MSE of $\hat{\Theta}_{(c,\varepsilon)}$. Since $\varepsilon_2 > 0$, $\text{tr}\, \hat{\Theta}_{(c_2,\varepsilon_2)}$ can be negative, we consider the estimator

$$(\text{tr}\, \hat{\Theta}_{(c_2,\varepsilon_2)})_+ = \max\{0, \text{tr}\, \hat{\Theta}_{(c_2,\varepsilon_2)}\}. \tag{2.19}$$

In the numerical results reported in Table II, we have considered the estimators $\text{tr}\, \hat{\Theta}_U$, $\text{tr}\, \hat{\Theta}_{\text{REML}}$, $(\text{tr}\, \hat{\Theta}_U)_+$, $(\text{tr}\, \hat{\Theta}_{(c_1,\varepsilon_1)})_+$, and $(\text{tr}\, \hat{\Theta}_{(c_2,\varepsilon_2)})_+$, where $c_1, \varepsilon_1, c_2, \varepsilon_2$ are given (2.12) and (2.18), and $(\text{tr}\, \hat{\Theta}_U)_+$ and $(\text{tr}\, \hat{\Theta}_{(c_1,\varepsilon_1)})_+$ are defined analogously to $(\text{tr}\, \hat{\Theta}_{(c_2,\varepsilon_2)})_+$ in (2.19). The

TABLE II

MSEs and Squared Biases (Based on 50,000 Simulations) of Different Estimators of tr $\Theta$ in the Model (1.2) for $p=2$; $q_1=5$, $q_2=5$; $\Sigma=I_2$; $\lambda=1$; $\Theta=\delta_1\begin{pmatrix}1&1\\1&1\end{pmatrix}$, and $\Theta=\delta_2 I_2$ for $\delta_1, \delta_2=0.5$, 1, and 4 (the Squared Biases Are Given in Parentheses)

| | $\delta_1$ | | | $\delta_2$ | | |
|---|---|---|---|---|---|---|
| | 0.5 | 1 | 4.0 | 0.5 | 1 | 4.0 |
| $\mathrm{tr}(\hat{\Theta}_U)$ | 2.7942 | 4.7773 | 33.3734 | 2.6066 | 4.0113 | 20.8769 |
| $\mathrm{tr}(\hat{\Theta}_{\mathrm{REML}})$ | 2.2448<br>(0.4979) | 4.1305<br>(0.3976) | 32.6933<br>(0.2719) | 1.9468<br>(0.4316) | 3.1141<br>(0.2578) | 19.7977<br>(0.0281) |
| $(\mathrm{tr}\,\hat{\Theta}_U)_+$ | 1.9198<br>(0.0600) | 4.0052<br>(0.0185) | 33.0535<br>(0) | 1.7736<br>(0.0560) | 3.3546<br>(0.0145) | 20.7681<br>(0.0001) |
| $(\mathrm{tr}\,\hat{\Theta}_{(c_1,\varepsilon_1)})_+$ | 1.0810<br>(0.0320) | 2.1249<br>(0.0187) | 20.3387<br>(3.5501) | 0.9970<br>(0.0312) | 1.7623<br>(0.0195) | 13.9558<br>(3.5148) |
| $(\mathrm{tr}\,\hat{\Theta}_{(c_2,\varepsilon_2)})_+$ | 1.4322<br>(0.0505) | 2.8423<br>(0) | 24.0235<br>(1.1235) | 1.3233<br>(0.0484) | 2.3654<br>(0) | 15.4018<br>(1.1054) |

parameter values considered are the same as in Table I and the computations in Table II are based on 50,000 simulations. From Table II, it is clear that a singificant reduction in MSE can be achieved over both tr $\hat{\Theta}_U$ and tr $\hat{\Theta}_{\mathrm{REML}}$, and $(\mathrm{tr}\,\hat{\Theta}_{(c_1,\varepsilon_1)})_+$ appears to be a satisfactory nonnegative estimator of tr $\Theta$.

*Nonnegative Estimation of* $\mathbf{a}'\Theta\mathbf{a}$, $\mathbf{a}\neq 0$

As before, we shall characterize $c$ and $\varepsilon$ so that $\mathbf{a}'\hat{\Theta}_{(c,\varepsilon)}\mathbf{a}$ has a uniformly smaller MSE comparted to $\mathbf{a}'\hat{\Theta}_U\mathbf{a}$ for estimating $\mathbf{a}'\Theta\mathbf{a}$. Note that the estimators we are considering are linear combinations of $\mathbf{a}'S_1\mathbf{a}$ and $\mathbf{a}'S_2\mathbf{a}$. From (1.2), it is clear that

$$\mathbf{a}'S_1\mathbf{a} \sim (\mathbf{a}'\Sigma\mathbf{a}+\lambda\mathbf{a}'\Theta\mathbf{a})\,\chi^2_{q_1} \quad \text{and} \quad \mathbf{a}'S_2\mathbf{a} \sim (\mathbf{a}'\Sigma\mathbf{a})\,\chi^2_{q_2}. \tag{2.20}$$

Consequently, the problem of estimating $\mathbf{a}'\Theta\mathbf{a}$ is a univariate problem in the model (2.20) that involves the two variance components $\mathbf{a}'\Sigma\mathbf{a}$ and $\mathbf{a}'\Theta\mathbf{a}$. Thus the results in the univariate case directly apply for the nonnegative estimation of $\mathbf{a}'\Theta\mathbf{a}$. In particular, the conditions (2.6) and (2.7) are necessary and sufficient for $\mathbf{a}'\hat{\Theta}_{(c,\varepsilon)}\mathbf{a}$ to have a uniformly smaller MSE comparted to $\mathbf{a}'\hat{\Theta}_U\mathbf{a}$.

In Table III, we have compared several competing estimators for estimating $\mathbf{1}'\Theta\mathbf{1}$. The estimators considered are $\mathbf{1}'\hat{\Theta}_U\mathbf{1}$, $\mathbf{1}'\hat{\Theta}_{\mathrm{REML}}\mathbf{1}$, $(\mathbf{1}'\hat{\Theta}_U\mathbf{1})_+$, and $(\mathbf{1}'\hat{\Theta}_{(c_1,\varepsilon_1)}\mathbf{1})_+$, where, for an estimator $\hat{\Theta}$ of $\Theta$, $(\mathbf{1}'\hat{\Theta}\mathbf{1})_+ = \max\{0, \mathbf{1}'\hat{\Theta}\mathbf{1}\}$. The same models and parameter values were considered as

TABLE III

MSEs and Squared Biases (Based on 50,000 Simulations) of Different Estimators of $\mathbf{1}'\Theta\mathbf{1}$ in the Model (1.2) for $p=2$; $q_1=5$, $q_2=5$; $\Sigma=I_2$; $\lambda=1$; $\Theta=\delta_1\binom{1\ 1}{1\ 1}$, and $\Theta=\delta_2 I_2$ for $\delta_1, \delta_2=0.5$, 1, and 4 (the Squared Biases Are Given in Parentheses)

| | $\delta_1$ | | | $\delta_2$ | | |
|---|---|---|---|---|---|---|
| | 0.5 | 1 | 4.0 | 0.5 | 1 | 4.0 |
| $\mathbf{1}'\hat{\Theta}_U\mathbf{1}$ | 7.9152 | 15.8332 | 130.1574 | 5.1507 | 7.9235 | 41.2293 |
| $\mathbf{1}'\hat{\Theta}_{\mathrm{REML}}\mathbf{1}$ | 6.1201 | 13.6800 | 128.7449 | 3.4458 | 5.9953 | 39.3690 |
| | (0.2802) | (0.1461) | (0.0300) | (0.4247) | (0.2506) | (0.0227) |
| $(\mathbf{1}'\hat{\Theta}_U\mathbf{1})_+$ | 6.1032 | 14.2766 | 129.5631 | 3.3160 | 6.1125 | 40.1326 |
| | (0.0836) | (0.0218) | (0) | (0.1847) | (0.0836) | (0.0024) |
| $(\mathbf{1}'\hat{\Theta}_{(c_1,\varepsilon_1)}\mathbf{1})_+$ | 3.2220 | 7.8444 | 83.3603 | 1.8138 | 3.2251 | 23.9348 |
| | (0.0061) | (0.4860) | (17.4442) | (0.0674) | (0.0060) | (3.5024) |

in Tables I and II, and the computations are based on 50,000 simulations. Once again, the computations indicate that some of the proposed estimators are superior to both $\mathbf{1}'\hat{\Theta}_U\mathbf{1}$ and $\mathbf{1}'\hat{\Theta}_{\mathrm{REML}}\mathbf{1}$. The estimator of $\mathbf{1}'\Theta\mathbf{1}$ that we recommend is $(\mathbf{1}'\hat{\Theta}_{(c_1,\varepsilon_1)}\mathbf{1})_+$.

## 3. Estimation of the Error Variance Component $\Sigma$

In this section we discuss the problem of estimation of the error variance component $\Sigma$, based on $S_1$ and $S_2$ defined in (1.2). As noted before, the UMVUE of $\Sigma$ is given by $\hat{\Sigma}_U = S_2/q_2$. To describe the restricted maximum likelihood estimator $\hat{\Sigma}_{\mathrm{REML}}$ of $\Sigma$, we use the representation $S_1 = T'\Lambda T$ and $S_2 = T'T$ used in (1.5), where $T$ is a nonsingular matrix and $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_p)$ is a diagonal matrix. Let $D = \mathrm{diag}(d_1, ... d_p)$ be a diagonal matrix with $d_i = 1/q_2$ if $\lambda_i/q_1 \geqslant 1/q_2$ and $d_i = (\lambda_i + 1)/(q_1 + q_2)$ if $\lambda_i/q_1 < 1/q_2$. Then $\hat{\Sigma}_{\mathrm{REML}}$ is given by (see Calvin and Dykstra, 1991b, p. 859)

$$\hat{\Sigma}_{\mathrm{REML}} = T'\,DT. \tag{3.1}$$

In the sequel we consider two types of invariant loss functions, the entropy loss and the squared error loss:

$$L_1(\hat{\Sigma}, \Sigma) = \mathrm{tr}\ \hat{\Sigma}\Sigma^{-1} - \ln|\hat{\Sigma}\Sigma^{-1}| - p \tag{3.2}$$

$$L_2(\hat{\Sigma}, \Sigma) = \mathrm{tr}(\hat{\Sigma}\Sigma^{-1} - I)^2. \tag{3.3}$$

Proceeding in the same spirit as in Mathew *et al.* (1992b), we claim that

$$\hat{\Sigma}_{(1)} = \begin{cases} \dfrac{S_1 + S_2}{q_1 + q_2 - (p-1)} & \text{if } |I + S_2^{-1} S_1| \leqslant \dfrac{q_1 + q_2 - (p-1)}{q_2} \\ \dfrac{S_2}{q_2} & \text{otherwise} \end{cases} \tag{3.4}$$

provides uniform risk improvement over $S_2/q_2$ under $L_1$. Note that $\hat{\Sigma}_{(1)}$ can equivalently be expressed as

$$\hat{\Sigma}_{(1)} = \begin{cases} \dfrac{T'(I+\Lambda)T}{q_1 + q_2 - (p-1)} & \text{if } \prod_{i=1}^{p} (1+\lambda_i) \leqslant \dfrac{q_1 + q_2 - (p-1)}{q_2} \\ \dfrac{T'T}{q_2} & \text{otherwise} \end{cases}. \tag{3.5}$$

Analogously, it is claimed that

$$\hat{\Sigma}_{(2)} = \begin{cases} \dfrac{S_1 + S_2}{q_1 + q_2 - (p-3)} & \text{if } |I + S_2^{-1} S_1| \leqslant \dfrac{q_1 + q_2 - (p-3)}{q_2 + 2} \\ \dfrac{S_2}{q_2 + 2} & \text{otherwise} \end{cases} \tag{3.6}$$

is uniformly better than $S_2/(q_2+2)$ (the best multiple of $S_2$ under $L_2$) under the squared error loss $L_2$, and hence is uniformly better than $\hat{\Sigma}_U$. $\hat{\Sigma}_{(2)}$ can equivalently be expressed as

$$\hat{\Sigma}_{(2)} = \begin{cases} \dfrac{T'(I+\Lambda)T}{q_1 + q_2 - (p-3)} & \text{if } \prod_{i=1}^{p} (1+\lambda_i) \leqslant \dfrac{q_1 + q_2 - (p-3)}{q_2 + 2} \\ \dfrac{T'T}{q_2 + 2} & \text{otherwise} \end{cases}. \tag{3.7}$$

We have been able to establish the risk dominance of $\hat{\Sigma}_{(1)}$ and $\hat{\Sigma}_{(2)}$ over $\hat{\Sigma}_U$ only in the bivariate case, i.e., $p = 2$. Unfortunately, the proof is purely algebraic in nature and lacks an immediate generalization for $p > 2$. It is interesting to compare $\hat{\Sigma}_{\text{REML}}$, $\hat{\Sigma}_{(1)}$ and $\hat{\Sigma}_{(2)}$, and note how a drastic modification of $\hat{\Sigma}_{\text{REML}}$ results in uniform risk improvement over $\hat{\Sigma}_U$.

To derive $\hat{\Sigma}_{(1)}$ and $\hat{\Sigma}_{(2)}$, we decompose $S_1$ and $S_2$ as

$$S_1 = T_1 T_1', \qquad S_2 = T_2 T_2', \tag{3.8}$$

where $T_1 = ((t_{ij(1)}))$ and $T_2 = ((t_{ij(2)}))$ are lower triangular matrices with positive diagonal elements. Consider now an estimator of $\Sigma$ of the form

$$\hat{\Sigma} = T_2 \psi(UU') T_2', \tag{3.9}$$

where $U = T_2^{-1}T_1$ and $\psi(\cdot)$ is a $p \times p$ p.d. matrix-valued function of the matrix argument $UU'$. $\hat{\Sigma}$ above can be motivated along the lines of Stein (1964), Sinha (1976), Shorrock and Zidek (1976), and Sinha and Ghosh (1987). Note that $\psi(UU') = I/q_2$ results in $\hat{\Sigma}_U$.

Now consider the loss function $L_1$ and the problem of optimal choice of $\psi(UU')$ in (3.9) to minimize the resultant risk of $\hat{\Sigma}$. A standard conditional argument (see Sinha and Ghosh, 1987) then yields that

$$\psi_{\text{opt}}(UU') = \{E_{\Sigma,\Theta}(T_2'T_2 \mid U)\}^{-1}. \tag{3.10}$$

The result given in Lemma 3.1, providing an upper bound of $\psi_{\text{opt}}(UU')$ is crucial for the proof of (3.4). We have been able to prove Lemma 3.1 only for $p = 2$.

LEMMA 3.1.

$$\psi_{\text{opt}}(UU') \leqslant \frac{|I + UU'|\, I_p}{q_1 + q_2 - (p-1)},$$

*whathever be $\Sigma$ and $\Theta$.*

Using the convexity of $L_1$, it then follows that given $\psi(UU') = I/q_2$ and hence $\hat{\Sigma}_U$, use of $\hat{\Sigma}_{(1)} = T_2\psi_{(1)}(UU')\,T_2'$ where

$$\psi_{(1)}(UU') = \begin{cases} \dfrac{I + UU'}{q_1 + q_2 - (p-1)} & \text{if } |I + UU'| \leqslant \dfrac{q_1 + q_2 - (p-1)}{q_2} \\ \dfrac{I}{q_2} & \text{otherwise} \end{cases} \tag{3.11}$$

provides uniform improvement over $\hat{\Sigma}_U$. Since $\hat{\Sigma}_{(1)}$ defined above coincides with the definition given in (3.4), we have proved the claim.

*Proof of Lemma* 3.1. *for* $p = 2$. Recall that $S_1 \sim W_p(\Sigma + \lambda\Theta, q_1)$ independent of $S_2 \sim W_p(\Sigma, q_2)$. Due to the invariance of the loss function $L_1$ (and also $L_2$), without any loss of generality we can take $\Sigma = I_p$ and $\Theta$ to be diagonal so that we can write

$$\{\Sigma + \lambda\Theta\}^{-1} = L = \begin{pmatrix} l_{11} & 0 \\ 0 & l_{22} \end{pmatrix}, \qquad 0 \leqslant l_{ii} \leqslant 1, \quad i = 1, 2.$$

Let $S \sim W_p(\Sigma, \nu)$, $\nu \geqslant p$, and write $S = TT'$ where $T$ is a lower triangular matrix with $T = ((t_{ij}))$, $t_{ii} > 0$, $t_{ij} = 0, j > i, i, j = 1, ..., p$. Then the p.d.f. of $T$ can be written as (see Anderson, 1984)

$$f(T) = Ke^{-(1/2)\operatorname{tr}\Sigma^{-1}TT'}\left(\prod_{i=1}^{p} t_{ii}^{\nu - i}\right), \tag{3.12}$$

where here and throughout below $K$ is used to denote a generic constant (in some cases $K$ may depend on some unknown parameters). Denote by $T_1 = ((t_{ij(1)}))$ and $T_2 = ((t_{ij(2)}))$ the appropriate lower triangular matrices with positive diagonal elements corresponding to $S_1$ and $S_2$, respectively. The joint p.d.f. of $T_1$ and $T_2$ can be obtained using (3.12). Making the transformation $(T_1, T_2) \to (U \equiv T_2^{-1} T_1, T_2)$ and noting that the Jacobian $= \prod_{i=1}^{p} t_{ii(2)}^{i}$, we get the joint p.d.f. of $U = ((u_{ij}))$ and $T_2$ as

$$f(U, T_2) = K e^{-(1/2)\operatorname{tr}[T_2 T_2' + L T_2 U U' T_2']} \left(\prod_{1}^{p} u_{ii}^{q_1 - i}\right)\left(\prod_{1}^{p} t_{ii(2)}^{q_1 + q_2 - i}\right),$$

which results in the following conditional p.d.f. of $T_2$, given $U$:

$$f(T_2 \mid u) = K \exp[-\tfrac{1}{2}\operatorname{tr}\{T_2 T_2' + L T_2 V T_2'\}] \left(\prod_{1}^{p} t_{ii(2)}^{q_1 + q_2 - i}\right). \tag{3.13}$$

Here $V = uu' = ((v_{ij}))$. We shall assume $p = 2$ in the remainder of the proof. It is easy to conclude from (3.13) that $t_{11(2)}$ is independent of $(t_{21(2)}, t_{22(2)})$, and

$$t_{11(2)} \sim f(t_{11(2)}) = K \exp\left(-\frac{t_{11(2)}^2}{2}(1 + v_{11} l_{11})\right) t_{11(2)}^{\nu_1},$$

$$t_{21(2)} \mid t_{22(2)} \sim N\left[-\frac{t_{22(2)} v_{12} l_{22}}{1 + v_{11} l_{22}}, \frac{1}{1 + v_{11} l_{22}}\right]$$

$$t_{22(2)} \sim f(t_{22(2)}) = K \exp\left(-\frac{t_{22(2)}^2}{2} \frac{|I + l_{22} V|}{1 + v_{11} l_{22}}\right) t_{22(2)}^{\nu_2}, \tag{3.14}$$

where $\nu_1 = q_1 + q_2 - 1$, $\nu_2 = q_1 + q_2 - 2$. Direct computation then yields

$$E(T_2' T_2 \mid u)$$
$$= \begin{bmatrix} \dfrac{1}{1 + v_{11} l_{22}} + \dfrac{\nu_1 + 1}{1 + v_{11} l_{11}} + \dfrac{v_{12}^2 l_{22}^2 (\nu_2 + 1)}{(1 + v_{11} l_{22})\,|I + l_{22} V|} & -\dfrac{v_{12} l_{22} (\nu_2 + 1)}{|I + l_{22} V|} \\ -\dfrac{v_{12} l_{22} (\nu_2 + 1)}{|I + l_{22} V|} & \dfrac{(\nu_2 + 1)(1 + v_{11} l_{22})}{|I + l_{22} V|} \end{bmatrix}. \tag{3.15}$$

To establish Lemma 3.1, we now show that

$$E(T_2' T_2 \mid u) \geqslant \frac{\nu_1}{|I + V|} \cdot I_2, \qquad \text{whatever be } l_{11}, l_{22}. \tag{3.16}$$

In order to prove (3.16), it is enough to show that the smaller eigenvalue of $E(T_2'T_2 \mid u)$ is greater than or equal to $v_1/|I+V|$, whatever be $l_{11}, l_{22}$, or equivalently,

$$\inf_{0 \leqslant l_{11}, l_{22} \leqslant 1} \inf_{\mathbf{x}:\, \|\mathbf{x}\|=1} [\mathbf{x}'E(T_2'T_2 \mid u)\,\mathbf{x}] \geqslant \frac{v_1}{|I+V|}. \tag{3.17}$$

Upon simplyfing $\mathbf{x}'E(T_2'T_2 \mid u)\,\mathbf{x}$ using (3.15), it is readily verified that

$$\mathbf{x}'E(T_2'T_2 \mid u)\,\mathbf{x} \geqslant v_1 \phi(x, l_{22}), \tag{3.18}$$

where

$$\phi(x, l_{22}) = \frac{x^2}{1+v_{11}} + \left\{\frac{x v_{12} l_{22}}{\sqrt{1+v_{11}l_{22}}} - \sqrt{1-x^2}\sqrt{1+v_{11}l_{22}}\right\}^2 \frac{1}{|I+V|}, \tag{3.19}$$

and we write $\mathbf{x} = (x, \sqrt{1-x^2})$, $0 \leqslant x \leqslant 1$ (using the fact that $\|\mathbf{x}\| = 1$). Thus, in order to establish (3.17), it is enough to show that

$$\phi(x, l_{22}) - \frac{1}{|I+V|} \geqslant 0,$$

for all $x$ and $l_{22}$ satisfying $0 \leqslant x \leqslant 1$ and $0 \leqslant l_{22} \leqslant 1$. Writing $A = v_{22}/(1+v_{11}) + v_{12}^2 l_{22}^2/(1+v_{11}l_{22})$, $t = x/\sqrt{1-x^2}$, and using $v_{11}v_{22} \geqslant v_{12}^2$, it can be verified that

$$\begin{aligned}\phi(x, l_{22}) - \frac{1}{|I+V|} &\geqslant \frac{1-x^2}{|I+V|}\{At^2 - 2v_{12}l_{22}t + v_{11}l_{22}\} \\ &= \frac{1-x^2}{|I+V|}\left\{A\left(t - \frac{v_{12}l_{22}}{A}\right)^2 + v_{11}l_{22} - \frac{v_{12}^2 l_{22}^2}{A}\right\}.\end{aligned} \tag{3.20}$$

The nonnegativity of the last expression in (3.20) can be established by showing that $v_{11}l_{22} - v_{12}^2 l_{22}^2/A \geqslant 0$ (using $v_{11}v_{22} \geqslant v_{12}^2$). This completes the proof of Lemma 3.1.

We next consider the loss function $L_2$. The risk of $\hat{\Sigma}$ defined in (3.9) can be computed as

$$\begin{aligned} &E_{\Sigma,\Theta}\{L_2(\hat{\Sigma}, \Sigma)\} \\ &\quad = E_{\Sigma,\Theta}[\operatorname{tr}(T_2\psi(UU')\,T_2'\Sigma^{-1} - I)^2] \\ &\quad = E_{I,\Theta^*}\operatorname{tr}[T_2\psi(UU')\,T_2'T_2\psi(UU')\,T_2' - 2T_2\psi(UU')\,T_2' + I] \\ &\quad = E_{I,\Theta^*}[\operatorname{tr}\psi(UU')\,E_{I,\Theta^*}\{T_2'T_2\psi(UU')\,T_2'T_2 \mid U\} \\ &\qquad - 2\operatorname{tr}\psi(UU')\,E_{I,\Theta^*}\{T_2'T_2 \mid U\} + p], \end{aligned} \tag{3.21}$$

where the inner conditional expectation is evaluated for fixed $U$, the outer expectation is evaluated with respect to $U$, and $\Theta^* = \lambda\Sigma^{-1/2}\Theta\Sigma^{-1/2}$. As in the previous case, the problem now is to compute $\psi_{\text{opt}}(UU')$ and identify an upper bound of it, similar to the one given in Lemma 3.1. We claim that $\psi_{\text{opt}}(UU')$ satisfies the inequality given below.

LEMMA 3.2.

$$\psi_{\text{opt}}(UU') \leqslant \frac{|I+UU'|}{q_1+q_2-(p-3)} I_p, \tag{3.22}$$

*whatever be $\Theta$ and $\Sigma$.*

A proof of Lemma 3.2 is given in Mathew *et al.* (1992c) for $p=2$. The proof is computationally involved and is omitted here. As before, using the convexity of $L_2$, it then follows that, given $\psi(UU') = I/q_2$ resulting in $\hat{\Sigma}_U$, use of $\hat{\Sigma}_{(2)} = T_2\psi_{(2)}(UU')\,T_2'$ where

$$\psi_{(2)}(UU') = \begin{cases} \dfrac{I+UU'}{q_1+q_2-(p-3)} & \text{if } |I+UU'| \leqslant \dfrac{q_1+q_2-(p-3)}{q_2+2} \\[2ex] \dfrac{I}{q_2+2} & \text{otherwise} \end{cases} \tag{3.23}$$

yields uniform improvement over $S_2/(q_2+2)$. Since $\hat{\Sigma}_{(2)}$ coincides with $\hat{\Sigma}_{(2)}$ defined in (3.6), the claim is established for $p=2$.

TABLE IV

Risks (Based on 50,000 Simulations) of Different Estimators of $\Sigma$ in the Model (1.2) for $p=2$; $q_1=5$, $q_2=5$; $\Sigma=I_2$; $\lambda=1$; $\Theta=\delta_1\begin{pmatrix}1&1\\1&1\end{pmatrix}$, and $\Theta=\delta_2 I_2$ for $\delta_1, \delta_2 = 0.01$, 0.5, 4, and 9

| | $\delta_1$ | | | | $\delta_2$ | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.01 | 0.5 | 4.0 | 9.0 | 0.01 | 0.5 | 4.0 | 9.0 |
| | (a) Risks of different estimators of $\Sigma$ for the loss function $L_1$ | | | | | | | |
| $\hat{\Sigma}_{(1)}$ | 0.7023 | 0.7074 | 0.7098 | 0.7100 | 0.7023 | 0.7076 | 0.7100 | 0.7100 |
| $\hat{\Sigma}_{\text{REML}}$ | 0.6176 | 0.6256 | 0.6475 | 0.6412 | 0.6176 | 0.6247 | 0.6801 | 0.6991 |
| $S_2/q_2$ | 0.7100 | 0.7100 | 0.7100 | 0.7100 | 0.7100 | 0.7100 | 0.7100 | 0.7100 |
| | (b) Risks of different estimators of $\Sigma$ for the loss function $L_2$ | | | | | | | |
| $\hat{\Sigma}_{(2)}$ | 0.7742 | 0.7775 | 0.7788 | 0.7789 | 0.7741 | 0.7775 | 0.7789 | 0.7789 |
| $\hat{\Sigma}_{\text{REML}}$ | 0.6557 | 0.7214 | 0.8350 | 0.8533 | 0.6557 | 0.7334 | 1.0477 | 1.1490 |
| $S_2/(q_2+2)$ | 0.7789 | 0.7789 | 0.7789 | 0.7789 | 0.7789 | 0.7789 | 0.7789 | 0.7789 |

In Table IV, we give the risks of different competing estimators of $\Sigma$ in the bivariate case for $q_1 = 5$, $q_2 = 5$. The parameter values considered are the same as in Tables I–III. In Table IVa, we have considered the loss function $L_1$ given in (3.2) and have reported the risks of $\hat{\Sigma}_{(1)}$, $\hat{\Sigma}_{\text{REML}}$, and $\hat{\Sigma}_U = S_2/q_2$. The loss function $L_2$, given in (3.3), is considered in Table IVb, and we have reported the risks of $\hat{\Sigma}_{(2)}$, $\hat{\Sigma}_{\text{REML}}$, and $S_2/(q_2+2)$. The risks of $S_2/q_2$ and $S_2/(q_2+2)$ are clearly independent of $\Theta$.

The numerical results in Table II indicate that $\hat{\Sigma}_{\text{REML}}$ does have some edge over both $\hat{\Sigma}_{(1)}$ and $S_2/q_2$ with respect to the loss function $L_1$. However, with respect to the loss function $L_2$, $\hat{\Sigma}_{\text{REML}}$ provides improvement over $S_2/(q_2+2)$ only for very small values of $\Theta$. For large values of $\Theta$, $\hat{\Sigma}_{\text{REML}}$ performs much worse than both $\hat{\Sigma}_{(2)}$ and $S_2/(q_2+2)$.

## Acknowledgment

Our sincere thanks are due to the referees for some helpful comments.

## References

[1] Amemiya, Y. (1985). What should be done when an estimated between-group covariance matrix is not nonnegative definite? *Amer. Statst.* **39** 112–117.

[2] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

[3] Calvin, J. A., and Dykstra, R. L. (1991a). Maximum likelihood estimation of a set of covariance matrices under Loewner order restrictions with applications to balanced multivariate variance components models. *Ann. Statist.* **19** 850–869.

[4] Calvin, J. A., and Dykstra, R. L. (1991b). Least squares estimation of covariance matrices in balanced multivariate variance components models. *J. Amer. Statist. Assoc.* **86** 388–395.

[5] Calvin, J. A., and Sedransk, J. (1991). Bayesian and frequentist predictive inference for the patterns of care studies. *J. Amer. Statist. Assoc.* **86** 36–48.

[6] Kelly, R. J., and Mathew, T. (1993). Improved estimators of variance components having smaller probability of negativity. *J. Roy. Statist. Soc. Ser. B* **55** 897–911.

[7] Mathew, T., Sinha, B. K., and Sutradhar, B. (1992a). Nonnegative estimation of random effects variance components in balanced mixed models. In *Nonparametric Statistics and Related Topics* (A. K. Md. Saleh, Ed.), pp. 281–295. North-Holland, New York.

[8] Mathew, T., Sinha, B. K., and Sutradhar, B. (1992b). Improved estimation of error variance in general balanced mixed models. *Statist. Decisions* **10** 227–238.

[9] Mathew, T., Niyogi, A., and Sinha, B. K. (1992c). "Improved Nonnegative Estimation of Variance Components in Balanced Multivariate Mixed Models." Technical Report, Department of Mathematics and Statistics, University of Maryland–Baltimore County, Baltimore, Maryland.

[10] MUIRHEAD, R. J. (1982). *Aspects if Multivariate Statical Theory*. Wiley, New York.

[11] SHORROCK, R. W., AND ZIDEK, J. V. (1976). An improved estimator of the generalized variance. *Ann. Statist.* **4** 629–638.

[12] SINHA, B. K. (1976). On improved estimators of the generalized variance. *J. Multivariate Anal.* **6** 617–625.

[13] SINHA, B. K., AND GHOSH, M. (1987). Inadmissiblity of the best equivariant estimators of the variance–covariance matrix, the precision matrix and the generalized variance under entropy loss. *Statist. Decisions* **5** 201–227.

[14] STEIN, C. (1964). Inadmissiblity of the usual estimator of the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160.