CIRJE-F-38

Improved Nonnegative Estimation of Multivariate Components of Variance

M. S. Srivastava and T. Kubokawa
(University of Toronto) (The University of Tokyo)

January 1999

Improved Nonnegative Estimation of Multivariate Components of Variance

M.S. Srivastava and T. Kubokawa
University of Toronto and *University of Tokyo*

In this paper, we consider a multivariate one-way random effect model with equal replications. We propose non-negative definite estimators for 'between' and 'within' components of variance. Under the Stein loss function / Kullback-Leibler distance function, these estimators are shown to be better than the corresponding unbiased estimators. In particular, it is shown that the proposed restricted maximum likelihood estimator performs better than the unbiased as well as the truncated estimators proposed in this paper. Minimax and order-preserving minimax estimators are also proposed.

Key words and phrases: Random effects model, Stein loss, minimax and unbiased estimators, restricted maximum likelihood estimator.

AMS subject classifications: Primary 62H12, 62F30, Secondary 62C12, 62C20.

1 Introduction

The estimation of variance components in univariate mixed linear models have been considered extensively in the literature and various results are available. For example, Rao and Kleffe (1988) provide an exhaustive account of Rao's MINQUE theory. Other important contributions are due to Thompson (1962), Patterson and Thompson (1971,1975), Searle (1971) and Harville (1977) who considered maximum likelihood and restricted maximum likelihood methods. However, since unbiased estimators of 'between' components of variance take negative values with positive probability, considerable attension has also been paid to provide positive estimators for 'between' components. Nonnegative estimators improving upon the unbiased estimators have also been derived by Mathew, Sinha and Sutradhar (1992) and Kubokawa (1995) from a frequentist-view point.

On the other hand, the estimation of variance components in multivariate mixed linear model did not receive such an attention primarily due to technical difficulties encountered in obtaining similar results. For example, Amemiya (1985) proposed a restricted maximum likelihood estimator for the 'between' components but it is not known whether it is better (in any sense) than the usual unbiased estimator other than it is n.n.d. Similarly, Calvin and Dykstra (1991) proposed estimators for the ordered covariances but nothing is known about the properties of these estimators. Calvin and Dykstra (1991) also mentions the computational difficulties encountered with the corresponding MINQUE theory given in Rao and Kleffe (1988). Recently Mathew, Niyogi and Sinha (1994) considered one-way random effect model with equal replications and proposed some shrinkage estimators but the dominance result over the unbiased estimator remained open. Thus, no analytical results are available in the literature for the dominance over $\widehat{\Sigma}_1^{UB}$.

In this paper, we also consider one-way random effect model with equal replications:

$$\boldsymbol{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{a}_i + \boldsymbol{e}_{ij}, \quad i = 1, \ldots, k, \ j = 1, \ldots, r, \tag{1.1}$$

where $\boldsymbol{a}_i$'s and $\boldsymbol{e}_{ij}$'s are independent random variables, $\boldsymbol{a}_i$ having p-variate normal distribution with mean $\mathbf{0}$ and covariance matrix Σ_A, $\mathcal{N}_p(\mathbf{0}, \Sigma_A)$ and $\boldsymbol{e}_{ij}$ having $\mathcal{N}_p(\mathbf{0}, \Sigma_1)$. Here $\boldsymbol{\mu} \in \boldsymbol{R}^p$ is an unknown common mean vector and Σ_A and Σ_1 are unknown covariance matrices. Let $\overline{\boldsymbol{y}}_{i.} = r^{-1}\sum_{j=1}^r \boldsymbol{y}_{ij}$, $\overline{\boldsymbol{y}}_{..} = (rk)^{-1}\sum_{i=1}^k\sum_{j=1}^r \boldsymbol{y}_{ij}$, $\boldsymbol{S}_1 = \sum_{i=1}^k\sum_{j=1}^r(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_{i.})(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_{i.})'$ and $\boldsymbol{S}_2 = r\sum_{i=1}^k(\overline{\boldsymbol{y}}_{i.} - \overline{\boldsymbol{y}}_{..})(\overline{\boldsymbol{y}}_{i.} - \overline{\boldsymbol{y}}_{..})'$. The statistics $\overline{\boldsymbol{y}}_{..}$, $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are the minimal sufficient and are mutually independently distributed as $\overline{\boldsymbol{y}}_{..} \sim \mathcal{N}_p(\boldsymbol{\mu}, (rk)^{-1}(\Sigma_1 + r\Sigma_A))$,

$$\boldsymbol{S}_1 \sim \mathcal{W}_p(\Sigma_1, n_1) \quad \text{and} \quad \boldsymbol{S}_2 \sim \mathcal{W}_p(\Sigma_2, n_2) \tag{1.2}$$

for

$$\Sigma_2 = \Sigma_1 + r\Sigma_A, \quad n_1 = k(r-1) \quad \text{and} \quad n_2 = k - 1, \tag{1.3}$$

where $\mathcal{W}_p(\Sigma_1, n_1)$ designates the p-variate Wishart distribution with expectation $n_1\Sigma_1$ and n_1 degrees of freedom.

In Section 2, the estimation of the 'within' multivariate component of variance is addressed under the Stein (or entropy) loss function. The usual unbiased estimator of Σ_1, $\widehat{\Sigma}_1^{UB} = n_1^{-1}\boldsymbol{S}_1$, can be improved on by using the information from the order restriction $\Sigma_1 \le \Sigma_2$. This issue was discussed by Mathew *et al.* (1994) who considered the estimator

$$\widehat{\Sigma}_1^{MNS} = \begin{cases} (n_1 + n_2 - p + 1)^{-1}(\boldsymbol{S}_1 + \boldsymbol{S}_2) & \text{if } |\boldsymbol{I} + \boldsymbol{S}_1^{-1}\boldsymbol{S}_2| \le (n_1 + n_2 - p + 1)/n_1 \\ n_1^{-1}\boldsymbol{S}_1 & \text{otherwise,} \end{cases} \tag{1.4}$$

and show that $\widehat{\Sigma}_1^{MNS}$ dominates $\widehat{\Sigma}_1^{UB}$ in the case of $p = 2$. But their arguments leading to the showing of dominance of (1.4) over the unbiased estimator are not clear to us. For example, from their Lemma 3.1, it appears to us that the only claim that can be made is that the estimator defined by

$$\widehat{\Sigma}_1^{*} = \begin{cases} (n_1 + n_2 - p + 1)^{-1}|\boldsymbol{I} + \boldsymbol{S}_1^{-1}\boldsymbol{S}_2|\boldsymbol{S}_1 & \text{if } |\boldsymbol{I} + \boldsymbol{S}_1^{-1}\boldsymbol{S}_2| \le (n_1 + n_2 - p + 1)/n_1 \\ n_1^{-1}\boldsymbol{S}_1 & \text{otherwise} \end{cases} \tag{1.5}$$

dominates $\widehat{\Sigma}_1^{UB}$ when $p = 2$.

The estimator (1.5) was obtained by using the so called 'pivot' $\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$ whose statistical properties are difficult to obtain. In this paper, we consider instead the statistic $\boldsymbol{S}_2^{-1/2}\boldsymbol{S}_1\boldsymbol{S}_2^{-1/2}$ and propose estimators of the type

$$\widehat{\Sigma}_1(\Psi) = \boldsymbol{S}_2^{1/2}\boldsymbol{P}\Psi(\Lambda)\boldsymbol{P}'\boldsymbol{S}_2^{1/2},$$

where $\boldsymbol{S}_2^{1/2}$ is a symmetric matrix such that $\boldsymbol{S}_2 = (\boldsymbol{S}_2^{1/2})^2$, $\Psi(\Lambda) = \mathrm{diag}(\psi_1(\Lambda), \dots, \psi_p(\Lambda))$ and $\boldsymbol{P}$ is an orthogonal $p \times p$ matrix such that

$$\begin{aligned}\boldsymbol{P}'\boldsymbol{S}_2^{-1/2}\boldsymbol{S}_1\boldsymbol{S}_2^{-1/2}\boldsymbol{P} &= \Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_p) \\ &= \mathrm{diag}(\lambda_i, i = 1, \dots, p).\end{aligned}$$

For example, if $\Psi(\Lambda) = (n_1 + n_2)^{-1}(\boldsymbol{I} + \Lambda)$ for $\Lambda \geq (n_1/n_2)\boldsymbol{I}$; $\Psi(\Lambda) = n_1^{-1}\Lambda$ otherwise, then

$$\widehat{\Sigma}_1(\Psi) = \begin{cases} (n_1 + n_2)^{-1}(\boldsymbol{S}_1 + \boldsymbol{S}_2) & \text{if } \boldsymbol{I} + \boldsymbol{S}_1^{-1}\boldsymbol{S}_2 \leq (n_1 + n_2)/n_1\boldsymbol{I} \\ n_1^{-1}\boldsymbol{S}_1 & \text{otherwise.} \end{cases} \tag{1.6}$$

Clearly for large n_2, the truncation in (1.6) begins later than in (1.4). But no dominance result is available. We, however, show in Section 2.1, Corollary 1 that if we modify (1.4) in which the truncation is without the determinant sign, then it dominates $\widehat{\Sigma}_1^{UB}$ for all p. This estimator is similar, in spirit to (1.6) and perhaps can be obtained from the pivot $\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$. However, we show in Corollary 1 that the estimator given by (1.6) dominates this estimator. In Section 2, we describe a general method for obtaining the estimators $\widehat{\Sigma}_1(\Psi)$ dominating another one $\widehat{\Sigma}_1(\Psi^0)$. From this result, we get estimators improving on the unbiased estimator $\widehat{\Sigma}_1^{UB}$ in terms of risk. One of the improved estimators is the so-called REML estimator, which can be also interpreted as an empirical Bayes rule. Using the general method, we provide minimax estimators dominating the minimax estimator given by James and Stein (1961).

In Section 3, it is shown that non-order-preserving estimators can be improved upon by the order-preserving estimators. This implies that the minimax estimators given in Section 2 can be further improved upon.

The problem of estimating the 'between' mutivariate component of variance Σ_A is treated in Section 4. The unbiased estimator of Σ_A is given by

$$\widehat{\Sigma}_A^{UB} = r^{-1}\left(n_2^{-1}\boldsymbol{S}_2 - n_1^{-1}\boldsymbol{S}_1\right),$$

which is not always non-negative definite (n.n.d.). Amemiya (1985) proposed an REML estimator which is n.n.d. for eliminating this undesirable property. However its superiority over the unbiased estimator has not been established from a decision-theoretical aspect. Mathew *et al.* (1994) considered another type of estimators, namely, linear combinations of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ and provided conditions under which the combined estimators are n.n.d. and better than the unbiased estimator relative to the quadratic loss function. In the univariate case, the mean sqaured error (MSE) has been usually employed as a criterion of comparing estimators of the 'between' component of variance. However we do not think the MSE is an appropriate measure in evaluating estimators of dispersion parameters because the MSE penalizes the under-estimate less than the over-estimate. As an alternative measure, we employ the Kullback-Leibler distance or Stein loss function and consider the estimation of Σ_A in the context of simultaneous estimation of Σ_1 and Σ_A. Under this measure, the results given in Sections 2 and 3 are directly applicable to get estimators improving on the unbiased estimators $(\widehat{\Sigma}_1^{UB}, \Sigma_A^{UB})$. From this result, it is shown that the REML estimators of (Σ_1, Σ_A) dominates the unbiased estimators. Also n.n.d. estimators superior to minimax estimators of Σ_1 and Σ_A are derived. The paper concludes in Section 5.

2 Estimation of Multivariate 'Within' Component of Variance

2.1 A general approach to improving estimators

Let $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ be independent random matrices, $\boldsymbol{S}_i \sim \mathcal{W}_p(\Sigma_i, n_i)$, $i = 1, 2$, with $\Sigma_1 \leq \Sigma_2$ where $\Sigma_1 \leq \Sigma_2$ denotes that $\Sigma_2 - \Sigma_1$ is n.n.d. Denote the parameter space by $\Omega = \{(\Sigma_1, \Sigma_2) \mid \Sigma_1 \leq \Sigma_2\}$. Suppose that we want to estimate Σ_1 relative to the Stein (or entropy) loss function

$$L(\widehat{\Sigma}_1 \Sigma_1^{-1}) = \mathrm{tr}\widehat{\Sigma}_1 \Sigma_1^{-1} - \log|\widehat{\Sigma}_1 \Sigma_1^{-1}| - p, \tag{2.1}$$

which was proposed by James and Stein (1961) and also can be derived by the Kullback-Leibler distance

$$\int \left\{ \log \frac{f(\boldsymbol{S}_1; \widehat{\Sigma}_1)}{f(\boldsymbol{S}_1; \Sigma_1)} \right\} f(\boldsymbol{S}_1; \widehat{\Sigma}_1) d\nu(\boldsymbol{S}_1)$$

where $f(\boldsymbol{S}_1; \Sigma_1)$ designates a density function of $\boldsymbol{S}_1$ with respect to measure $\nu(\cdot)$. Every estimator $\widehat{\Sigma}_1$ is evaluated by the risk function $R_1(\omega; \widehat{\Sigma}_1) = E_\omega[L(\widehat{\Sigma}_1 \Sigma_1^{-1})]$ for $\omega \in \Omega$.

Let $\boldsymbol{S}_2^{1/2}$ be a symmetric matrix such that $\boldsymbol{S}_2 = (\boldsymbol{S}_2^{1/2})^2$ and let $\boldsymbol{P}$ be an orthogonal $p \times p$ matrix such that

$$\boldsymbol{P}' \boldsymbol{S}_2^{-1/2} \boldsymbol{S}_1 \boldsymbol{S}_2^{-1/2} \boldsymbol{P} = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p),$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$. We consider estimators of the form

$$\widehat{\Sigma}_1(\Psi) = \boldsymbol{S}_2^{1/2} \boldsymbol{P} \Psi(\Lambda) \boldsymbol{P}' \boldsymbol{S}_2^{1/2} \tag{2.2}$$

where $\Psi(\Lambda) = \mathrm{diag}(\psi_1(\Lambda), \ldots, \psi_p(\Lambda))$ for nonnegative function $\psi(\Lambda)$. For given estimator $\widehat{\Sigma}_1(\Psi)$, we define two types of truncation rules $[\Psi(\Lambda)]^{TR}$ and $[\Psi(\Lambda)]^{TR*}$ by

$$\begin{aligned} [\Psi(\Lambda)]^{TR} &= \mathrm{diag}(\psi_1^{TR}(\Lambda), \ldots, \psi_p^{TR}(\Lambda)), \\ \psi_i^{TR}(\Lambda) &= \min\left\{\psi_i(\Lambda), \frac{\lambda_i + 1}{n_1 + n_2}\right\}, \quad i = 1, \ldots, p, \end{aligned} \tag{2.3}$$

and

$$\begin{aligned}
[\Psi(\Lambda)]^{TR*} &= \mathrm{diag}(\psi_1^{TR*}(\Lambda), \ldots, \psi_p^{TR*}(\Lambda)), \\
\psi^{TR*}(\Lambda) &= \begin{cases} (n_1+n_2)^{-1}(\lambda_i+1) & \text{if } (n_1+n_2)^{-1}(\Lambda+\boldsymbol{I}) \le \Psi(\Lambda) \\ \psi_i(\Lambda) & \text{otherwise.} \end{cases}
\end{aligned} \tag{2.4}$$

Then the corresponding truncated estimators are written as

$$\begin{aligned}
\widehat{\Sigma}_1([\Psi]^{TR}) &= \boldsymbol{S}_2^{1/2}\boldsymbol{P}\mathrm{diag}(\psi_1^{TR}(\Lambda), \ldots, \psi_p^{TR}(\Lambda))\boldsymbol{P}'\boldsymbol{S}_2^{1/2}, \\
\widehat{\Sigma}_1([\Psi]^{TR*}) &= \begin{cases} (n_1+n_2)^{-1}(\boldsymbol{S}_1+\boldsymbol{S}_2) & \text{if } (n_1+n_2)^{-1}(\boldsymbol{S}_1+\boldsymbol{S}_2) \le \widehat{\Sigma}_1(\Psi(\Lambda)) \\ \widehat{\Sigma}_1(\Psi(\Lambda)) & \text{otherwise.} \end{cases}
\end{aligned} \tag{2.5}$$

Note that each diagonal element is truncated componentwise in $\widehat{\Sigma}_1([\Psi]^{TR})$ while the diagonal matrix is truncated for $\widehat{\Sigma}_1([\Psi]^{TR*})$. We get the following general dominance results.

Theorem 1.
(1) The estimator $\widehat{\Sigma}_1([\Psi]^{TR})$ dominates $\widehat{\Sigma}_1(\Psi)$ relative to the Stein loss (2.1) if $P[[\Psi(\Lambda)]^{TR} \neq \Psi(\Lambda)] > 0$ at some $\omega \in \Omega$.
(2) The estimator $\widehat{\Sigma}_1([\Psi]^{TR})$ dominates $\widehat{\Sigma}_1([\Psi]^{TR*})$ relative to the Stein loss (2.1) if $P[[\Psi(\Lambda)]^{TR} \neq [\Psi(\Lambda)]^{TR*}] > 0$ at some $\omega \in \Omega$.

Proof. Without any loss of generality, let $\Sigma_1 = \boldsymbol{I}$ and $\Sigma_2 = \Theta = \mathrm{diag}(\theta_1, \ldots, \theta_p)$ with $\theta_1 \ge 1, \ldots, \theta_p \ge 1$. The joint density of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ is

$$const.|\boldsymbol{S}_1|^{(n_1-p-1)/2}|\boldsymbol{S}_2|^{(n_2-p-1)/2}|\Theta|^{-n_2/2}\mathrm{etr}\left[-\frac{1}{2}(\boldsymbol{S}_1+\Theta^{-1}\boldsymbol{S}_2)\right].$$

Making the transformation $\mathbf{F} = \boldsymbol{S}_2^{-1/2}\boldsymbol{S}_1\boldsymbol{S}_2^{-1/2}$ with $J(\boldsymbol{S}_1 \to \mathbf{F}) = |\boldsymbol{S}_2|^{(p+1)/2}$ gives the joint density of $\mathbf{F}$ and $\boldsymbol{S}_2$:

$$f_{F,S_2}(\mathbf{F}, \boldsymbol{S}_2) = const.|\mathbf{F}|^{(n_1-p-1)/2}|\boldsymbol{S}_2|^{(n_1+n_2-p-1)/2}|\Theta|^{-n_2/2}\mathrm{etr}\left[-\frac{1}{2}(\mathbf{F}+\Theta^{-1})\boldsymbol{S}_2\right]. \tag{2.6}$$

Making the transformation $\mathbf{F} = \boldsymbol{P}\Lambda\boldsymbol{P}'$, we see that the joint density of $(\Lambda, \boldsymbol{P}, \boldsymbol{S}_2)$ is written by

$$f_{\Lambda,P,S_2}(\Lambda, \boldsymbol{P}, \boldsymbol{S}_2) = const.f_p(\boldsymbol{P})g(\Lambda)|\boldsymbol{S}_2|^{(n_1+n_2-p-1)/2}|\Theta|^{-n_2/2}\mathrm{etr}\left[-\frac{1}{2}(\boldsymbol{P}\Lambda\boldsymbol{P}'+\Theta^{-1})\boldsymbol{S}_2\right],$$

where $f_p(\boldsymbol{P}) = J(\boldsymbol{P}'d\boldsymbol{P} \to d\boldsymbol{P})$ and $g(\Lambda)$ is a function of Λ (see Srivastava and Khatri (1979, p.31-32)). Hence the conditional distribution of $\boldsymbol{S}_2$ given $(\Lambda, \boldsymbol{P})$ is

$$\boldsymbol{S}_2 \,\Big|\, (\Lambda, \boldsymbol{P}) \sim W_p\left(\left(\boldsymbol{P}\Lambda\boldsymbol{P}'+\Theta^{-1}\right)^{-1}, n_1+n_2\right),$$

which yields the conditional expectation of $\boldsymbol{S}_2$ given $(\Lambda, \boldsymbol{P})$,

$$E[\boldsymbol{S}_2 | \Lambda, \boldsymbol{P}] = (n_1 + n_2)\left(\boldsymbol{P}\Lambda\boldsymbol{P}' + \Theta^{-1}\right)^{-1}. \tag{2.7}$$

For the proof of part (1), we write the difference of the risk functions of $\widehat{\Sigma}_1(\Psi)$ and $\widehat{\Sigma}_1([\Psi]^{TR})$ as

$$\begin{aligned} & R_1(\Theta, \widehat{\Sigma}_1(\Psi)) - R_1(\Theta, \widehat{\Sigma}_1([\Psi]^{TR})) \\ = \ & E_\Theta\left[\text{tr}(\boldsymbol{P}\Psi(\Lambda)\boldsymbol{P}' - \boldsymbol{P}[\Psi(\Lambda)]^{TR}\boldsymbol{P}')\boldsymbol{S}_2 - \log|\Psi(\Lambda)\{[\Psi(\Lambda)]^{TR}\}^{-1}|\right] \\ = \ & E_\Theta^{\Lambda,P}\left[\text{tr}(\Psi(\Lambda) - [\Psi(\Lambda)]^{TR})\boldsymbol{P}'E_\Theta[\boldsymbol{S}_2|\Lambda, \boldsymbol{P}]\boldsymbol{P} - \log|\Psi(\Lambda)\{[\Psi(\Lambda)]^{TR}\}^{-1}|\right]. \end{aligned} \tag{2.8}$$

From (2.7) and the fact thst $\Psi(\Lambda) \geq [\Psi(\Lambda)]^{TR}$, it follows that the r.h.s. in (2.8) is greater than or equal to

$$\begin{aligned} & E_\Theta\left[\text{tr}\{\Psi(\Lambda) - [\Psi(\Lambda)]^{TR}\}(n_1+n_2)(\Lambda + \boldsymbol{I})^{-1} - \log|\Psi(\Lambda)\{[\Psi(\Lambda)]^{TR}\}^{-1}|\right] \\ = \ & \sum_{i=1}^p E_\Theta\left[\left\{\left(\psi_i(\Lambda) - \frac{\lambda_i+1}{n_1+n_2}\right)\frac{n_1+n_2}{\lambda_i+1} - \log\psi_i(\Lambda)\frac{n_1+n_2}{\lambda_i+1}\right\}\right. \\ & \qquad\qquad\qquad \left.\times I\left(\psi_i(\Lambda) > \frac{\lambda_i+1}{n_1+n_2}\right)\right] \\ = \ & \sum_{i=1}^p E_\Theta\left[\left\{\psi_i\frac{n_1+n_2}{\lambda_i+1} - \log\psi_i\frac{n_1+n_2}{\lambda_i+1} - 1\right\} I(\psi_i(\Lambda) > \frac{\lambda_i+1}{n_1+n_2})\right] \\ \geq \ & 0, \end{aligned} \tag{2.9}$$

which proves part (1).

For the proof of part (2), note that $(n_1+n_2)^{-1}(\Lambda + \boldsymbol{I}) \leq \Psi(\Lambda)$ is equivalent to the condition that $(n_1+n_2)^{-1}(\lambda_i + 1) \leq \psi_i(\Lambda)$ for every i, and that under this condition, $\psi_i^{TR}(\Lambda) = (n_1+n_2)^{-1}(\lambda_i+1)$ for every i. Since $P[[\Psi(\Lambda)]^{TR} \neq [\Psi(\Lambda)]^{TR*}] > 0$ for some $\omega \in \Omega$, there exists an index set J such that

$$P_\omega[\psi_j^{TR}(\Lambda) < \psi_j^{TR*}(\Lambda)] = P_\omega[(n_1+n_2)^{-1}(\lambda_i+1) < \psi_j^{TR*}(\Lambda)] > 0$$

at some $\omega \in \Omega$ for any $j \in J$. By the same arguments as in (2.8) and (2.9), we have

$$\begin{aligned} & R_1(\Theta, \widehat{\Sigma}_1([\Psi]^{TR*})) - R_1(\Theta, \widehat{\Sigma}_1([\Psi]^{TR})) \\ \geq \ & \sum_{j\in J} E_\Theta\left[\left\{\left(\psi_j(\Lambda) - \frac{\lambda_i+1}{n_1+n_2}\right)\frac{n_1+n_2}{\lambda_i+1} - \log\psi_j(\Lambda)\frac{n_1+n_2}{\lambda_i+1}\right\}\right. \\ & \qquad\qquad\qquad \left.\times I\left(\psi_j(\Lambda) > \frac{\lambda_i+1}{n_1+n_2}\right)\right], \end{aligned}$$

which is nonnegative, and the proof of Theorem 1 is complete. □□

Corollary 1. *The estimator in (1.4) modified as*

$$\widehat{\Sigma}_1^S(a) = \begin{cases} (n_1+n_2-a)^{-1}(\boldsymbol{S}_1+\boldsymbol{S}_2) & \text{if } \boldsymbol{I}+\boldsymbol{S}_1^{-1}\boldsymbol{S}_2 \le (n_1+n_2-a)/n_1 \\ n_1^{-1}\boldsymbol{S}_1 & \text{otherwise} \end{cases}$$

where $0 < a < n_2$, *dominates* $\widehat{\Sigma}_1^{UB}$. *Furthermore* $\widehat{\Sigma}_1^S(0)$ *dominates* $\widehat{\Sigma}_1^S(a)$.

Proof. Define the set A by

$$\begin{aligned} A &= \left\{\boldsymbol{I}+\boldsymbol{S}_1^{-1}\boldsymbol{S}_2 \le (n_1+n_2-a)/n_1\right\} \\ &= \left\{\frac{\lambda_i}{\lambda_i+1}\frac{n_1+n_2-a}{n_1} > 1\right\}. \end{aligned}$$

Then the same arguments as in the proof of Theorem 1 gives that

$$\begin{aligned} & R_1(\Theta, \widehat{\Sigma}_1^{UB}) - R_1(\Theta, \widehat{\Sigma}_1^S(a)) \\ \ge & \sum_{i=1}^p E_\Theta\left[\left\{\left(\frac{\lambda_i}{n_1} - \frac{\lambda_i+1}{n_1+n_2-a}\right)\frac{n_1+n_2}{\lambda_i+1} - \log\frac{\lambda_i}{n_1}\frac{n_1+n_2}{\lambda_i+1}\right\} I_A\right] \\ = & \sum_{i=1}^p E_\Theta\left[\left\{\frac{\lambda_i}{\lambda_i+1}\frac{n_1+n_2-a}{n_1} - \log\frac{\lambda_i}{\lambda_i+1}\frac{n_1+n_2-a}{n_1} - 1\right\} I_A\right. \\ & \left. + \frac{a}{n_1+n_2-a}\left\{\frac{\lambda_i}{\lambda_i+1}\frac{n_1+n_2-a}{n_1} - 1\right\} I_A\right], \end{aligned}$$

which is greater than or equal to zero, and the first part of Corollary 1 is proved.

For the proof of the second part, define the set B by

$$B = \left\{\frac{\lambda_i}{\lambda_i+1}\frac{n_1+n_2-a}{n_1} > 1\right\},$$

and denote the complement of A by A^c. Similarly to the above arguments, the risk difference is written as

$$\begin{aligned} & R_1(\Theta, \widehat{\Sigma}_1^S(a)) - R_1(\Theta, \widehat{\Sigma}_1^S(0)) \\ \ge & \sum_{i=1}^p E_\Theta\left[\left\{\left(\frac{\lambda_i+1}{n_1+n_2-a} - \frac{\lambda_i+1}{n_1+n_2-a}\right)\frac{n_1+n_2}{\lambda_i+1} - \log\frac{n_1+n_2}{n_1+n_2-a}\right\} I_A\right] \\ & + \sum_{i=1}^p E_\Theta\left[\left\{\left(\frac{\lambda_i}{n_1} - \frac{\lambda_i+1}{n_1+n_2}\right)\frac{n_1+n_2}{\lambda_i+1} - \log\frac{\lambda_i}{n_1}\frac{n_1+n_2}{\lambda_i+1}\right\} I_{B\cap A^c}\right], \end{aligned}$$

which is greater than or equal to zero, and the proof of Corollary 1 is complete. □□

2.2 Improvements on the unbiased estimator

Since $\boldsymbol{S}_1 = \boldsymbol{S}_2^{1/2}\boldsymbol{S}_2^{-1/2}\boldsymbol{S}_1\boldsymbol{S}_2^{-1/2}\boldsymbol{S}_2^{1/2} = \boldsymbol{S}_2^{1/2}\boldsymbol{P}\Lambda\boldsymbol{P}'\boldsymbol{S}_2^{1/2}$, the unbiased estimator $\widehat{\Sigma}_1^{UB} = n_1^{-1}\boldsymbol{S}_1$ can be expressed in the same manner as (2.2) by

$$\widehat{\Sigma}_1^{UB} = \widehat{\Sigma}_1(\Psi^{UB}),$$

where

$$\Psi^{UB} = \mathrm{diag}(n_1^{-1}\lambda_1, \ldots, n_1^{-1}\lambda_p).$$

The truncation rules given in Section 2.1 produce the estimators

$$\widehat{\Sigma}_1^{REML} = \widehat{\Sigma}_1([\Psi^{UB}]^{TR}), \tag{2.10}$$

where

$$[\Psi^{UB}]^{TR} = \mathrm{diag}\left(\min\left\{\frac{\lambda_1}{n_1}, \frac{\lambda_1+1}{n_1+n_2}\right\}, \ldots, \min\left\{\frac{\lambda_p}{n_1}, \frac{\lambda_p+1}{n_1+n_2}\right\}\right),$$

and

$$\begin{aligned}\widehat{\Sigma}_1^{USTR} &= \widehat{\Sigma}_1([\Psi^{UB}]^{TR*}) \qquad (2.11)\\ &= \begin{cases}(n_1+n_2)^{-1}(\boldsymbol{S}_1+\boldsymbol{S}_2) & \text{if } (n_1+n_2)^{-1}(\boldsymbol{S}_1+\boldsymbol{S}_2) \le n_1^{-1}\boldsymbol{S}_1\\ n_1^{-1}\boldsymbol{S}_1 & \text{otherwise.}\end{cases}\end{aligned}$$

For instance, suppose that

$$\lambda_1 \ge n_1/n_2, \ldots, \lambda_q \ge n_1/n_2, \lambda_{q+1} < n_1/n_2, \ldots, \lambda_p < n_1/n_2$$

for some q. Then $[\Psi^{UB}]^{TR}$ takes the value

$$[\Psi^{UB}]^{TR} = \mathrm{diag}\left(\frac{\lambda_1+1}{n_1+n_2}, \ldots, \frac{\lambda_q+1}{n_1+n_2}, \frac{\lambda_{q+1}}{n_1}, \ldots, \frac{\lambda_p}{n_1}\right),$$

while

$$[\Psi^{UB}]^{TR*} = \mathrm{diag}\left(\frac{\lambda_1}{n_1}, \ldots, \frac{\lambda_p}{n_1}\right).$$

From Theorem 1, we get

Corollary 2. *The estimator $\widehat{\Sigma}_1^{REML}$ dominates the estimator $\widehat{\Sigma}_1^{USTR}$ which improves on the unbiased one Σ_1^{UB} relative to the Stein loss (2.1).*

The estimator $\widehat{\Sigma}_1^{REML}$ is known to be the restricted (or residual) maximum likelihood (REML) estimator of Σ_1 under the constraint $\Sigma_1 \le \Sigma_2$. Corollary 2 thus implies that the REML estimator is superior not only to the unbiased estimator but also to $\widehat{\Sigma}_1^{USTR}$ given in (2.3), although it appears to have a natural form.

It is interesting to note that the REML estimator $\widehat{\Sigma}_1^{REML}$ can also be derived as an empirical Bayes rule. Let $\boldsymbol{\eta} = \Sigma_2^{-1}$ and $\boldsymbol{\xi} = \Sigma_2^{1/2}\Sigma_1^{-1}\Sigma_2^{1/2}$. Suppose that $\boldsymbol{\eta}$ has non-informative prior distribution $|\boldsymbol{\eta}|^{-(p+1)/2}d\nu(\boldsymbol{\eta})$ for some measure $\nu(\cdot)$ and that $\boldsymbol{\xi}$ is unknown. The joint density of $(\boldsymbol{\eta}, \boldsymbol{S}_1, \boldsymbol{S}_2)$ has the form

$$const.|\boldsymbol{\eta}|^{(n_1+n_2-p-1)/2}|\boldsymbol{S}_1|^{(n_1-p-1)/2}|\boldsymbol{S}_2|^{(n_2-p-1)/2}|\boldsymbol{\xi}|^{n_1/2}\mathrm{etr}\left[-\frac{1}{2}\left(\boldsymbol{\xi}^{1/2}\boldsymbol{S}_1\boldsymbol{\xi}^{1/2}+\boldsymbol{S}_2\right)\boldsymbol{\eta}\right],$$

so that the posterior density given $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ and the marginal density of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are given by

$$\begin{aligned}
(\text{posterior}) &\propto |\boldsymbol{\eta}|^{(n_1+n_2-p-1)/2}\mathrm{etr}\left[-\frac{1}{2}\left(\boldsymbol{\xi}^{1/2}\boldsymbol{S}_1\boldsymbol{\xi}^{1/2}+\boldsymbol{S}_2\right)\boldsymbol{\eta}\right],\\
(\text{marginal}) &\propto |\boldsymbol{\xi}|^{n_1/2}|\boldsymbol{\xi}^{1/2}\boldsymbol{S}_1\boldsymbol{\xi}^{1/2}+\boldsymbol{S}_2|^{-(n_1+n_2)/2}|\boldsymbol{S}_1|^{(n_1-p-1)/2}|\boldsymbol{S}_2|^{(n_2-p-1)/2}.
\end{aligned}$$

We thus get the Bayes estimator of Σ_1 under the Stein loss (2.1)

$$\begin{aligned}
\widehat{\Sigma}_1^B(\boldsymbol{\xi}) &= \boldsymbol{\xi}^{-1/2}(E[\boldsymbol{\eta}|\boldsymbol{S}_1,\boldsymbol{S}_2])^{-1}\boldsymbol{\xi}^{-1/2}\\
&= (n_1+n_2)^{-1}\boldsymbol{\xi}^{-1/2}(\boldsymbol{\xi}^{1/2}\boldsymbol{S}_1\boldsymbol{\xi}^{1/2}+\boldsymbol{S}_2)\boldsymbol{\xi}^{-1/2}\\
&= (n_1+n_2)^{-1}(\boldsymbol{S}_1+\boldsymbol{\xi}^{-1/2}\boldsymbol{S}_2\boldsymbol{\xi}^{-1/2}).
\end{aligned}$$

Since $\boldsymbol{\xi}$ is unknown, $\boldsymbol{\xi}$ needs to be estimated from the marginal density. Putting $\boldsymbol{\beta} = \boldsymbol{S}_1^{-1/2}\boldsymbol{\xi}^{-1/2}\boldsymbol{S}_2\boldsymbol{\xi}^{-1/2}\boldsymbol{S}_1^{-1/2}$, the maximum likelihood estimator of $\boldsymbol{\beta}$ can be derived by maximising $|\boldsymbol{\beta}|^{n_2/2}|\boldsymbol{I}+\boldsymbol{\beta}|^{-(n_1+n_2)/2}$ subject to the order restriction $\boldsymbol{\beta} \le \boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$ since $\boldsymbol{\xi} \ge \boldsymbol{I}$. The resulting MLE of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \boldsymbol{Q}\mathrm{diag}\left(\min\left\{\frac{n_2}{n_1}, \frac{1}{\lambda_i}\right\}, i = 1,\ldots,p\right)\boldsymbol{Q}'$$

where $\boldsymbol{Q}$ is an orthogonal $p \times p$ matrix such that

$$\boldsymbol{Q}'\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}\boldsymbol{Q} = \mathrm{diag}\left(\lambda_1^{-1},\ldots,\lambda_p^{-1}\right).$$

Putting $\hat{\boldsymbol{\beta}}$ or $\hat{\boldsymbol{\xi}}$ into the Bayes Estimator $\widehat{\Sigma}_1^B(\boldsymbol{\xi})$, we obtain the empirical Bayes estimator

$$\begin{aligned}
\widehat{\Sigma}_1^B(\hat{\boldsymbol{\xi}}) &= \frac{1}{n_1+n_2}\boldsymbol{S}_1^{1/2}\boldsymbol{Q}\left\{\mathrm{diag}\left(\min\left\{\frac{n_2}{n_1}, \frac{1}{\lambda_i}\right\}, i = 1,\ldots,p\right)+\boldsymbol{I}\right\}\boldsymbol{Q}'\boldsymbol{S}_1^{1/2}\\
&= \frac{1}{n_1+n_2}\boldsymbol{S}_1^{1/2}\boldsymbol{Q}\mathrm{diag}\left(\min\left\{\frac{1}{n_1}, \frac{\lambda_i+1}{(n_1+n_2)\lambda_i}\right\}, i = 1,\ldots,p\right)\boldsymbol{Q}'\boldsymbol{S}_1^{1/2}.
\end{aligned}$$

Here note that orthogonal matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ satisfy

$$\begin{cases}
\boldsymbol{S}_2^{-1/2}\boldsymbol{S}_1\boldsymbol{S}_2^{-1/2} = (\boldsymbol{S}_2^{-1/2}\boldsymbol{S}_1^{1/2})(\boldsymbol{S}_2^{-1/2}\boldsymbol{S}_1^{1/2})' = \boldsymbol{P}\Lambda\boldsymbol{P}'\\
\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2} = (\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2^{1/2})(\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2^{1/2})' = \boldsymbol{Q}\Lambda^{-1}\boldsymbol{Q}'.
\end{cases}$$

Then we have that

$$S_2^{-1/2}S_1^{1/2} = P\Lambda^{1/2}Q' \quad \text{or} \quad S_2^{1/2}P = S_1^{1/2}Q\Lambda^{-1/2}.$$

Hence $\widehat{\Sigma}_1(\hat{\boldsymbol{\xi}})$ is rewritten as

$$\widehat{\Sigma}_1(\hat{\boldsymbol{\xi}}) = S_2^{1/2}P\text{diag}\left(\min\left\{\frac{\lambda_i}{n_1}, \frac{\lambda_i+1}{n_1+n_2}\right\}, i=1,\ldots,p\right)P'S_2^{1/2},$$

which is identical to the REML estimator $\widehat{\Sigma}_1^{REML}$. Hence the REML estimator can be interpreted as the empirical Bayes rule.

2.3 Improvements on the minimax estimator

Historically, the first interesting event in estimation of the covariance matrix was brought by James and Stein (1961), who established the non-minimaxity of the unbiased estimator $\widehat{\Sigma}_1^{UB}$ and presented the minimax estimator of the form

$$\widehat{\Sigma}_1^{JS} = T_1 D^m T_1',$$

where T_1 is a lower triangular $p \times p$ matrix such that $S_1 = T_1T_1'$, and D^m is the diagonal matrix given by $D^m = \text{diag}(d_1, \ldots, d_p)$ for

$$d_i = (n_1 + p + 1 - 2i)^{-1}, \quad i = 1, \ldots, p.$$

We now obtain a minimax estimator (improving on $\widehat{\Sigma}_1^{JS}$) using the information on S_2. Let us consider the estimator of the form

$$\widehat{\Sigma}_1^m = \widehat{\Sigma}_1(\Psi^m) = S_2^{1/2}P\Psi^m(\Lambda)P'S_2^{1/2} \tag{2.12}$$

where

$$\Psi^m(\Lambda) = \text{diag}(d_1\lambda_1, \ldots, d_p\lambda_p).$$

We first demonstare the minimaxity of $\widehat{\Sigma}_1^m$.

Proposition 1. *The estimator $\widehat{\Sigma}_1^m$ is a minimax estimator improving on $\widehat{\Sigma}_1^{JS}$ relative to the Stein loss (2.1).*

Proof. Recall that $\mathbf{F} = S_2^{-1/2}S_1S_2^{-1/2} = P\Lambda P'$ and that $S_1 \sim \mathcal{W}_p(I, n_1)$. Then it is seen that the conditional distribution of $\mathbf{F}$ given S_2 has $\mathcal{W}_p(\Sigma_*, n_1)$ for $\Sigma_* = S_2^{-1}$. Then the risk function of $\widehat{\Sigma}_1^m$ is represented by

$$R_1(\Theta, \widehat{\Sigma}_1^m) = E^{S_2}\left[E^{F|S_2}\left[\text{tr}P\Psi^m(\Lambda)P'\Sigma_*^{-1} - \log|P\Psi^m(\Lambda)P'\Sigma_*^{-1}| - p\middle|S_2\right]\right], \tag{2.13}$$

so that given $\boldsymbol{S}_2$, conditionally $\boldsymbol{P}\Psi^m\boldsymbol{P}'$ corresponds to the Stein's orthogonally invariant minimax estimator of Σ_*. Then from the results of Stein (1977) and Dey and Srinivasan (1985), it follows that the conditional expectation $E^{F|S_2}[\cdot|\boldsymbol{S}_2]$ given in (2.13) is less than the constant minimax risk of $\widehat{\Sigma}_1^{JS}$, which proves Proposition 1. □□

Now, using the truncation rules given in Section 2.1, we get the estimators

$$\widehat{\Sigma}_1^{MTR} = \widehat{\Sigma}_1([\Psi^m]^{TR}), \tag{2.14}$$

where

$$[\Psi^m(\Lambda)]^{TR} = \mathrm{diag}\left(\min\left\{d_1\lambda_1, \frac{\lambda_1+1}{n_1+n_2}\right\}, \ldots, \min\left\{d_p\lambda_p, \frac{\lambda_p+1}{n_1+n_2}\right\}\right),$$

and

$$\begin{aligned}\widehat{\Sigma}_1^{MSTR} &= \widehat{\Sigma}_1([\Psi^m]^{TR*}) \\ &= \begin{cases}(n_1+n_2)^{-1}(\boldsymbol{S}_1+\boldsymbol{S}_2) & \text{if } (n_1+n_2)^{-1}(\boldsymbol{S}_1+\boldsymbol{S}_2) \le \widehat{\Sigma}_1^m \\ \widehat{\Sigma}_1^m & \text{otherwise.}\end{cases}\end{aligned}$$

From Theorem 1 and Proposition 1, we can get

Corollary 3. *The estimator $\widehat{\Sigma}_1^{MTR}$ is a minimax estimator improving on $\widehat{\Sigma}_1^{MSTR}$ which dominates $\widehat{\Sigma}_1^m$ and $\widehat{\Sigma}_1^{JS}$ relative to the Stein loss (2.1).*

3 Dominance Results by Order-Preserving Estimation

We consider the general type of estimators given by (2.2), namely,

$$\widehat{\Sigma}_1(\Psi) = \boldsymbol{S}_2^{1/2}\boldsymbol{P}\Psi(\Lambda)\boldsymbol{P}'\boldsymbol{S}_2^{1/2}, \quad \Psi(\Lambda) = \mathrm{diag}(\psi_1(\Lambda), \ldots, \psi_p(\Lambda)).$$

For the diagonal elements $\psi_1(\Lambda), \ldots, \psi_p(\Lambda)$, it is quite natural to satisfy the condition

$$\psi_1(\Lambda) \ge \psi_2(\Lambda) \ge \ldots \ge \psi_p(\Lambda) \quad \text{for any } \Lambda,$$

which is called *order-preserving* in Sheena and Takemura (1992). The minimax and improved estimators $\widehat{\Sigma}_1^{MTR}$ and $\widehat{\Sigma}_1^{MSTR}$ given in the previous section do not satisfy the order-preserving condition.

In this section, we show that non-order-preserving estimators can be improved on by the order-preserving estimators. We first write the risk function of $\widehat{\Sigma}_1(\Psi)$ as

$$\begin{aligned}R_1(\Theta, \widehat{\Sigma}_1(\Psi)) &= E_\Theta[\mathrm{tr}\boldsymbol{P}\Psi(\Lambda)\boldsymbol{P}'(\boldsymbol{P}\Lambda\boldsymbol{P}' + \Theta^{-1})^{-1} - \log|\Psi(\Lambda)| - \log|\boldsymbol{S}_2| - p] \\ &= E_\Theta[\mathrm{tr}\Psi(\Lambda)(\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P})^{-1} - \log|\Psi(\Lambda)| - \log|\boldsymbol{S}_2| - p]. \end{aligned} \tag{3.1}$$

Let $\boldsymbol{B} = \Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}$ and denote the (i,i)-diagonal element of $\boldsymbol{B}^{-1}$ by $\boldsymbol{B}^{ii}$. Then the following lemma is essential for proving the required result.

Lemma 1. *Let $E[\cdot|\Lambda]$ be a conditional expectation with respect to $\boldsymbol{P}$ given Λ. For $i < j$,*

$$E[\boldsymbol{B}^{ii}|\Lambda] \leq E[\boldsymbol{B}^{jj}|\Lambda]. \tag{3.2}$$

Proof. From (2.6), it is seen that $\mathbf{F} = \boldsymbol{S}_2^{-1/2}\boldsymbol{S}_1\boldsymbol{S}_2^{-1/2}$ has the density

$$const.|\mathbf{F}|^{(n_1-p-1)/2}|\mathbf{F} + \Theta^{-1}|^{-(n_1+n_2-p-1)/2}|\Theta^{-1}|^{n_2/2},$$

so that the joint density of $(\Lambda, \boldsymbol{P})$ is given by

$$const.f_p(\boldsymbol{P})g(\Lambda)|\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}|^{-(n_1+n_2-p-1)/2}|\Theta|^{-n_2/2},$$

where $f_p(\boldsymbol{P})$ ia a Jacobian and $g(\Lambda)$ is a function of Λ. Hence the inequality (3.2) is equivalent to

$$\int_{O(p)} (\boldsymbol{B}^{jj} - \boldsymbol{B}^{ii})|\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}|^{-(n_1+n_2-p-1)/2}d\mu(\boldsymbol{P}) \geq 0,$$

where $\mu(\cdot)$ designates the invariant probability measure on the groups of p-dimensional orthogonal matrices $O(p)$. Without any loss of generality, we demonstrates the case where $j = 2$ and $i = 1$, that is,

$$\int_{O(p)} (\boldsymbol{B}^{22} - \boldsymbol{B}^{11})|\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}|^{-(n_1+n_2-p-1)/2}d\mu(\boldsymbol{P}) \geq 0. \tag{3.3}$$

Let $\boldsymbol{B}^f_{ii}$, $i = 1, 2$, be the cofactor determinants corresponding to the element $\boldsymbol{B}_{ii}$. Then $\boldsymbol{B}^{22} - \boldsymbol{B}^{11} = (\boldsymbol{B}^f_{22} - \boldsymbol{B}^f_{11})/|\boldsymbol{B}|$, and (3.3) can be written as

$$\int_{O(p)} (\boldsymbol{B}^f_{22} - \boldsymbol{B}^f_{11})|\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}|^{-(n_1+n_2-p+1)/2}d\mu(\boldsymbol{P}) \geq 0. \tag{3.4}$$

Note that $\mu(\boldsymbol{P})$ is invariant with respect to permutation of columns of $\boldsymbol{P}$. By interchanging 1 and 2, the left-hand side of (3.4) can be written as

$$-\int_{O(p)} (\boldsymbol{B}^f_{22} - \boldsymbol{B}^f_{11})|\Lambda^* + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}|^{-(n_1+n_2-p+1)/2}d\mu(\boldsymbol{P}) \geq 0, \tag{3.5}$$

where $\Lambda^* = \mathrm{diag}(\lambda_2, \lambda_1, \lambda_3, \ldots, \lambda_p)$. Adding (3.4) and (3.5), we see that for $\alpha = -(n_1 + n_2 - p + 1)/2$, $E[\boldsymbol{B}^{22}|\Lambda] \geq E[\boldsymbol{B}^{11}|\Lambda]$ if and only if

$$\int_{O(p)} (\boldsymbol{B}^f_{22} - \boldsymbol{B}^f_{11})\left\{|\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}|^{\alpha} - |\Lambda^* + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}|^{\alpha}\right\}d\mu(\boldsymbol{P}) \geq 0. \tag{3.6}$$

Let us decompose $\boldsymbol{P}'\Theta^{-1}\boldsymbol{P}$ as

$$\boldsymbol{P}'\Theta^{-1}\boldsymbol{P} = \begin{pmatrix} a_{11} & a_{12} & \boldsymbol{a}'_{13} \\ a_{12} & a_{22} & \boldsymbol{a}'_{23} \\ \boldsymbol{a}_{13} & \boldsymbol{a}_{23} & \boldsymbol{A}_{33} \end{pmatrix}.$$

Then we have

$$\begin{aligned} \boldsymbol{B}^f_{22} - \boldsymbol{B}^f_{11} &= \begin{vmatrix} a_{11}+\lambda_1 & \boldsymbol{a}'_{13} \\ \boldsymbol{a}_{13} & \boldsymbol{A}_{33}+\Lambda_3 \end{vmatrix} - \begin{vmatrix} a_{22}+\lambda_2 & \boldsymbol{a}'_{23} \\ \boldsymbol{a}_{23} & \boldsymbol{A}_{33}+\Lambda_3 \end{vmatrix} \\ &= (\lambda_1-\lambda_2)|\boldsymbol{A}_{33}+\Lambda_3| + \begin{vmatrix} a_{11} & \boldsymbol{a}'_{13} \\ \boldsymbol{a}_{13} & \boldsymbol{A}_{33}+\Lambda_3 \end{vmatrix} - \begin{vmatrix} a_{22} & \boldsymbol{a}'_{23} \\ \boldsymbol{a}_{23} & \boldsymbol{A}_{33}+\Lambda_3 \end{vmatrix}, \end{aligned}$$

where $\Lambda_3 = \mathrm{diag}(\lambda_3, \dots, \lambda_p)$. On the other hand, for $x = \lambda_1 - \lambda_2$,

$$\begin{aligned} |\Lambda^* + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}| &= \begin{vmatrix} \lambda_1 + a_{11} - x & a_{12} & \boldsymbol{a}'_{13} \\ a_{12} & \lambda_2 + a_{22} + x & \boldsymbol{a}'_{23} \\ \boldsymbol{a}_{13} & \boldsymbol{a}_{23} & \boldsymbol{A}_{33}+\Lambda_3 \end{vmatrix} \\ &= |\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}| + x\begin{vmatrix} \lambda_1 + a_{11} & \boldsymbol{a}'_{13} \\ \boldsymbol{a}_{13} & \boldsymbol{A}_{33}+\Lambda_3 \end{vmatrix} \\ &\quad - x\begin{vmatrix} \lambda_2 + a_{22} & \boldsymbol{a}'_{23} \\ \boldsymbol{a}_{23} & \boldsymbol{A}_{33}+\Lambda_3 \end{vmatrix} - x^2|\boldsymbol{A}_{33}+\Lambda_3| \\ &= |\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}| + |\boldsymbol{A}_{33}+\Lambda_3|(x\lambda_1 - x\lambda_2 - x^2) + k_{P,\Lambda}x \\ &= |\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}| + k_{P,\Lambda}(\lambda_1-\lambda_2), \end{aligned}$$

where

$$k_{P,\Lambda} = k(\boldsymbol{P}'\Theta^{-1}\boldsymbol{P}, \Lambda_3) = \begin{vmatrix} a_{11} & \boldsymbol{a}'_{13} \\ \boldsymbol{a}_{13} & \boldsymbol{A}_{33}+\Lambda_3 \end{vmatrix} - \begin{vmatrix} a_{22} & \boldsymbol{a}'_{23} \\ \boldsymbol{a}_{23} & \boldsymbol{A}_{33}+\Lambda_3 \end{vmatrix}.$$

Therefore the inequality (3.6) is represented by

$$\begin{aligned} \int_{O(p)} &\left\{(\lambda_1-\lambda_2)|\boldsymbol{A}_{33}+\Lambda_3| + k_{P,\Lambda}\right\}\left\{|\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}|^\alpha\right. \\ &\left. -(|\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}| + k_{P,\Lambda}(\lambda_1-\lambda_2))^\alpha\right\} d\mu(\boldsymbol{P}) \geq 0. \end{aligned}$$

Since $|\boldsymbol{A}_{33}+\Lambda_3|$ does not depend on the above permutation of exchanging 1 and 2,

$$\begin{aligned} \int_{O(p)} &(\lambda_1-\lambda_2)|\boldsymbol{A}_{33}+\Lambda_3|\left\{|\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}|^\alpha\right. \\ &\left. -(|\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}| + k_{P,\Lambda}(\lambda_1-\lambda_2))^\alpha\right\} d\mu(\boldsymbol{P}) = 0. \end{aligned}$$

Noting that $(|\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}| + k_{P,\Lambda}(\lambda_1-\lambda_2))^\alpha$ is a decreasing function of $k_{P,\Lambda}$ for $\alpha = -(n_1+n_2-p+1)/2$, we see that

$$k_{P,\Lambda}\left\{|\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}|^\alpha - (|\Lambda + \boldsymbol{P}'\Theta^{-1}\boldsymbol{P}| + k_{P,\Lambda}(\lambda_1-\lambda_2))^\alpha\right\} \geq 0,$$

which establishes the inequality (3.6), and Lemma 1 is proved. □□

Now we demonstrate that non-order-preserving estimators can be improved upon by the order-preserving estimators. Let $\widehat{\Sigma}_1(\Psi)$ be a non-order-preserving estimator. Let $\psi_i^O(\Lambda)$ be the i-th largest element in $(\psi_1(\Lambda),\ldots,\psi_p(\Lambda))$, so that $\psi_1^O(\Lambda)\geq\ldots\geq\psi_p^O(\Lambda)$. Note that $(\psi_1^O,\ldots,\psi_p^O)$ majorizes $(\psi_1,\ldots,\psi_p)$, that is,

$$\sum_{i=1}^{j}\psi_i^O\geq\sum_{i=1}^{j}\psi_i\quad\text{for}\quad 1\leq j\leq p-1\quad\text{and}\quad\sum_{i=1}^{p}\psi_i^O=\sum_{i=1}^{p}\psi_i. \tag{3.7}$$

Let $\widehat{\Sigma}_1(\Psi^O)=\boldsymbol{S}_2^{1/2}\boldsymbol{P}\Psi^O(\Lambda)\boldsymbol{P}'\boldsymbol{S}_2^{1/2}$ for $\Psi^O(\Lambda)=\mathrm{diag}(\psi_1^O(\Lambda),\ldots,\psi_p^O(\Lambda))$. Then we get

Theorem 2. *If $P_\omega[\Psi(\Lambda)\neq\Psi^O(\Lambda)]>0$ for some $\omega\in\Omega$, then $\widehat{\Sigma}_1(\Psi)$ is dominated by the order-preserving estimator $\widehat{\Sigma}_1(\Psi^O)$ relative to the Stein loss (2.1).*

Proof. The risk difference is written as

$$\begin{aligned} R_1(\Theta,\widehat{\Sigma}_1(\Psi^O))-R_1(\Theta,\widehat{\Sigma}_1(\Psi)) &= E_\Theta[\mathrm{tr}(\Psi^O(\Lambda)-\Psi(\Lambda))\boldsymbol{B}^{-1}] \\ &= E_\Theta^\Lambda\left[\sum_{i=1}^{p}(\psi_i^O(\Lambda)-\psi_i(\Lambda))E[\boldsymbol{B}^{ii}|\Lambda]\right]. \end{aligned} \tag{3.8}$$

Following Sheena and Takemura (1992), we use the Abel's identity to get the equation

$$\begin{aligned} &\sum_{i=1}^{p}(\psi_i^O-\psi_i)E[\boldsymbol{B}^{ii}\big|\Lambda] \\ &\quad= (\psi_1^O-\psi_1)(E[\boldsymbol{B}^{11}\big|\Lambda]-E[\boldsymbol{B}^{22}\big|\Lambda]) \\ &\qquad+(\psi_1^O+\psi_2^O-\psi_1-\psi_2)(E[\boldsymbol{B}^{22}\big|\Lambda]-E[\boldsymbol{B}^{33}\big|\Lambda]) \\ &\qquad+\cdots+ \\ &\qquad+(\psi_1^O+\cdots\psi_{p-1}^O-\psi_1-\cdots-\psi_{p-1})(E[\boldsymbol{B}^{p-1,p-1}\big|\Lambda]-E[\boldsymbol{B}^{pp}\big|\Lambda]), \end{aligned}$$

which can be seen to be negative from Lemma 1 and (3.7). Hence from (3.8), Theorem 2 is proved. □□

Applying Theorem 2 to $\widehat{\Sigma}_1^{MTR}$ and $\widehat{\Sigma}_1^{MSTR}$, we obtain the order-preserving estimators improving on them. For instance, the order-preserving estimator of $\widehat{\Sigma}_1^{MTR}$ is given by

$$\widehat{\Sigma}_1^{MTRO}=\boldsymbol{S}_2^{1/2}\boldsymbol{P}[[\Psi^m]^{TR}]^O\boldsymbol{P}'\boldsymbol{S}_2^{1/2},$$

where $[[\Psi^m]^{TR}]^O=\mathrm{diag}(\psi_1^{MTRO},\ldots,\psi_p^{MTRO})$ and ψ_i^{MTRO} is the i-th largest in the diagonal elements $\min\{d_i\lambda_i,(n_1+n_2)^{-1}(\lambda_i+1)\}$, $i=1,\ldots,p$.

4 Estimation of the 'Between' Multivariate Component of Variance

In this section, we consider the estimation of the 'between' multivariate component of variance in the context of the simultaneous estimation of 'within' and 'between' components.

Recall that as described in (1.1), (1.2) and (1.3), $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are independent random matrices having $\mathcal{W}_p(\Sigma_1, n_1)$ and $\mathcal{W}_p(\Sigma_2, n_2)$, respectively, for $\Sigma_2 = \Sigma_1 + r\Sigma_A$. We want to estimate Σ_A based on $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ and to discuss the preference of estimators in a decision-theoretic framework. The parametric structure $\Sigma_2 = \Sigma_1 + r\Sigma_A$ means that estimators of Σ_A can be provided through estimation of both Σ_1 and Σ_2. This suggests that the estimation of Σ_A may be considered in the context of the simultaneous estimation of (Σ_1, Σ_2).

We thus consider the problem of estimating (Σ_1, Σ_2) simultaneously relative to the Kullback-Leibler loss function

$$\begin{aligned} & L_{KL}(\widehat{\Sigma}_1, \widehat{\Sigma}_A; \Sigma_1, \Sigma_A) \\ &= n_1\left\{\mathrm{tr}\widehat{\Sigma}_1\Sigma_1^{-1} - \log|\widehat{\Sigma}_1\Sigma_1^{-1}| - p\right\} \qquad (4.1) \\ &\quad +n_2\left\{\mathrm{tr}(\widehat{\Sigma}_1 + r\widehat{\Sigma}_A)(\Sigma_1 + r\Sigma_A)^{-1} - \log|(\widehat{\Sigma}_1 + r\widehat{\Sigma}_A)(\Sigma_1 + r\Sigma_A)^{-1}| - p\right\}, \end{aligned}$$

which can be really derived from the Kullback-Leibler distance

$$\int \left[\log\left\{f(\boldsymbol{S}_1, \boldsymbol{S}_2; \widehat{\Sigma}_1, \widehat{\Sigma}_A)/f(\boldsymbol{S}_1, \boldsymbol{S}_2; \Sigma_1, \Sigma_A)\right\}\right] f(\boldsymbol{S}_1, \boldsymbol{S}_2; \widehat{\Sigma}_1, \widehat{\Sigma}_A)d\nu(\boldsymbol{S}_1)d\nu(\boldsymbol{S}_2)$$

for joint density function $f(\boldsymbol{S}_1, \boldsymbol{S}_2; \Sigma_1, \Sigma_A)$.

When Σ_1 and $\Sigma_2 = \Sigma_1 + r\Sigma_A$ are estimated by $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$, it is quite natural to take the form $\widehat{\Sigma}_A = r^{-1}(\widehat{\Sigma}_2 - \widehat{\Sigma}_1)$ as an estimator of Σ_A. As long as such types of estimators are treated, the risk function of $(\widehat{\Sigma}_1, \widehat{\Sigma}_A)$ relative to the Kullback-Leibler loss (4.1) is written as

$$\begin{aligned} R_{KL}(\omega; \widehat{\Sigma}_1, \widehat{\Sigma}_A) &= E_\omega\left[L_{KL}(\widehat{\Sigma}_1, \widehat{\Sigma}_A; \Sigma_1, \Sigma_A)\right] \\ &= n_1R_1(\omega; \widehat{\Sigma}_1) + n_2R_2(\omega; \widehat{\Sigma}_2), \end{aligned}$$

where $\omega = (\Sigma_1, \Sigma_1 + r\Sigma_A) \in \Omega$ and

$$\begin{aligned} R_1(\omega; \widehat{\Sigma}_1) &= E_\omega\left[\mathrm{tr}\widehat{\Sigma}_1\Sigma_1^{-1} - \log|\widehat{\Sigma}_1\Sigma_1^{-1}| - p\right] \\ R_2(\omega; \widehat{\Sigma}_2) &= E_\omega\left[\mathrm{tr}\widehat{\Sigma}_2\Sigma_2^{-1} - \log|\widehat{\Sigma}_2\Sigma_2^{-1}| - p\right]. \end{aligned}$$

Hence the original problem under the loss (4.1) is decomposed into two problems of estimating Σ_1 and Σ_2 in terms of the risk functions $R_1(\omega; \widehat{\Sigma}_1)$ and $R_2(\omega; \widehat{\Sigma}_2)$, respectively.

Since the estimation of Σ_1 in terms of the risk $R_1(\omega; \widehat{\Sigma}_1)$ has been treated in previous sections, we need only to consider the estimation of Σ_2 under the risk $R_2(\omega; \widehat{\Sigma}_2)$.

Let $\boldsymbol{S}_1^{1/2}$ be a symmetric matrix such that $\boldsymbol{S}_1 = (\boldsymbol{S}_1^{1/2})^2$ and let $\boldsymbol{Q}$ be an orthogonal $p \times p$ matrix such that

$$\boldsymbol{Q}'\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}\boldsymbol{Q} = \Lambda^{-1} = \mathrm{diag}(\lambda_1^{-1}, \ldots, \lambda_p^{-1}),$$

where $\lambda_1^{-1} \le \ldots \le \lambda_p^{-1}$. The diagonal matrix Λ is also defined in Section 2.1 as

$$\boldsymbol{P}'\boldsymbol{S}_2^{-1/2}\boldsymbol{S}_1\boldsymbol{S}_2^{-1/2}\boldsymbol{P} = \Lambda,$$

so that we note that the following relation holds:

$$\boldsymbol{S}_2^{1/2}\boldsymbol{P} = \boldsymbol{S}_1^{1/2}\boldsymbol{Q}\Lambda^{-1/2}. \tag{4.2}$$

We consider the estimators of the form

$$\widehat{\Sigma}_2(\boldsymbol{\Phi}) = \boldsymbol{S}_1^{1/2}\boldsymbol{Q}\boldsymbol{\Phi}(\Lambda)\boldsymbol{Q}'\boldsymbol{S}_1^{1/2}, \tag{4.3}$$

where $\boldsymbol{\Phi}(\Lambda) = \mathrm{diag}(\phi_1(\Lambda), \ldots, \phi_p(\Lambda))$. From (4.2), it is seen that the estimator $\widehat{\Sigma}_2(\boldsymbol{\Phi})$ is also represented as

$$\widehat{\Sigma}_2(\boldsymbol{\Phi}) = \boldsymbol{S}_2^{1/2}\boldsymbol{P}\Lambda^{1/2}\boldsymbol{\Phi}(\Lambda)\Lambda^{1/2}\boldsymbol{P}'\boldsymbol{S}_2^{1/2}. \tag{4.4}$$

We shall provide general conditions for the dominance of $\widehat{\Sigma}_2(\boldsymbol{\Phi})$ given by (4.3) in terms of the risk $R_2(\omega; \widehat{\Sigma}_2)$. Making the transformations, we can suppose that $\boldsymbol{S}_1 \sim \mathcal{W}_p(\Theta^{-1}, n_1)$ and $\boldsymbol{S}_2 \sim \mathcal{W}_p(\boldsymbol{I}, n_2)$ with any loss of generality, where $\Theta^{-1} = \mathrm{diag}(\theta_1^{-1}, \ldots, \theta_p^{-1})$ for $\theta_1^{-1} \le 1, \ldots, \theta_p^{-1} \le 1$. Therefore we can apply the results directly to get the improvements on $\widehat{\Sigma}_2(\boldsymbol{\Phi})$. The corresponding truncation rules are described as

$$\begin{aligned} \{\boldsymbol{\Phi}(\Lambda)\}^{TR} &= \mathrm{diag}\left(\phi_1^{TR}(\Lambda), \ldots, \phi_p^{TR}(\Lambda)\right), \\ \phi_i^{TR}(\Lambda) &= \max\left\{\phi_i(\Lambda), \frac{\lambda_i^{-1}+1}{n_1+n_2}\right\}, \quad i = 1, \ldots, p, \end{aligned} \tag{4.5}$$

and

$$\begin{aligned} \{\boldsymbol{\Phi}(\Lambda)\}^{TR*} &= \mathrm{diag}\left(\phi_1^{TR*}, \ldots, \phi_p^{TR*}(\Lambda)\right), \\ \phi_i^{TR*}(\Lambda) &= \begin{cases} (n_1+n_2)^{-1}(\lambda_i^{-1}+1) & \text{if } (n_1+n_2)^{-1}(\Lambda+\boldsymbol{I}) \ge \boldsymbol{\Phi}(\Lambda) \\ \phi_i(\Lambda) & \text{otherwise.} \end{cases} \end{aligned} \tag{4.6}$$

Then the corresponding truncated estimators are given by

$$\begin{aligned} \widehat{\Sigma}_2(\{\boldsymbol{\Phi}\}^{TR}) &= \boldsymbol{S}_1^{1/2}\boldsymbol{Q}\mathrm{diag}\left(\phi_1^{TR}(\Lambda), \ldots, \phi_p^{TR}(\Lambda)\right)\boldsymbol{Q}'\boldsymbol{S}_1^{1/2}, \\ \widehat{\Sigma}_2(\{\boldsymbol{\Phi}\}^{TR*}) &= \begin{cases} (n_1+n_2)^{-1}(\boldsymbol{S}_1+\boldsymbol{S}_2) & \text{if } (n_1+n_2)^{-1}(\boldsymbol{S}_1+\boldsymbol{S}_2) \ge \widehat{\Sigma}_2(\boldsymbol{\Phi}(\Lambda)) \\ \widehat{\Sigma}_2(\boldsymbol{\Phi}(\Lambda)) & \text{otherwise.} \end{cases} \end{aligned} \tag{4.7}$$

Similar to Theorem 1, we can verify that $\widehat{\Sigma}_2(\{\Phi\}^{TR})$ dominates $\widehat{\Sigma}_2(\{\Phi\}^{TR*})$ which is better than $\widehat{\Sigma}_2(\Phi)$ in terms of the risk $R_2(\omega;\widehat{\Sigma}_2)$.

Using these truncation rules, we can get several truncated estimators being better than unbiased or minimax estimators. For instance, applying the truncation rule $\{\Phi\}^{TR}$ to the unbiased estimator

$$\widehat{\Sigma}_2^{UB} = n_2^{-1}\boldsymbol{S}_2 = \boldsymbol{S}_1^{1/2}\boldsymbol{Q}\Phi^{UB}\boldsymbol{Q}'\boldsymbol{S}_1^{1/2}$$

for $\Phi^{UB} = \mathrm{diag}((n_2\lambda_1)^{-1},\ldots,(n_2\lambda_p)^{-1})$, we obtain the REML estimator

$$\widehat{\Sigma}_2^{REML} = \widehat{\Sigma}_2(\{\Phi^{UB}\}^{TR}), \tag{4.8}$$

improving upon $\widehat{\Sigma}_2^{UB}$, where

$$\{\Phi^{UB}\}^{TR} = \mathrm{diag}\left(\max\left\{\frac{\lambda_1^{-1}}{n_2}, \frac{\lambda_1^{-1}+1}{n_1+n_2}\right\},\ldots,\max\left\{\frac{\lambda_p^{-1}}{n_2}, \frac{\lambda_p^{-1}+1}{n_1+n_2}\right\}\right).$$

Also the minimax estimator corresponded to (2.12) for Σ_2 is given by

$$\widehat{\Sigma}_2^{m} = \widehat{\Sigma}_2(\Phi^m) = \boldsymbol{S}_1^{1/2}\boldsymbol{Q}\Phi^m(\Lambda)\boldsymbol{Q}'\boldsymbol{S}_1^{1/2},$$

where

$$\Phi^m(\Lambda) = \mathrm{diag}\left(\frac{e_p}{\lambda_1},\ldots,\frac{e_1}{\lambda_p}\right),$$

for $e_i = (n_2+p+1-2i)^{-1}$. It should be noted that the order of $e_1,\ldots,e_p$ in $\Phi^m(\Lambda)$ is reveresed to the case of $\Psi^m(\Lambda)$ in (2.12) because $\lambda_p^{-1}\geq\ldots\geq\lambda_1^{-1}$. Applying the truncation rule yields

$$\widehat{\Sigma}_2^{MTR} = \widehat{\Sigma}_2(\{\Phi^m\}^{TR}), \tag{4.9}$$

improving on $\widehat{\Sigma}_2^{m}$, where

$$\{\Phi^m(\Lambda)\}^{TR} = \mathrm{diag}\left(\max\left\{\frac{e_p}{\lambda_1}, \frac{\lambda_1^{-1}+1}{n_1+n_2}\right\},\ldots,\max\left\{\frac{e_1}{\lambda_p}, \frac{\lambda_p^{-1}+1}{n_1+n_2}\right\}\right).$$

We now construct estimators of Σ_A along the manner that $\widehat{\Sigma}_A = r^{-1}(\widehat{\Sigma}_2-\widehat{\Sigma}_1)$. It will be interesting to know the kind of nonnegative estimators that can be obtained by combining truncated estimators of Σ_1 and Σ_2. Combining $\widehat{\Sigma}_1([\Psi]^{TR})$ given by (2.5) and $\widehat{\Sigma}_2(\{\Phi\}^{TR})$ given by (4.7), and noting the expression (4.4), we get the estimator of Σ_A of the form

$$\begin{aligned}\widehat{\Sigma}_A([\Psi]^{TR},\{\Phi\}^{TR}) &= r^{-1}\left(\widehat{\Sigma}_2([\Psi]^{TR})-\widehat{\Sigma}_1(\{\Phi\}^{TR}\right)\\ &= r^{-1}\boldsymbol{S}_2^{1/2}\boldsymbol{P}\left\{\Lambda^{1/2}\{\Phi(\Lambda)\}^{TR}\Lambda^{1/2}-[\Psi(\Lambda)]^{TR}\right\}\boldsymbol{P}'\boldsymbol{S}_2^{1/2},\end{aligned}$$

where

$$
\begin{aligned}
& \Lambda^{1/2}\{\Phi(\Lambda)\}^{TR}\Lambda^{1/2} - [\Psi(\Lambda)]^{TR} \\
= & \ \mathrm{diag}\left(\max\left\{\phi_i(\Lambda)\lambda_i, \frac{\lambda_i+1}{n_1+n_2}\right\} - \min\left\{\psi_i(\Lambda), \frac{\lambda_i+1}{n_1+n_2}\right\}\right). \qquad (4.10)
\end{aligned}
$$

In the case of combining the REML estimators $\widehat{\Sigma}_1^{REML}$ and $\widehat{\Sigma}_2^{REML}$, the i-th diagonal element in (4.10) is

$$
\max\left\{\frac{1}{n_2}, \frac{\lambda_i+1}{n_1+n_2}\right\} - \min\left\{\frac{\lambda_i}{n_1}, \frac{\lambda_i+1}{n_1+n_2}\right\} = \max\left\{\frac{1}{n_2} - \frac{\lambda_i}{n_1}, 0\right\},
$$

which gives the estimator

$$
\begin{aligned}
\widehat{\Sigma}_A^{REML} &= r^{-1}(\widehat{\Sigma}_2^{REML} - \widehat{\Sigma}_1^{REML}) \\
&= r^{-1}\boldsymbol{S}_2^{1/2}\boldsymbol{P}\mathrm{diag}\left(\max\left\{\frac{1}{n_2} - \frac{\lambda_i}{n_1}, 0\right\}, i = 1, \dots, p\right)\boldsymbol{P}'\boldsymbol{S}_2^{1/2},
\end{aligned}
$$

which is n.n.d. This REML estimator of Σ_A is similar to the one proposed by Amemiya (1985). We thus get n.n.d. estimators $(\widehat{\Sigma}_1^{REML}, \widehat{\Sigma}_A^{REML})$ improving on $(\widehat{\Sigma}_1^{UB}, \widehat{\Sigma}_A^{UB})$ relative to the Kullback-Leibler loss (4.1).

In the case of combining improved minimax estimators $\widehat{\Sigma}_1^{MTR}$ and $\widehat{\Sigma}_2^{MTR}$, the i-th diagonal element in (4.10) is

$$
\begin{aligned}
& \max\left\{e_{p-i+1}, \frac{\lambda_i+1}{n_1+n_2}\right\} - \min\left\{d_i\lambda_i, \frac{\lambda_i+1}{n_1+n_2}\right\} \\
= & \ \max\left\{\frac{1}{n_2-(p+1-2i)} - \frac{\lambda}{n_1+p+1-2i}, 0\right\},
\end{aligned}
$$

which gives the estimator

$$
\begin{aligned}
& \widehat{\Sigma}_A^{MTR} = r^{-1}\left(\widehat{\Sigma}_2^{MTR} - \widehat{\Sigma}_1^{MTR}\right) \\
= & \ r^{-1}\boldsymbol{S}_2^{1/2}\boldsymbol{P}\mathrm{diag}\left(\max\left\{\frac{1}{n_2-(p+1-2i)} - \frac{\lambda_i}{n_1+p+1-2i}, 0\right\}, i = 1, \dots, p\right)\boldsymbol{P}'\boldsymbol{S}_2^{1/2}
\end{aligned}
$$

which is also n.n.d. In the sequel we get n.n.d. estimators $(\widehat{\Sigma}_1^{MTR}, \widehat{\Sigma}_A^{MTR})$ improving on $(\widehat{\Sigma}_1^{m}, \widehat{\Sigma}_A^{m})$ in terms of the risk $R_{KL}(\omega; \widehat{\Sigma}_1, \widehat{\Sigma}_A)$ where $\widehat{\Sigma}_A^{m} = r^{-1}(\widehat{\Sigma}_2^{m} - \widehat{\Sigma}_1^{m})$. Comparing two n.n.d. estimators $\widehat{\Sigma}_A^{REML}$ and Σ_A^{MTR}, we can note that for $i > (<)(p+1)/2$,

$$
\frac{1}{n_2-(p+1-2i)} - \frac{\lambda_i}{n_1+p+1-2i} > (<)\frac{1}{n_2} - \frac{\lambda_i}{n_1},
$$

which implies that

$$P\left[\frac{1}{n_2-(p+1-2i)}-\frac{\lambda_i}{n_1+p+1-2i}>0\right]>(<)P\left[\frac{1}{n_2}-\frac{\lambda_i}{n_1}>0\right].$$

Hence we cannot compare them in the sense of maximizing the probability that they are positive-definite.

5 Concluding Remarks

In this paper we have proposed n.n.d. estimators for the 'between' and 'within' covariance matrices. We considered a natural 'pivot' $\boldsymbol{S}_2^{-1/2}\boldsymbol{S}_1\boldsymbol{S}_2^{-1/2}$ instead of $\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$, the latter is even difficult to handle. Although it can be shown that the results of this paper hold for any factorization of $\boldsymbol{S}_2$, the symmetric factorization is easier to handle and from practical viewpoint can easily be obtained from any statistical packages. The restricted maximum likelihood estimators have been shown to perform better than the unbiased and the truncated estimators proposed in this paper. The proposed truncated estimators, however, are natural estimators and somewhat simpler to implement than restricted maximum likelihood estimators. This estimator also dominates the estimator proposed by Mathew *et al.* (1994) and modified by us.

Acknowledgements. The research of the first author was supported in part by Natural Sciences and Engineering Research Council of Canada. The research of the second author was supported in part by a grant from the Center for International Research on the Japanese Economy, the University of Tokyo, and by the Ministry of Education, Japan, Grant No. 09780214. This work was done during the visit of the second author to the University of Toronto, 1998 summer.

REFERENCES

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

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

Dey, D., and Srinivasan, C. (1985). Estimation of covariance matrix under Stein's loss. *Ann. Statist.*, **13**, 1581-1591.

Harville, W. (1977). Maximum likelihood approaches to variance components estimation and to related problems. *J. Amer. Statist. Assoc.*, **72**, 320-338.

James, W., and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* , **1**, 361-379. University of California Press, Berkeley.

Kubokawa, T. (1995). Estimation of variance components in mixed linear models. *J. Multivariate Anal.*, **53**, 210-236.

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

Mathew, T., Niyogi, A., and Sinha, B.K. (1994). Improved nonnegative estimation of variance components in balanced multivariate mixed models. *J. Multivariate Anal.*, **51**, 83-101.

Patterson, H.D., and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545-554.

Patterson, H.D., and Thompson, R. (1975). Maximum likelihood estimation of components of variance. In *Proceedings Eight International Biometric Conference* (L.C.A. Corsten and T. Postelnica, eds), 197-207. Biometric Soc., Washington, D.C.

Rao, C.R., and Kleffe, J. (1988). *Estimation of Variance Components and Applications.* North-Holland, Amsterdam.

Searle, S.R. (1971). *Linear Models.* Wiley, New York.

Sheena, Y., and Takemura, A. (1992). Inadmissibility of non-order-preserving orthogonally invariant estimators of the covariance matrix in the case of Stein's loss. *J. Multivariate Anal.*, **41**, 117-131.

Srivastava, M.S., and Khatri, C.G. (1979). *An Introduction to Multivariate Statistics.* North-Holland, New York.

Stein, C. (1977). Lectures on multivariate estimation theory. In *Investigation on Statistical Estimation Theory* I, 4-65. Zapiski Nauchych Seminarov LOMI im. V.A. Steklova AN SSSR vol. 74, Leningrad (in Russian).

Thompson, W.A., Jr. (1962). The problem of negative estimates of variance components. *Ann. Math. Statist.*, **33**, 273-289.