The Annals of Statistics
1999, Vol. 27, No. 6, 2008–2032

IMPROVED NONNEGATIVE ESTIMATION OF MULTIVARIATE COMPONENTS OF VARIANCE

BY M. S. SRIVASTAVA[1] AND T. KUBOKAWA[2]

University of Toronto and University of Tokyo

In this paper, we consider a multivariate one-way random effect model with equal replications. We propose nonnegative definite estimators for "between" and "within" components of variance. Under the Stein loss function, it is shown that the proposed estimators of the "within" component dominate the best unbiased estimator. Restricted maximum likelihood, truncated and order-preserving minimax estimators are also proposed. A Monte Carlo simulation is carried out to choose among these estimators.

For estimating the "between" component, we consider the Stein loss function for jointly estimating the two positive definite matrices ("within" and "within" plus "between") and obtain estimators for the "between" component dominating the best unbiased estimator. Other estimators as considered for "within" are also proposed. A Monte Carlo simulation is carried out to choose among these estimators.

1. Introduction. The estimation of variance components in univariate mixed linear models has 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. The unbiased estimators of "between" components of variance, however, take negative values with positive probability and considerable attention has been paid to provide positive estimators for between components. Herbach (1959) considered the balanced one-way layout and provided maximum likelihood estimators free from the above defect. Other important contributions in maximum likelihood and restricted maximum likelihood estimators are due to Thompson (1962), Patterson and Thompson (1971, 1975), Searle (1971) and Harville (1977). Estimators which are not only nonnegative but also improve on the unbiased estimators have been derived from a frequentist viewpoint by Mathew, Sinha and Sutradhar (1992) and Kubokawa (1995).

On the other hand, the estimation of variance components in the multivariate mixed linear model did not receive such attention, primarily due to technical difficulties encountered even in the balanced one-way random effect model with equal replications. For example, maximum likelihood estimators have been proposed in the literature in various forms but it is not known if

Received August 1998; revised September 3, 1999.

[1]Supported in part by Natural Sciences and Engineering Research Council of Canada.

[2]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, Grants 09780214 and 11680320.

AMS 1991 *subject classifications*. Primary 62H12, 62F30; secondary 62C12, 62C20.

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

these estimators dominate the usual uniformly minimum variance unbiased estimators other than being nonnegative definite. A brief review of the restricted maximum likelihood estimators appears in Anderson, Anderson and Olkin (1986). Bock and Vandenberg (1968) introduced simple estimators for the components which were later shown by Bock and Petersen (1975) to be maximum likelihood estimators. Earlier, Klotz and Putter (1969) had considered maximum likelihood estimation of the components but this solution is in a different form than Bock and Petersen (1975). Other contributors in this area of maximum likelihood estimation and testing are Rao (1983), Amemiya and Fuller (1984), Amemiya (1985) and Anderson, Anderson and Olkin (1986).

Calvin and Dykstra (1991) proposed restricted maximum likelihood (REML) estimators for the ordered covariances, but nothing is known about the properties of these estimators. Calvin and Dykstra (1991) also mention the computational difficulties encountered with the corresponding MINQUE theory given in Rao and Kleffe (1988). Recently Mathew, Niyogi and Sinha (1994) considered a 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 the best unbiased estimator.

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

$$\mathbf{y}_{ij} = \boldsymbol{\mu} + \mathbf{a}_i + \mathbf{e}_{ij}, \qquad i = 1, \dots, k, \;\; j = 1, \dots, r, \tag{1.1}$$

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

$$\mathbf{S}_1 \sim \mathscr{W}_p(\boldsymbol{\Sigma}_1, n_1) \quad \text{and} \quad \mathbf{S}_2 \sim \mathscr{W}_p(\boldsymbol{\Sigma}_2, n_2) \tag{1.2}$$

for

$$\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_1 + r\boldsymbol{\Sigma}_A, \qquad n_1 = k(r-1) \quad \text{and} \quad n_2 = k - 1, \tag{1.3}$$

where $\mathscr{W}_p(\boldsymbol{\Sigma}_1, n_1)$ designates the p-variate Wishart distribution with expectation $n_1\boldsymbol{\Sigma}_1$ 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 $\boldsymbol{\Sigma}_1$, $\widehat{\boldsymbol{\Sigma}}_1^{UB} = n_1^{-1}\mathbf{S}_1$, can be improved on by using the information from the order restriction $\boldsymbol{\Sigma}_1 \le \boldsymbol{\Sigma}_2$. This issue was discussed by Mathew, Niyogi and Sinha (1994) who considered the estimator

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

and showed that $\widehat{\boldsymbol{\Sigma}}_1^{MNS}$ dominates $\widehat{\boldsymbol{\Sigma}}_1^{UB}$ 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

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

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

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

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

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

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

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

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

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 $\widehat{\boldsymbol{\Sigma}}_1^{UB}$ for all p. This estimator is similar, in spirit to (1.6) and perhaps can be obtained from the pivot $\mathbf{S}_1^{-1/2}\mathbf{S}_2\mathbf{S}_1^{-1/2}$. 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 $\widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi})$ dominating another one $\widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi}^0)$. From this result, we get estimators improving on the unbiased estimator $\widehat{\boldsymbol{\Sigma}}_1^{UB}$ 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). Some of the minimax estimators are nonorder-preserving and they can be further improved upon by the order-preserving estimators as suggested by Stein (1975, 1977), Lin and Perlman (1985) and Sheena and Takemura (1992).

The problem of estimating the "between" multivariate component of variance $\boldsymbol{\Sigma}_A$ is treated in Section 3. The unbiased estimator of $\boldsymbol{\Sigma}_A$ is given by

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

which is not always nonnegative definite (n.n.d.). The (restricted) maximum likelihood estimators, on the other hand, are n.n.d. However, its superiority over the unbiased estimator has not been established from a decision-theoretical aspect. Mathew, Niyogi and Sinha (1994) considered another type of estimators, namely, linear combinations of $\mathbf{S}_1$ and $\mathbf{S}_2$ 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 squared 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 underestimate less than the overestimate. As an alternative measure, we employ the Kullback–Leibler distance or Stein loss function and consider the estimation of $\mathbf{\Sigma}_A$ in the context of simultaneous estimation of $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_A$. Under this measure, the results given in Section 2 are directly applicable to get estimators improving on the unbiased estimators $(\widehat{\mathbf{\Sigma}}_1^{UB}, \mathbf{\Sigma}_A^{UB})$. From this result, it is shown that the REML estimators of $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_A)$ dominate the unbiased estimators. Also n.n.d. estimators superior to minimax estimators of $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_A$ are derived. Monte Carlo simulations are carried out in Section 4 to choose among different estimators. The paper concludes in Section 5.

2. Estimation of the multivariate "within" component of variance.

2.1. *A general approach to improving estimators.* Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be independent random matrices, $\mathbf{S}_i \sim \mathscr{W}_p(\mathbf{\Sigma}_i, n_i)$, $i = 1, 2$, with $\mathbf{\Sigma}_1 \le \mathbf{\Sigma}_2$ where $\mathbf{\Sigma}_1 \le \mathbf{\Sigma}_2$ denotes that $\mathbf{\Sigma}_2 - \mathbf{\Sigma}_1$ is n.n.d. Denote the parameter space by $\Omega = \{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \,|\, \mathbf{\Sigma}_1 \le \mathbf{\Sigma}_2\}$. Suppose that we want to estimate $\mathbf{\Sigma}_1$ relative to the Stein (or entropy) loss function

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

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

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

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

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

$$\mathbf{P}'\mathbf{S}_2^{-1/2}\mathbf{S}_1\mathbf{S}_2^{-1/2}\mathbf{P} = \mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_p),$$

where $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$. We consider estimators of the form

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

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

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

and

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

Then the corresponding truncated estimators are written as

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

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

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

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

PROOF. Without any loss of generality, let $\boldsymbol{\Sigma}_1 = \mathbf{I}$ and $\boldsymbol{\Sigma}_2 = \boldsymbol{\Theta} = \operatorname{diag}(\theta_1, \dots, \theta_p)$ with $\theta_1 \ge 1, \dots, \theta_p \ge 1$. The joint density of $\mathbf{S}_1$ and $\mathbf{S}_2$ is

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

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

$$
\begin{aligned}
&f_{F,S_2}(\mathbf{F}, \mathbf{S}_2)\\
&\quad = \text{const.}|\mathbf{F}|^{(n_1-p-1)/2}|\mathbf{S}_2|^{(n_1+n_2-p-1)/2}|\boldsymbol{\Theta}|^{-n_2/2}\operatorname{etr}\left[-\tfrac{1}{2}(\mathbf{F}+\boldsymbol{\Theta}^{-1})\mathbf{S}_2\right].
\end{aligned}
\tag{2.6}
$$

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

$$
\begin{aligned}
&f_{\Lambda,P,S_2}(\boldsymbol{\Lambda}, \mathbf{P}, \mathbf{S}_2)\\
&\quad = \text{const.} f_p(\mathbf{P})g(\boldsymbol{\Lambda})|\mathbf{S}_2|^{(n_1+n_2-p-1)/2}|\boldsymbol{\Theta}|^{-n_2/2}\operatorname{etr}\left[-\tfrac{1}{2}(\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'+\boldsymbol{\Theta}^{-1})\mathbf{S}_2\right],
\end{aligned}
$$

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

$$\mathbf{S}_2 \mid (\mathbf{\Lambda}, \mathbf{P}) \sim \mathscr{W}_p\big((\mathbf{P\Lambda P}' + \mathbf{\Theta}^{-1})^{-1}, n_1 + n_2\big),$$

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

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

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

$$\begin{aligned} &R_1\big(\mathbf{\Theta}, \widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi})\big) - R_1\big(\mathbf{\Theta}, \widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}]^{TR})\big) \\ &\quad = E_\Theta\big[\operatorname{tr}(\mathbf{P\Psi}(\mathbf{\Lambda})\mathbf{P}' - \mathbf{P}[\mathbf{\Psi}(\mathbf{\Lambda})]^{TR}\mathbf{P}')\mathbf{S}_2 - \log|\mathbf{\Psi}(\mathbf{\Lambda})\{[\mathbf{\Psi}(\mathbf{\Lambda})]^{TR}\}^{-1}|\big] \\ &\quad = E_\Theta^{\Lambda, P}\big[\operatorname{tr}(\mathbf{\Psi}(\mathbf{\Lambda}) - [\mathbf{\Psi}(\mathbf{\Lambda})]^{TR})\mathbf{P}'E_\Theta[\mathbf{S}_2 \mid \mathbf{\Lambda}, \mathbf{P}]\mathbf{P} \\ &\qquad\qquad - \log|\mathbf{\Psi}(\mathbf{\Lambda})\{[\mathbf{\Psi}(\mathbf{\Lambda})]^{TR}\}^{-1}|\big]. \end{aligned} \tag{2.8}$$

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

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

which proves part (i).

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

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

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

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

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

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

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

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

PROOF. Define the set A by

$$\begin{aligned} A &= \{\mathbf{I}+\mathbf{S}_1^{-1}\mathbf{S}_2 \le (n_1+n_2-a)/n_1\mathbf{I}\} \\ &= \left\{\frac{\lambda_i}{\lambda_i+1}\frac{n_1+n_2-a}{n_1} \ge 1,\ i=1,\dots,p\right\}. \end{aligned}$$

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

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

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

$$B = \left\{\frac{\lambda_i}{\lambda_i+1}\frac{n_1+n_2}{n_1} \ge 1,\ i=1,\dots,p\right\},$$

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

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

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

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

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

where

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

The truncation rules given in Section 2.1 produce the estimators

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

where

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

and

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

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 $[\boldsymbol{\Psi}^{UB}]^{TR}$ takes the value

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

while

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

From Theorem 1, we get the following corollary.

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

It may be noted that Hara (1999) has also recently obtained this result by a different method.

The estimator $\widehat{\boldsymbol{\Sigma}}_1^{REML}$ is known to be the restricted (or residual) maximum likelihood (REML) estimator of $\boldsymbol{\Sigma}_1$ under the constraint $\boldsymbol{\Sigma}_1 \le \boldsymbol{\Sigma}_2$. Corollary 2 implies that the REML estimator is superior not only to the unbiased estimator but also to $\widehat{\boldsymbol{\Sigma}}_1^{UTR*}$ and $\widehat{\boldsymbol{\Sigma}}_1^{UTR*}(a)$, although $\widehat{\boldsymbol{\Sigma}}_1^{UTR*}$ appears to have a natural form.

It is interesting to note that the REML estimator $\widehat{\mathbf{\Sigma}}_1^{REML}$ can also be derived as an empirical Bayes rule. Let $\boldsymbol{\eta} = \mathbf{\Sigma}_2^{-1}$ and $\boldsymbol{\xi} = \mathbf{\Sigma}_2^{1/2}\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_2^{1/2}$. Suppose that $\boldsymbol{\eta}$ has noninformative 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 $(\boldsymbol{\eta}, \mathbf{S}_1, \mathbf{S}_2)$ has the form

$$\text{const.}|\boldsymbol{\eta}|^{(n_1+n_2-p-1)/2}|\mathbf{S}_1|^{(n_1-p-1)/2}|\mathbf{S}_2|^{(n_2-p-1)/2}|\boldsymbol{\xi}|^{n_1/2}\ \operatorname{etr}\big[-\tfrac{1}{2}(\boldsymbol{\xi}^{1/2}\mathbf{S}_1\boldsymbol{\xi}^{1/2} + \mathbf{S}_2)\boldsymbol{\eta}\big],$$

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

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

We thus get the Bayes estimator of $\mathbf{\Sigma}_1$ under the Stein loss (2.1),

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

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

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

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

$$\mathbf{Q}'\mathbf{S}_1^{-1/2}\mathbf{S}_2\mathbf{S}_1^{-1/2}\mathbf{Q} = \operatorname{diag}(\lambda_1^{-1}, \ldots, \lambda_p^{-1}).$$

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

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

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

$$\begin{aligned}
\mathbf{S}_2^{-1/2}\mathbf{S}_1\mathbf{S}_2^{-1/2} &= (\mathbf{S}_2^{-1/2}\mathbf{S}_1^{1/2})(\mathbf{S}_2^{-1/2}\mathbf{S}_1^{1/2})' = \mathbf{P}\mathbf{\Lambda}\mathbf{P}',\\
\mathbf{S}_1^{-1/2}\mathbf{S}_2\mathbf{S}_1^{-1/2} &= (\mathbf{S}_1^{-1/2}\mathbf{S}_2^{1/2})(\mathbf{S}_1^{-1/2}\mathbf{S}_2^{1/2})' = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}'.
\end{aligned}$$

Then we have that

$$\mathbf{S}_2^{-1/2}\mathbf{S}_1^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{Q}' \quad \text{or} \quad \mathbf{S}_2^{1/2}\mathbf{P} = \mathbf{S}_1^{1/2}\mathbf{Q}\mathbf{\Lambda}^{-1/2}.$$

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

$$\widehat{\boldsymbol{\Sigma}}_1(\widehat{\boldsymbol{\xi}}) = \mathbf{S}_2^{1/2}\mathbf{P}\ \mathrm{diag}\left(\min\left\{\frac{\lambda_i}{n_1}, \frac{\lambda_i+1}{n_1+n_2}\right\},\ i=1,\dots,p\right)\mathbf{P}'\mathbf{S}_2^{1/2},$$

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

2.3. *Improvements on the minimax estimator.* Historically, the first interesting result in estimation of the covariance matrix was provided by James and Stein (1961), who established the nonminimaxity of the unbiased estimator $\widehat{\boldsymbol{\Sigma}}_1^{UB}$ and presented the minimax estimator of the form

$$\widehat{\boldsymbol{\Sigma}}_1^{JS} = \mathbf{T}_1\mathbf{D}^m\mathbf{T}_1',$$

where $\mathbf{T}_1$ is a lower triangular $p \times p$ matrix such that $\mathbf{S}_1 = \mathbf{T}_1\mathbf{T}_1'$, and $\mathbf{D}^m$ is the diagonal matrix given by $\mathbf{D}^m = \mathrm{diag}(d_1, \dots, d_p)$ for

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

It is known that the James–Stein minimax estimator $\widehat{\boldsymbol{\Sigma}}_1^{JS}$ has a drawback that it depends on the coordinate system. For modifying this property, several orthogonally equivariant minimax estimators have been proposed in the literature. The orthogonally equivariant estimators are generally given by

$$\widetilde{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi}) = \mathbf{R}\boldsymbol{\Psi}(\mathbf{L})\mathbf{R}', \tag{2.12}$$

where $\mathbf{R}$ is an orthogonal matrix such that $\mathbf{S}_1 = \mathbf{RLR}'$ and $\mathbf{L} = \mathrm{diag}(\ell_1, \dots, \ell_p)$ for eigenvalues $\ell_1 \geq \cdots \geq \ell_p$. It is important to note that the estimator $\widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi})$ is equivariant under the scale transformation $\mathbf{S}_1 \to \mathbf{AS}_1\mathbf{A}'$ and $\mathbf{S}_2 \to \mathbf{AS}_2\mathbf{A}'$ for any $p \times p$ nonsingular matrix $\mathbf{A}$, so that $\widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi})$ does not depend on the coordinate system.

PROPOSITION 1. *If the orthogonally equivariant estimator* $\widetilde{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi}(\mathbf{L}))$ *is minimax, then for the same function* $\boldsymbol{\Psi}(\cdot)$, $\widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi}(\boldsymbol{\Lambda}))$ *is scale-equivariant, minimax and improving on* $\widehat{\boldsymbol{\Sigma}}_1^{JS}$ *relative to the Stein loss* (2.1).

PROOF. Recall that $\mathbf{F} = \mathbf{S}_2^{-1/2}\mathbf{S}_1\mathbf{S}_2^{-1/2} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$ and that $\mathbf{S}_1 \sim \mathscr{W}_p(\mathbf{I}, n_1)$. Then it is seen that the conditional distribution of $\mathbf{F}$ given $\mathbf{S}_2$ has $\mathscr{W}_p(\boldsymbol{\Sigma}_*, n_1)$ for $\boldsymbol{\Sigma}_* = \mathbf{S}_2^{-1}$. Then the risk function of $\widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi})$ is represented by

$$\begin{aligned} &R_1(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi})) \\ &\quad = E^{S_2}\big[E^{F\,|\,S_2}[\mathrm{tr}\,\mathbf{P}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\mathbf{P}'\boldsymbol{\Sigma}_*^{-1} - \log|\mathbf{P}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\mathbf{P}'\boldsymbol{\Sigma}_*^{-1}| - p \,|\, \mathbf{S}_2]\big], \end{aligned} \tag{2.13}$$

so that given $\mathbf{S}_2$, conditionally $\mathbf{P}\boldsymbol{\Psi}\mathbf{P}'$ corresponds to the Stein's orthogonally invariant estimator $\widetilde{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi})$ of $\boldsymbol{\Sigma}_*$ with $\mathbf{S}_1 \sim \mathscr{W}(\boldsymbol{\Sigma}^*, n)$. Hence the minimaxity of $\widetilde{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi})$ implies the minimaxity of $\widehat{\boldsymbol{\Sigma}}_1(\boldsymbol{\Psi})$, which proves Proposition 1. □

From Proposition 1, we can obtain some minimax estimators by using the results derived previously for the estimation of $\mathbf{\Sigma}_1$.

2.3.1. *Stein-type estimator.* Let $\widehat{\mathbf{\Sigma}}_1^S = \widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi}^S)$ for

$$\mathbf{\Psi}^S(\mathbf{\Lambda}) = \operatorname{diag}(d_1\lambda_1, \dots, d_p\lambda_p). \tag{2.14}$$

The minimaxity of $\widehat{\mathbf{\Sigma}}_1^S$ follows from the result of Dey and Srinivasan (1985), who also give another orthogonally equivariant estimator beating $\widetilde{\mathbf{\Sigma}}_1^S$ for $p \geq 3$.

2.3.2. *Takemura-type estimator.* Stein (1956), Eaton (1970) and Takemura (1984) gave an orthogonally equivariant and improved estimator, which can be represented in our problem as

$$\widehat{\mathbf{\Sigma}}^T = \mathbf{S}_2^{1/2}\left\{\int_{O(p)} \mathbf{\Gamma}\mathbf{U}_\Gamma\mathbf{D}_m\mathbf{U}_\Gamma'\mathbf{\Gamma}'\, d\mu(\mathbf{\Gamma})\right\}\mathbf{S}_2^{1/2}, \tag{2.15}$$

where $\mathbf{U}_\Gamma \in G_T^+$ with $\mathbf{U}_\Gamma\mathbf{U}_\Gamma' = \mathbf{\Gamma}'\mathbf{F}\mathbf{\Gamma}$ for $\mathbf{F} = \mathbf{S}_2^{-1/2}\mathbf{S}_1\mathbf{S}_2^{-1/2} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$. Takemura (1984) provided another expression as $\widehat{\mathbf{\Sigma}}^T = \widehat{\mathbf{\Sigma}}(\mathbf{\Psi}^T)$ for $\mathbf{\Psi}^T(\mathbf{\Lambda}) = \operatorname{diag}(\psi_1^T, \dots, \psi_p^T)$, where

$$(\psi_1^T, \dots, \psi_p^T)' = \operatorname{diag}(\lambda_1, \dots, \lambda_p)\mathbf{W}(\mathbf{\Lambda})(d_1, \dots, d_p)', \tag{2.16}$$

for $p \times p$ doubly stochastic matrix $\mathbf{W}(\mathbf{\Lambda})$. Also Takemura (1984) gave exact expressions for $\mathbf{\Psi}^T(\mathbf{\Lambda})$ for $p = 2$ and 3. For instance,

$$\psi_1^T = \lambda_1\left(\frac{\sqrt{\lambda_1}}{\sqrt{\lambda_1}+\sqrt{\lambda_2}}d_1 + \frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}+\sqrt{\lambda_2}}d_2\right),$$
$$\psi_2^T = \lambda_2\left(\frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}+\sqrt{\lambda_2}}d_1 + \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_1}+\sqrt{\lambda_2}}d_2\right)$$

for $p = 2$. However, the explicit calculation of $\mathbf{W}(\mathbf{\Lambda})$ for $p > 3$ remains an intractable problem.

2.3.3. *Perron-type estimator.* Perron (1992) gave an approximation to $\mathbf{W}(\mathbf{\Lambda})$, say $\widetilde{\mathbf{W}}(\mathbf{\Lambda})$, with a doubly stochastic property, and showed the minimaxity of the approximated estimator. Let

$$\tilde{w}_{ij}(\mathbf{\Lambda}) = \frac{\operatorname{tr}_{j-1}(\mathbf{\Lambda}_i)}{\operatorname{tr}_{j-1}(\mathbf{\Lambda})} - \frac{\operatorname{tr}_j(\mathbf{\Lambda}_i)}{\operatorname{tr}_j(\mathbf{\Lambda})},$$

for

$$\operatorname{tr}_j(\mathbf{\Lambda}) = \begin{cases} 1, & \text{if } j = 0, \\ \displaystyle\sum_{1\leq i_1<\cdots<i_j\leq p}\prod_{k=1}^{j}\lambda_{i_k}, & \text{if } j = 1, \dots, p, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\mathbf{\Lambda}_i = \operatorname{diag}(\lambda_1, \dots, \lambda_{i-1}, 0, \lambda_{i+1}, \dots, \lambda_p).$$

Let $\widetilde{\mathbf{W}}(\mathbf{\Lambda}) = (\tilde{w}_{ij})$ and put

$$(\psi_1^P, \ldots, \psi_p^P)' = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)\widetilde{\mathbf{W}}(\mathbf{\Lambda})(d_1, \ldots, d_p)'. \tag{2.17}$$

For $p = 2$, they are given by

$$\psi_1^P = \lambda_1\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}d_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2}d_2\right),$$

$$\psi_2^P = \lambda_1\left(\frac{\lambda_2}{\lambda_1 + \lambda_2}d_1 + \frac{\lambda_1}{\lambda_1 + \lambda_2}d_2\right).$$

Then the result of Perron (1992) implies the minimaxity of the scale equivariant estimator $\widehat{\mathbf{\Sigma}}_1^P = \widehat{\Sigma}_1(\mathbf{\Psi}^P)$ for $\mathbf{\Psi}^P = \operatorname{diag}(\psi_1^P, \ldots, \psi_p^P)$.

2.3.4. *Haff-type estimator.* Let

$$\widehat{\mathbf{\Sigma}}_1^H = \frac{1}{n_1}\left(\mathbf{S}_1 + \frac{a_0}{\operatorname{tr}\mathbf{S}_1^{-1}\mathbf{S}_2}\mathbf{S}_2\right). \tag{2.18}$$

From the result of Haff (1980), it can be verified that $\widehat{\mathbf{\Sigma}}^H$ dominates the unbiased estimator $\widehat{\mathbf{\Sigma}}_1^{UB}$ when $0 < a_0 \leq 2(p-1)/n_1$. Then $\widehat{\mathbf{\Sigma}}_1^H$ is expressed as $\widehat{\mathbf{\Sigma}}_1^H = \widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi}^H)$ by letting $\mathbf{\Psi}^H = n_1^{-1}\mathbf{\Lambda} + a_0(\operatorname{tr}\mathbf{\Lambda}^{-1})^{-1}\mathbf{I}$.

Yang and Berger (1994) derived an orthogonally equivariant estimator as a Bayes rule against the reference prior distribution, and we can construct a scale equivariant one corresponding to it. Since it is difficult to express the estimator in an explicit form, we shall not consider this estimator in this paper. However, for some numerical investigations, see Sugiura and Ishibayashi (1997).

Now, applying the truncation rule (2.3) given in Section 2.1 to the Stein-type estimator $\widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi}^S)$, we get the corresponding improved estimator,

$$\widehat{\mathbf{\Sigma}}_1^{STR} = \widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}^S]^{TR}), \tag{2.19}$$

where

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

and

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

Similarly, we get truncated estimators

$$\widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}^T]^{TR}), \qquad \widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}^P]^{TR}) \quad \text{and} \quad \widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}^H]^{TR})$$

corresponding to the above scale-equivariant estimators. From Theorem 1 and Proposition 1, we can get the following corollary.

COROLLARY 3. *For* $\mathbf{\Psi} = \mathbf{\Psi}^S$, $\mathbf{\Psi}^T$ *and* $\mathbf{\Psi}^P$, *the estimator* $\widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}]^{TR})$ *is scale-equivariant, minimax and improving on the corresponding estimator* $\widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi})$ *and truncated estimator* $\widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}]^{TR*})$ *relative to the Stein loss* (2.1). *Also* $\widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}^H]^{TR})$ *dominates* $\widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi}^H)$ *and* $\widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}^H]^{TR*})$.

It should be noted that Corollary 3 does not imply the dominance of $\widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}]^{TR})$ over $\widetilde{\mathbf{\Sigma}}_1(\mathbf{\Psi})$, but states the dominance of $\widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}]^{TR})$ over $\widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi})$. Although $\widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi})$ is not identical to $\widetilde{\mathbf{\Sigma}}_1(\mathbf{\Psi})$, if $\widetilde{\mathbf{\Sigma}}_1(\mathbf{\Psi})$ is a superior minimax estimator, $\widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi})$ inherits the same good risk properties with minimaxity and improvement. Corollary 3 states that these minimax estimators can be further improved on by $\widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}]^{TR})$ by employing the information in $\mathbf{S}_2$.

2.4. *Dominance results by order-preserving estimation.* We consider the general type of estimators given by (2.2), namely,

$$\widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi}) = \mathbf{S}_2^{1/2}\mathbf{P}\mathbf{\Psi}(\mathbf{\Lambda})\mathbf{P}'\mathbf{S}_2^{1/2}, \qquad \mathbf{\Psi}(\mathbf{\Lambda}) = \operatorname{diag}(\psi_1(\mathbf{\Lambda}), \dots, \psi_p(\mathbf{\Lambda})).$$

For the diagonal elements $\psi_1(\mathbf{\Lambda}), \dots, \psi_p(\mathbf{\Lambda})$, it is quite natural to satisfy the condition

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

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

We here show that nonorder-preserving estimators can be improved on by the order-preserving estimators. We first write the risk function of $\widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi})$ as

$$\begin{aligned} &R_1(\mathbf{\Theta}, \widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi})) \\ &\quad = E_\Theta\big[\operatorname{tr}\mathbf{P}\mathbf{\Psi}(\mathbf{\Lambda})\mathbf{P}'(\mathbf{P}\mathbf{\Lambda}\mathbf{P}' + \mathbf{\Theta}^{-1})^{-1} - \log|\mathbf{\Psi}(\mathbf{\Lambda})| - \log|\mathbf{S}_2| - p\big] \\ &\quad = E_\Theta\big[\operatorname{tr}\ \mathbf{\Psi}(\mathbf{\Lambda})(\mathbf{\Lambda} + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P})^{-1} - \log|\mathbf{\Psi}(\mathbf{\Lambda})| - \log|\mathbf{S}_2| - p\big]. \end{aligned} \tag{2.20}$$

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

LEMMA 1. *Let* $E[\cdot \mid \mathbf{\Lambda}]$ *be a conditional expectation with respect to* $\mathbf{P}$ *given* $\mathbf{\Lambda}$. *For* $i < j$,

$$E[\mathbf{B}^{ii} \mid \mathbf{\Lambda}] \le E[\mathbf{B}^{jj} \mid \mathbf{\Lambda}]. \tag{2.21}$$

Let $\widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi})$ be a nonorder-preserving estimator. Let $\psi_i^O(\mathbf{\Lambda})$ be the ith largest element in $(\psi_1(\mathbf{\Lambda}), \dots, \psi_p(\mathbf{\Lambda}))$, so that $\psi_1^O(\mathbf{\Lambda}) \ge \cdots \ge \psi_p^O(\mathbf{\Lambda})$. Note that

$(\psi_1^O, \dots, \psi_p^O)$ majorizes $(\psi_1, \dots, \psi_p)$, that is,

$$\sum_{i=1}^{j} \psi_i^O \ge \sum_{i=1}^{j} \psi_i \text{ for } 1 \le j \le p-1 \text{ and } \sum_{i=1}^{p} \psi_i^O = \sum_{i=1}^{p} \psi_i. \tag{2.22}$$

Let $\widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi}^O) = \mathbf{S}_2^{1/2}\mathbf{P}\mathbf{\Psi}^O(\mathbf{\Lambda})\mathbf{P}'\mathbf{S}_2^{1/2}$ for $\mathbf{\Psi}^O(\mathbf{\Lambda}) = \operatorname{diag}(\psi_1^O(\mathbf{\Lambda}), \dots, \psi_p^O(\mathbf{\Lambda}))$. Then we get the theorem.

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

PROOF. The risk difference is written as

$$\begin{aligned} R_1(\mathbf{\Theta}, \widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi}^O)) - R_1(\mathbf{\Theta}, \widehat{\mathbf{\Sigma}}_1(\mathbf{\Psi})) &= E_\Theta[\operatorname{tr}(\mathbf{\Psi}^O(\mathbf{\Lambda}) - \mathbf{\Psi}(\mathbf{\Lambda}))\mathbf{B}^{-1}] \\ &= E_\Theta^\Lambda\left[\sum_{i=1}^{p}(\psi_i^O(\mathbf{\Lambda}) - \psi_i(\mathbf{\Lambda}))E[\mathbf{B}^{ii} \mid \mathbf{\Lambda}]\right]. \end{aligned} \tag{2.23}$$

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

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

which can be seen to be negative from Lemma 1 and (2.22). Hence from (2.23), Theorem 2 is proved. □

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

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

where $[[\mathbf{\Psi}^S]^{TR}]^O = \operatorname{diag}(\psi_1^{STRO}, \dots, \psi_p^{STRO})$ and ψ_i^{STRO} is the ith largest in the diagonal elements $\min\{d_i\lambda_i, (n_1+n_2)^{-1}(\lambda_i+1)\}$, $i = 1, \dots, p$.

3. 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), $\mathbf{S}_1$ and $\mathbf{S}_2$ are independent random matrices having $\mathscr{W}_p(\mathbf{\Sigma}_1, n_1)$ and $\mathscr{W}_p(\mathbf{\Sigma}_2, n_2)$, respectively, for $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_1 + r\mathbf{\Sigma}_A$. We want to estimate $\mathbf{\Sigma}_A$ based on $\mathbf{S}_1$ and $\mathbf{S}_2$ and to discuss the preference of estimators in a decision-theoretic framework. The parametric structure $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_1 + r\mathbf{\Sigma}_A$ means that estimators of $\mathbf{\Sigma}_A$ can be provided through estimation of both $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$. This suggests that the estimation of $\mathbf{\Sigma}_A$ may be considered in the context of the simultaneous estimation of $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$.

We thus consider the problem of estimating $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ simultaneously relative to the Kullback–Leibler loss function,

$$\begin{aligned} L_{KL}(\widehat{\mathbf{\Sigma}}_1, \widehat{\mathbf{\Sigma}}_A; \mathbf{\Sigma}_1, \mathbf{\Sigma}_A) &= n_1 L(\widehat{\mathbf{\Sigma}}_1 \mathbf{\Sigma}_1^{-1}) \\ &\quad + n_2 L\big((\widehat{\mathbf{\Sigma}}_1 + r\widehat{\mathbf{\Sigma}}_A)(\mathbf{\Sigma}_1 + r\mathbf{\Sigma}_A)^{-1}\big), \end{aligned} \tag{3.1}$$

for the function $L(\cdot)$ given by (2.1). This loss can be really derived from the Kullback–Leibler distance

$$\int [\log\{f(\mathbf{S}_1, \mathbf{S}_2; \widehat{\mathbf{\Sigma}}_1, \widehat{\mathbf{\Sigma}}_A)/f(\mathbf{S}_1, \mathbf{S}_2; \mathbf{\Sigma}_1, \mathbf{\Sigma}_A)\}] f(\mathbf{S}_1, \mathbf{S}_2; \widehat{\mathbf{\Sigma}}_1, \widehat{\mathbf{\Sigma}}_A)\, d\nu(\mathbf{S}_1)\, d\nu(\mathbf{S}_2)$$

for joint density function $f(\mathbf{S}_1, \mathbf{S}_2; \mathbf{\Sigma}_1, \mathbf{\Sigma}_A)$.

When $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_1 + r\mathbf{\Sigma}_A$ are estimated by $\widehat{\mathbf{\Sigma}}_1$ and $\widehat{\mathbf{\Sigma}}_2$, it is quite natural to take the form $\widehat{\mathbf{\Sigma}}_A = r^{-1}(\widehat{\mathbf{\Sigma}}_2 - \widehat{\mathbf{\Sigma}}_1)$ as an estimator of $\mathbf{\Sigma}_A$. As long as such types of estimators are treated, the risk function of $(\widehat{\mathbf{\Sigma}}_1, \widehat{\mathbf{\Sigma}}_A)$ relative to the Kullback–Leibler loss (3.1) is written as

$$\begin{aligned} R_{KL}(\omega; \widehat{\mathbf{\Sigma}}_1, \widehat{\mathbf{\Sigma}}_A) &= E_\omega\big[L_{KL}(\widehat{\mathbf{\Sigma}}_1, \widehat{\mathbf{\Sigma}}_A; \mathbf{\Sigma}_1, \mathbf{\Sigma}_A)\big] \\ &= n_1 R_1(\omega; \widehat{\mathbf{\Sigma}}_1) + n_2 R_2(\omega; \widehat{\mathbf{\Sigma}}_2), \end{aligned}$$

where $\omega = (\mathbf{\Sigma}_1, \mathbf{\Sigma}_1 + r\mathbf{\Sigma}_A) \in \Omega$ and

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

Hence the original problem under the loss (3.1) is decomposed into two problems of estimating $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ in terms of the risk functions $R_1(\omega; \widehat{\mathbf{\Sigma}}_1)$ and $R_2(\omega; \widehat{\mathbf{\Sigma}}_2)$, respectively.

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

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

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

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

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

so that we note that the following relation holds:

$$\mathbf{S}_2^{1/2}\mathbf{P} = \mathbf{S}_1^{1/2}\mathbf{Q}\mathbf{\Lambda}^{-1/2}. \tag{3.2}$$

We consider the estimators of the form

$$\widehat{\mathbf{\Sigma}}_2(\mathbf{\Phi}) = \mathbf{S}_1^{1/2}\mathbf{Q}\mathbf{\Phi}(\mathbf{\Lambda})\mathbf{Q}'\mathbf{S}_1^{1/2}, \tag{3.3}$$

where $\mathbf{\Phi}(\mathbf{\Lambda}) = \mathrm{diag}(\phi_1(\mathbf{\Lambda}), \ldots, \phi_p(\mathbf{\Lambda}))$. From (3.2), it is seen that the estimator $\widehat{\mathbf{\Sigma}}_2(\mathbf{\Phi})$ is also represented as

$$\widehat{\mathbf{\Sigma}}_2(\mathbf{\Phi}) = \mathbf{S}_2^{1/2}\mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{\Phi}(\mathbf{\Lambda})\mathbf{\Lambda}^{1/2}\mathbf{P}'\mathbf{S}_2^{1/2}. \tag{3.4}$$

We shall provide general conditions for the dominance of $\widehat{\mathbf{\Sigma}}_2(\mathbf{\Phi})$ given by (3.3) in terms of the risk $R_2(\omega; \widehat{\mathbf{\Sigma}}_2)$. Making the transformations, we can suppose that $\mathbf{S}_1 \sim \mathscr{W}_p(\mathbf{\Theta}^{-1}, n_1)$ and $\mathbf{S}_2 \sim \mathscr{W}_p(\mathbf{I}, n_2)$ without any loss of generality, where $\mathbf{\Theta}^{-1} = \mathrm{diag}\ (\theta_1^{-1}, \ldots, \theta_p^{-1})$ for $\theta_1^{-1} \le 1, \ldots, \theta_p^{-1} \le 1$. Therefore we can apply the results directly to get the improvements on $\widehat{\mathbf{\Sigma}}_2(\mathbf{\Phi})$. The truncation rules corresponding to (2.3) and (2.4) are described as

$$\begin{aligned} \{\mathbf{\Phi}(\mathbf{\Lambda})\}^{TR} &= \mathrm{diag}\big(\phi_1^{TR}(\mathbf{\Lambda}), \ldots, \phi_p^{TR}(\mathbf{\Lambda})\big), \\ \phi_i^{TR}(\mathbf{\Lambda}) &= \max\left\{\phi_i(\mathbf{\Lambda}), \frac{\lambda_i^{-1}+1}{n_1+n_2}\right\}, \qquad i = 1, \ldots, p \end{aligned} \tag{3.5}$$

and

$$\begin{aligned} \{\mathbf{\Phi}(\mathbf{\Lambda})\}^{TR*} &= \mathrm{diag}\big(\phi_1^{TR*}, \ldots, \phi_p^{TR*}(\mathbf{\Lambda})\big), \\ \phi_i^{TR*}(\mathbf{\Lambda}) &= \begin{cases} (n_1+n_2)^{-1}(\lambda_i^{-1}+1), & \text{if } (n_1+n_2)^{-1}(\mathbf{\Lambda}+\mathbf{I}) \ge \mathbf{\Phi}(\mathbf{\Lambda}), \\ \phi_i(\mathbf{\Lambda}), & \text{otherwise,} \end{cases} \end{aligned} \tag{3.6}$$

respectively, and the resulting truncated estimators are given by

$$\begin{aligned} \widehat{\mathbf{\Sigma}}_2(\{\mathbf{\Phi}\}^{TR}) &= \mathbf{S}_1^{1/2}\mathbf{Q}\ \mathrm{diag}\big(\phi_1^{TR}(\mathbf{\Lambda}), \ldots, \phi_p^{TR}(\mathbf{\Lambda})\big)\mathbf{Q}'\mathbf{S}_1^{1/2}, \\ \widehat{\mathbf{\Sigma}}_2(\{\mathbf{\Phi}\}^{TR*}) &= \begin{cases} (n_1+n_2)^{-1}(\mathbf{S}_1+\mathbf{S}_2), \\ \qquad\qquad \text{if } (n_1+n_2)^{-1}(\mathbf{S}_1+\mathbf{S}_2) \ge \widehat{\mathbf{\Sigma}}_2(\mathbf{\Phi}(\mathbf{\Lambda})), \\ \widehat{\mathbf{\Sigma}}_2(\mathbf{\Phi}(\mathbf{\Lambda})), \qquad \text{otherwise.} \end{cases} \end{aligned} \tag{3.7}$$

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

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

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

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

$$\widehat{\mathbf{\Sigma}}_2^{REML} = \widehat{\mathbf{\Sigma}}_2(\{\mathbf{\Phi}^{UB}\}^{TR}), \tag{3.8}$$

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

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

Also the Stein type minimax estimator corresponding to (2.14) for $\boldsymbol{\Sigma}_2$ is given by

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

where

$$\boldsymbol{\Phi}^S(\boldsymbol{\Lambda}) = \operatorname{diag}\left(\frac{e_p}{\lambda_1}, \dots, \frac{e_1}{\lambda_p}\right),$$

for $e_i = (n_2 + p + 1 - 2i)^{-1}$. It should be noted that the order of $e_1, \dots, e_p$ in $\boldsymbol{\Phi}^S(\boldsymbol{\Lambda})$ is reversed to the case of $\boldsymbol{\Psi}^S(\boldsymbol{\Lambda})$ in (2.14) because $\lambda_p^{-1} \geq \cdots \geq \lambda_1^{-1}$. Applying the truncation rule yields

$$\widehat{\boldsymbol{\Sigma}}_2^{STR} = \widehat{\boldsymbol{\Sigma}}_2(\{\boldsymbol{\Phi}^S\}^{TR}), \tag{3.9}$$

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

$$\{\boldsymbol{\Phi}^S(\boldsymbol{\Lambda})\}^{TR} = \operatorname{diag}\left(\max\left\{\frac{e_p}{\lambda_1}, \frac{\lambda_1^{-1}+1}{n_1+n_2}\right\}, \dots, \max\left\{\frac{e_1}{\lambda_p}, \frac{\lambda_p^{-1}+1}{n_1+n_2}\right\}\right).$$

We now construct estimators of $\boldsymbol{\Sigma}_A$ along the manner that $\widehat{\boldsymbol{\Sigma}}_A = r^{-1}(\widehat{\boldsymbol{\Sigma}}_2 - \widehat{\boldsymbol{\Sigma}}_1)$. It will be interesting to know the kind of nonnegative estimators that can be obtained by combining truncated estimators of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. Combining $\widehat{\boldsymbol{\Sigma}}_1([\boldsymbol{\Psi}]^{TR})$ given by (2.5) and $\widehat{\boldsymbol{\Sigma}}_2(\{\boldsymbol{\Phi}\}^{TR})$ given by (3.7), and noting the expression (3.4), we get the estimator of $\boldsymbol{\Sigma}_A$ of the form

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

where

$$\begin{aligned}&\boldsymbol{\Lambda}^{1/2}\{\boldsymbol{\Phi}(\boldsymbol{\Lambda})\}^{TR}\boldsymbol{\Lambda}^{1/2} - [\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\\ &\quad= \operatorname{diag}\left(\max\left\{\phi_i(\boldsymbol{\Lambda})\lambda_i, \frac{\lambda_i+1}{n_1+n_2}\right\} - \min\left\{\psi_i(\boldsymbol{\Lambda}), \frac{\lambda_i+1}{n_1+n_2}\right\}\right).\end{aligned} \tag{3.10}$$

In the case of combining the REML estimators $\widehat{\boldsymbol{\Sigma}}_1^{REML}$ and $\widehat{\boldsymbol{\Sigma}}_2^{REML}$, the ith diagonal element in (3.10) 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 n.n.d. estimator

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

This REML estimator of $\boldsymbol{\Sigma}_A$ is similar to the one proposed by Amemiya (1985). We thus get n.n.d. estimators $(\widehat{\boldsymbol{\Sigma}}_1^{REML}, \widehat{\boldsymbol{\Sigma}}_A^{REML})$ improving on $(\widehat{\boldsymbol{\Sigma}}_1^{UB}, \widehat{\boldsymbol{\Sigma}}_A^{UB})$ relative to the Kullback–Leibler loss (3.1). It may be noted that Hara (1999) has also recently obtained this dominance result by a different method.

In the case of combining improved minimax estimators $\widehat{\boldsymbol{\Sigma}}_1^{STR} = \widehat{\boldsymbol{\Sigma}}_1([\boldsymbol{\Psi}^S]^{TR})$ and $\widehat{\boldsymbol{\Sigma}}_2^{STR}$, the ith diagonal element in (3.10) is

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

which gives the estimator

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

which is also n.n.d. In the sequel we get n.n.d. estimators $(\widehat{\boldsymbol{\Sigma}}_1^{STR}, \widehat{\boldsymbol{\Sigma}}_A^{STR})$ improving on $(\widehat{\boldsymbol{\Sigma}}_1^{JS}, \widehat{\boldsymbol{\Sigma}}_A^{JS})$ in terms of the risk $R_{KL}(\omega; \widehat{\boldsymbol{\Sigma}}_1, \widehat{\boldsymbol{\Sigma}}_A)$ where $\widehat{\boldsymbol{\Sigma}}_A^{JS} = r^{-1}(\widehat{\boldsymbol{\Sigma}}_2^{JS} - \widehat{\boldsymbol{\Sigma}}_1^{JS})$ for the James–Stein minimax estimator $\widehat{\boldsymbol{\Sigma}}_2^{JS}$ of $\boldsymbol{\Sigma}_2$. Comparing two n.n.d. estimators $\widehat{\boldsymbol{\Sigma}}_A^{REML}$ and $\boldsymbol{\Sigma}_A^{STR}$, we 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.

4. Simulation results. In this section, we investigate the risk behavior of several estimators proposed in this paper through Monte Carlo simulation. The reported risks are the average of 50,000 replications for $p = 2$ and $r = 2$. We choose $n_1 = 6$, $n_2 = 7$ and the covariance matrices as $\mathbf{\Sigma}_1 = \mathrm{diag}(1, a)$ and $\mathbf{\Sigma}_2 = \mathrm{diag}((b+1)/2, ab)$ for various values of $b \geq 1$ and a is chosen as 1 and 10 in our simulation.

We first investigate the risk performance of the estimators of $\mathbf{\Sigma}_1$. For the sake of simplicity, we shall denote the estimators $\widehat{\mathbf{\Sigma}}_1^{UB}$, $\widehat{\mathbf{\Sigma}}_1^{JS}$ and $\widehat{\mathbf{\Sigma}}_1^{REML}$ by UB, JS and REML, respectively; the original estimators $\widetilde{\mathbf{\Sigma}}_1(\Psi^S)$, $\widetilde{\mathbf{\Sigma}}_1(\Psi^T)$ and $\widetilde{\mathbf{\Sigma}}_1(\Psi^P)$ and their induced estimators $\widehat{\mathbf{\Sigma}}_1(\Psi^S)$, $\widehat{\mathbf{\Sigma}}_1(\Psi^T)$ and $\widehat{\mathbf{\Sigma}}_1(\Psi^P)$ by S, T, P and S*, T*, P*, respectively. We shall denote the truncated estimators $\widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}^S]^{TR})$, $\widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}^T]^{TR})$ and $\widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}^P]^{TR})$ by STR, TTR and PTR, respectively. Similarly, the estimators $\widehat{\mathbf{\Sigma}}_1^H$ and $\widehat{\mathbf{\Sigma}}_1([\mathbf{\Psi}^H]^{TR})$ with $a_0 = (p-1)/n$ will be denoted by H* and HTR, respectively. Table 1 reports the values of the risks of the estimators UB, JS, REML, S, T, P, S*, T*, P*, H*, STR, TTR, PTR and HTR for $b = 1$, 3, 5, 11 and 21. Also the risk behaviors of the estimators UB, JS, REML, S*, STR, TTR, PTR and HTR are given in Figure 1 for $1 \leq b \leq 31$.

Table 1 and Figure 1 reveal that:

1. There are no dominance relations between the estimators S, T, P and STR, TTR, PTR.
2. HTR $\succ$ STR $\succ$ PTR $\succ$ TTR $\succ$ REML $\succ$ JS $\succ$ UB when b is close to one, where $\delta_A \succ \delta_B$ means that δ_A is better than δ_B.
3. STR $\succ$ PTR $\succ$ HTR $\succ$ TTR $\succ$ JS $\succ$ REML $\succ$ UB when b is larger than 6.
4. The risks of STR, TTR and PTR approach those of S*, T* and P*, respectively, as b increases.
5. REML is not as good as the truncated minimax estimators STR, TTR and PTR.

TABLE 1

Risks of the estimators UB, JS, S, T, P, REML, S*, *STR,* T*, *TTR,* P*, *PTR,* H* *and HTR in estimation of* $\mathbf{\Sigma}_1$

a	**UB**	**JS**	**S**	**T**	**P**
1	0.485	0.464	0.374	0.445	0.417
10	0.485	0.464	0.452	0.450	0.444

b	**REML**	**S***	**STR**	**T***	**TTR**	**P***	**PTR**	**H***	**HTR**
1	0.421	0.413	0.378	0.447	0.395	0.428	0.386	0.446	0.357
3	0.445	0.413	0.397	0.447	0.419	0.429	0.407	0.446	0.394
5	0.462	0.416	0.408	0.447	0.432	0.429	0.419	0.447	0.418
11	0.478	0.418	0.416	0.447	0.445	0.430	0.427	0.447	0.439
21	0.483	0.416	0.418	0.447	0.447	0.430	0.429	0.447	0.445

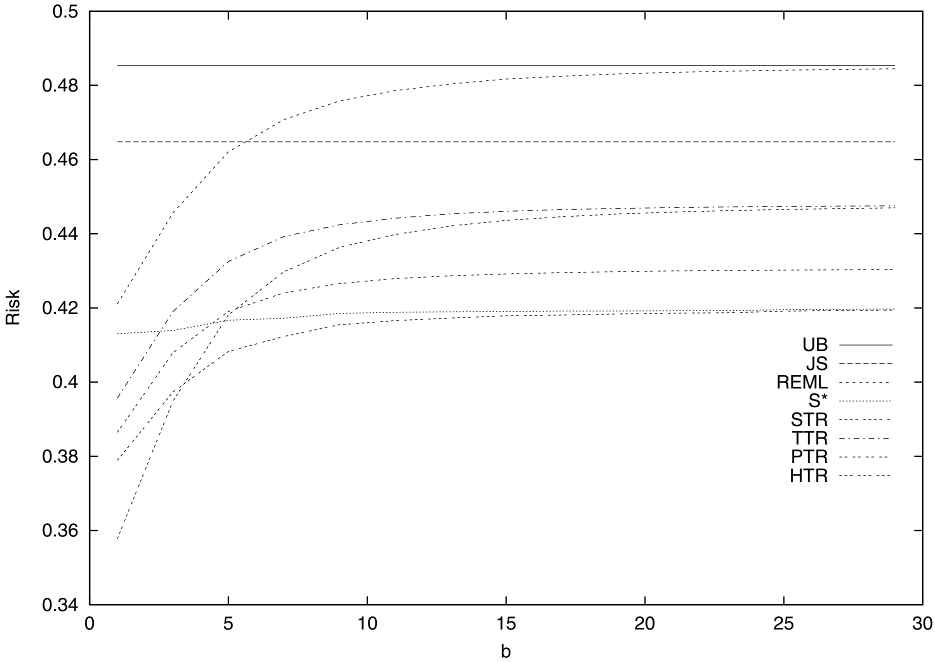

FIG. 1. *Risks of the estimator UB, JS, REML, S^*, STR, TTR, PTR and HTR in estimation of* $\mathbf{\Sigma}_1$.

For significant risk reduction near $\mathbf{\Sigma}_2 = \mathbf{I}$, the truncated Stein type estimator STR or its order-preserving estimator will be recommended for practical use.

The risk performance in simultaneous estimation of $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_A)$ is investigated in Figure 2.

For simplicity, denote $(\widehat{\mathbf{\Sigma}}_1^{UB}, \widehat{\mathbf{\Sigma}}_A^{UB})$, $(\widehat{\mathbf{\Sigma}}_1^{JS}, \widehat{\mathbf{\Sigma}}_A^{JS})$, $(\widehat{\mathbf{\Sigma}}_1^{REML}, \widehat{\mathbf{\Sigma}}_A^{REML})$ and $(\widehat{\mathbf{\Sigma}}_1^{STR}, \widehat{\mathbf{\Sigma}}_A^{STR})$ by UB, JS, REML and STR, respectively. Figure 2 reports their risk behaviors for $1 \le b \le 31$, and reveals that STR has the best risk behavior of the four.

Finally, we investigate the risk behavior of the estimators of the "between" component $\mathbf{\Sigma}_A$. We have proposed the two estimators $\widehat{\mathbf{\Sigma}}_A^{REML}$ and $\widehat{\mathbf{\Sigma}}_A^{STR}$ for $\mathbf{\Sigma}_A$ in Section 3. It would be desirable to compare the performance of these estimators with the one proposed by Mathew, Niyogi and Sinha (1994) and given by

$$\widehat{\mathbf{\Sigma}}_A^{MNS} = r^{-1}\mathbf{S}_2^{1/2}\mathbf{P}\,\mathrm{diag}\left(\max\left\{\frac{1}{n_2+2} - \frac{\lambda_i}{n_1+2}, 0\right\},\ i = 1, \dots, p\right)\mathbf{P}'\mathbf{S}_2^{1/2};$$

this has a good risk performance under the quadratic loss

$$L_1(\widehat{\mathbf{\Sigma}}_A, \mathbf{\Sigma}_A) = \mathrm{tr}\big(\widehat{\mathbf{\Sigma}}_A\mathbf{\Sigma}_A^{-1} - \mathbf{I}\big)^2$$

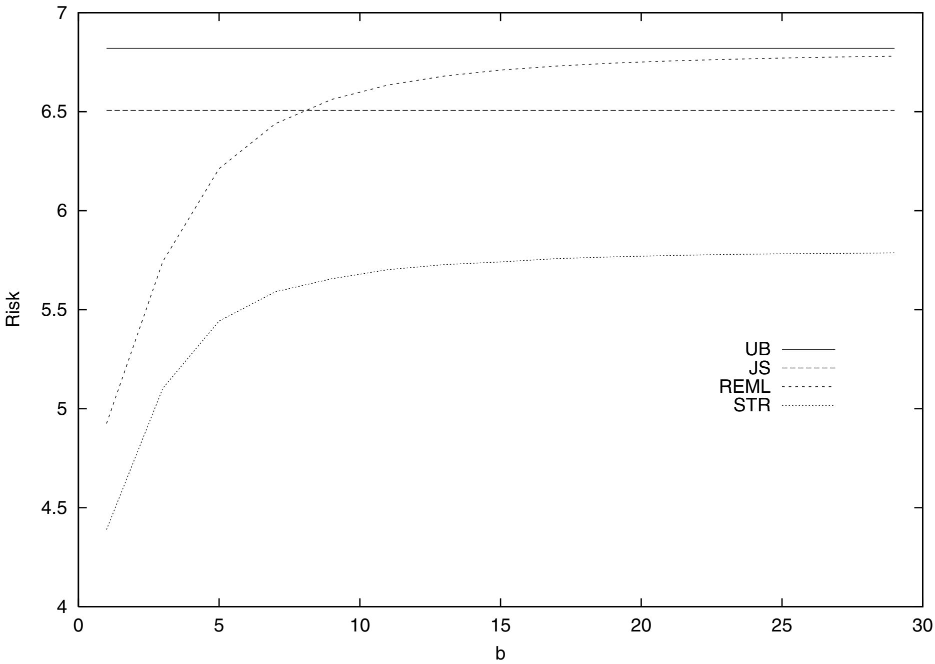

FIG. 2. *Risks of the estimators UB, JS, REML and STR in simultaneous estimation of* $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_A)$.

since the coefficients $(n_2+2)^{-1}$ and $(n_1+2)^{-1}$ are the best in a sense under the quadratic loss. Thus we compare these three nonnegative definite estimators relative to the L_1 loss and the following loss functions:

$$L_2(\widehat{\mathbf{\Sigma}}_A, \mathbf{\Sigma}_A) = \left\{\mathrm{tr}\big(\widehat{\mathbf{\Sigma}}_A \mathbf{\Sigma}_A^{-1} - \mathbf{I}\big)^2\right\}^{1/2},$$

$$L_3(\widehat{\mathbf{\Sigma}}_A, \mathbf{\Sigma}_A) = (1+d)^{-1}\ \mathrm{tr}(\widehat{\mathbf{\Sigma}}_A \mathbf{\Sigma}_A^{-1} - \mathbf{I}) - \log|\widehat{\mathbf{\Sigma}}_A \mathbf{\Sigma}_A^{-1} + d\mathbf{I}| + 2\log(1+d),$$

where $L_3(\widehat{\mathbf{\Sigma}}_A, \mathbf{\Sigma}_A)$ is convex and $L_3(\mathbf{\Sigma}_A, \mathbf{\Sigma}_A) = 0$ for positive constant d. Table 2 reports the results of the simulation experiments for $p = 2$, $r = 2$, $n_1 = 6$, $n_2 = 7$, $\mathbf{\Sigma}_1 = \mathbf{I}$ and $\mathbf{\Sigma}_A = c\mathbf{I}$, $c = 0.05$, 0.1, 0.2, 0.4, 1.0, 2.0 and $d = 0.0001$. From Table 2, we see that $\widehat{\mathbf{\Sigma}}_A^{MNS} \succ \widehat{\mathbf{\Sigma}}_A^{STR} \succ \widehat{\mathbf{\Sigma}}_A^{REML}$ for the loss functions L_1 and L_2.

For the loss function L_3, it is concluded that $\widehat{\mathbf{\Sigma}}_A^{STR}$ is superior to both $\widehat{\mathbf{\Sigma}}_A^{MNS}$ and $\widehat{\mathbf{\Sigma}}_A^{REML}$. The numerical results in Table 2, for the loss function L_3, suggest that the performance of $\widehat{\mathbf{\Sigma}}_A^{REML}$ and $\widehat{\mathbf{\Sigma}}_A^{MNS}$ is somewhat mixed. It appears that $\widehat{\mathbf{\Sigma}}_A^{MNS} \succ \widehat{\mathbf{\Sigma}}_A^{REML}$, most of the time.

5. Concluding remarks. In this paper we have proposed n.n.d. estimators for the "between" and "within" covariance matrices. We considered a nat-

TABLE 2

Risks of the estimators $\widehat{\boldsymbol{\Sigma}}_A^{MNS}$, $\widehat{\boldsymbol{\Sigma}}_A^{REML}$ and $\widehat{\boldsymbol{\Sigma}}_A^{STR}$, denoted by MNS, REML, STR, under the L_1, L_2 and L_3 loss functions

	L_1			L_2			L_3		
c	**MNS**	**REML**	**STR**	**MNS**	**REML**	**STR**	**MNS**	**REML**	**STR**
0.05	60.3	11.5	72.2	2.24	2.62	2.24	12.2	14.2	12.3
0.1	17.1	33.4	21.1	1.70	1.98	1.73	9.47	10.3	9.20
0.2	5.73	11.1	7.30	1.36	1.56	1.40	7.71	8.07	7.05
0.4	2.53	4.79	3.21	1.16	1.31	1.19	6.13	6.22	5.07
1.0	1.32	2.29	1.59	1.01	1.11	1.02	3.79	3.71	2.65
2.0	1.00	1.57	1.16	0.95	1.02	0.95	2.23	2.12	1.43

ural "pivot" $\mathbf{S}_2^{-1/2}\mathbf{S}_1\mathbf{S}_2^{-1/2}$ instead of $\mathbf{S}_1^{-1/2}\mathbf{S}_2\mathbf{S}_1^{-1/2}$; the latter is even difficult to handle. Although it can be shown that the results of this paper hold for any factorization of $\mathbf{S}_2$, the symmetric factorization is easier to handle and from a practical viewpoint can easily be obtained from any statistical packages.

Monte Carlo simulation suggests that for both "between" and "within" components, the Stein type truncated estimators STR perform best.

APPENDIX

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

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

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

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

where $f_p(\mathbf{P})$ is a Jacobian and $g(\boldsymbol{\Lambda})$ is a function of $\boldsymbol{\Lambda}$. Hence inequality (2.21) is equivalent to

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

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

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

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

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

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

$$\text{(A.3)} \qquad -\int_{O(p)} (\mathbf{B}_{22}^f - \mathbf{B}_{11}^f)|\mathbf{\Lambda}^* + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}|^{-(n_1+n_2-p+1)/2}\, d\mu(\mathbf{P}) \geq 0,$$

where $\mathbf{\Lambda}^* = \operatorname{diag}(\lambda_2, \lambda_1, \lambda_3, \ldots, \lambda_p)$. Adding (A.2) and (A.3), we see that for $\alpha = -(n_1 + n_2 - p + 1)/2$, $E[\mathbf{B}^{22} \mid \mathbf{\Lambda}] \geq E[\mathbf{B}^{11} \mid \mathbf{\Lambda}]$ if and only if

$$\text{(A.4)} \qquad \int_{O(p)} (\mathbf{B}_{22}^f - \mathbf{B}_{11}^f)\{|\mathbf{\Lambda} + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}|^{\alpha} - |\mathbf{\Lambda}^* + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}|^{\alpha}\}\, d\mu(\mathbf{P}) \geq 0.$$

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

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

Then we have

$$\begin{aligned} \mathbf{B}_{22}^f - \mathbf{B}_{11}^f &= \begin{vmatrix} a_{11} + \lambda_1 & \mathbf{a}_{13}' \\ \mathbf{a}_{13} & \mathbf{A}_{33} + \mathbf{\Lambda}_3 \end{vmatrix} - \begin{vmatrix} a_{22} + \lambda_2 & \mathbf{a}_{23}' \\ \mathbf{a}_{23} & \mathbf{A}_{33} + \mathbf{\Lambda}_3 \end{vmatrix} \\ &= (\lambda_1 - \lambda_2)|\mathbf{A}_{33} + \mathbf{\Lambda}_3| + \begin{vmatrix} a_{11} & \mathbf{a}_{13}' \\ \mathbf{a}_{13} & \mathbf{A}_{33} + \mathbf{\Lambda}_3 \end{vmatrix} - \begin{vmatrix} a_{22} & \mathbf{a}_{23}' \\ \mathbf{a}_{23} & \mathbf{A}_{33} + \mathbf{\Lambda}_3 \end{vmatrix}, \end{aligned}$$

where $\mathbf{\Lambda}_3 = \operatorname{diag}(\lambda_3, \ldots, \lambda_p)$. On the other hand, for $x = \lambda_1 - \lambda_2$,

$$\begin{aligned} |\mathbf{\Lambda}^* + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}| &= \begin{vmatrix} \lambda_1 + a_{11} - x & a_{12} & \mathbf{a}_{13}' \\ a_{12} & \lambda_2 + a_{22} + x & \mathbf{a}_{23}' \\ \mathbf{a}_{13} & \mathbf{a}_{23} & \mathbf{A}_{33} + \mathbf{\Lambda}_3 \end{vmatrix} \\ &= |\mathbf{\Lambda} + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}| + x\begin{vmatrix} \lambda_1 + a_{11} & \mathbf{a}_{13}' \\ \mathbf{a}_{13} & \mathbf{A}_{33} + \mathbf{\Lambda}_3 \end{vmatrix} \\ &\quad -x\begin{vmatrix} \lambda_2 + a_{22} & \mathbf{a}_{23}' \\ \mathbf{a}_{23} & \mathbf{A}_{33} + \mathbf{\Lambda}_3 \end{vmatrix} - x^2|\mathbf{A}_{33} + \mathbf{\Lambda}_3| \\ &= \big|\mathbf{\Lambda} + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}\big| + \big|\mathbf{A}_{33} + \mathbf{\Lambda}_3\big|(x\lambda_1 - x\lambda_2 - x^2) + k_{P,\Lambda}x \\ &= |\mathbf{\Lambda} + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}| + k_{P,\Lambda}(\lambda_1 - \lambda_2), \end{aligned}$$

where

$$k_{P,\Lambda} = k(\mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}, \mathbf{\Lambda}_3) = \begin{vmatrix} a_{11} & \mathbf{a}_{13}' \\ \mathbf{a}_{13} & \mathbf{A}_{33} + \mathbf{\Lambda}_3 \end{vmatrix} - \begin{vmatrix} a_{22} & \mathbf{a}_{23}' \\ \mathbf{a}_{23} & \mathbf{A}_{33} + \mathbf{\Lambda}_3 \end{vmatrix}.$$

Therefore inequality (A.4) is represented by

$$\begin{aligned} &\int_{O(p)} \{(\lambda_1 - \lambda_2)|\mathbf{A}_{33} + \mathbf{\Lambda}_3| + k_{P,\Lambda}\} \\ &\qquad \times \Big\{\big|\mathbf{\Lambda} + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}\big|^{\alpha} - \big(|\mathbf{\Lambda} + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}| + k_{P,\Lambda}(\lambda_1 - \lambda_2)\big)^{\alpha}\Big\}\, d\mu(\mathbf{P}) \geq 0. \end{aligned}$$

Since $|\mathbf{A}_{33} + \mathbf{\Lambda}_3|$ does not depend on the above permutation of exchanging 1 and 2,

$$\int_{O(p)} (\lambda_1 - \lambda_2)|\mathbf{A}_{33} + \mathbf{\Lambda}_3| \\ \times \left\{ \left|\mathbf{\Lambda} + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}\right|^{\alpha} - \left(\left|\mathbf{\Lambda} + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}\right| + k_{P,\Lambda}(\lambda_1 - \lambda_2)\right)^{\alpha} \right\} d\mu(\mathbf{P}) = 0.$$

Since $(|\mathbf{\Lambda} + \mathbf{P}'\mathbf{\Theta}^{-1}\mathbf{P}| + k_{P,\Lambda}(\lambda_1 - \lambda_2))^{\alpha}$ is a decreasing function of $k_{P,\Lambda}$ for $\lambda_1 - \lambda_2 > 0$ and $\alpha = -(n_1 + n_2 - p + 1)/2$, it is seen that $k_{P,\Lambda} \geq$ (resp. $<$) 0 if and only if

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

so that for any $k_{P,\Lambda}$,

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

This establishes inequality (A.4), and Lemma 1 is proved. □

Acknowledgments. We thank the Editor, the Associate Editors and the referee for their constructive suggestions and comments and for providing some references.

This work was done during the visit of the second author to the University of Toronto, summer, 1998.

REFERENCES

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

AMEMIYA, Y. and FULLER, W. (1984). Estimation for the multivariate errors-in-variables model with estimated error covariance matrix. *Ann. Statist.* **12** 497–509.

ANDERSON, B. M., ANDERSON, T. W. and OLKIN, I. (1986). Maximum likelihood estimators and likelihood ratio criteria in multivariate components of variance. *Ann. Statist.* **14** 405–417.

BOCK, R. D. and PETERSEN, A. C. (1975). A multivariate correction for attenuation. *Biometrika* **62** 673–678.

BOCK, R. D. and VANDENBERG, S. G. (1968). Components of heritable variation in mental test scores. In *Progress in Human Behavior Genetics* (S. G. Vandenberg, ed.) 233–260. Johns Hopkins Univ. Press.

CALVIN, J. A. and DYKSTRA, R. L. (1991). 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.

DEY, D. and SRINIVASAN, C. (1985). Estimation of covariance matrix under Stein's loss. *Ann. Statist.* **13** 1581–1591.

EATON, M. L. (1970). Some problems in covariance estimation. Technical Report 49, Stanford Univ.

HAFF, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* **8** 586–597.

HARA, H. (1999). Estimation of covariance matrix and mean squared error for shrinkage estimators in multivariate normal distribution. Ph.D. dissertation, Faculty of Engineering, Univ. Tokyo.

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

HERBACH, L. H. (1959). Properties of model II-type analysis of variance tests. *Ann. Math. Statist.* **30** 939–959.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379. Univ. California Press, Berkeley.

KLOTZ, J. and PUTTER, J. (1969). Maximum likelihood estimation of multivariate covariance components for the balanced one-way layout. *Ann. Math. Statist.* **40** 1100–1105.

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

LIN, S. P. and PERLMAN, M. D. (1985). A Monte Carlo comparison of four estimators of a covariance matrix. In *Multivariate Analysis* (P. R. Krishnaiah, ed.) **6** 411–429. North-Holland, Amsterdam.

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, Amsterdam.

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 Eighth International Biometric Conference* (L. C. A. Corsten and T. Postelnica, eds.) 197–207. Biometric Society, Washington, D.C.

PERRON, F. (1992). Minimax estimators of a covariance matrix. *J. Multivariate Anal.* **43** 16–28.

RAO, C. R. (1983). Likelihood ratio tests for relationships between two covariance matrices. In *Studies in Econometrics, Time Series, and Multivariate Statistics* (S. Karlin, T. Amemiya, and I. A. Goodman, eds.) 529–543. Academic Press, New York.

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, Amsterdam.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206. Univ. California Press, Berkeley.

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 **74**, Leningrad (in Russian).

SUGIURA, N. and ISHIBAYASHI, H. (1997). Reference prior Bayes estimator for bivariate normal covariance matrix with risk comparison. *Comm. Statist. Theory Methods* **26** 2203–2221.

TAKEMURA, A. (1984). An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population. *Tsukuba J. Math.* **8** 367–376.

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

YANG, R. and BERGER, J. O. (1994). Estimation of a covariance matrix using the reference prior. *Ann. Statist.* **22** 1195–1211.

DEPARTMENT OF STATISTICS
UNIVERSITY OF TORONTO
100 ST. GEORGE STREET
TORONTO, ONTARIO M5S 3G3
CANADA
E-MAIL: srivasta@utstat.toronto.edu

FACULTY OF ECONOMICS
UNIVERSITY OF TOKYO
7-3-1 BUNKYO-KU
TOKYO 113
JAPAN