

CIRJE-F-60

**Estimation of Variance and  
Covariance Components in  
Elliptically Contoured Distributions**

Tatsuya Kubokawa  
University of Tokyo

September 1999

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.

# Estimation of Variance and Covariance Components in Elliptically Contoured Distributions

TATSUYA KUBOKAWA  
*University of Tokyo*

In this paper, we consider the problems of estimating the univariate and multivariate components of variance in *elliptically contoured distribution (ECD) model* in a decision-theoretic setup. Empirical Bayes or generalized Bayes estimators and several other positive or nonnegative (definite) estimators improving upon usual ANOVA (unbiased) estimators of the variance components are obtained. The robustness of these dominance results is also investigated, and it is shown that all dominance results under normal models remain true within a specific class of distributions.

*AMS 1991 subject classifications.* Primary 62C15; secondary 62F11, 62H12, 62J05.

*Key words and phrases:* Elliptically contoured distribution, robustness of improvement, variance components model, mixed linear regression model, statistical decision theory, point estimation.

## 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 attention 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), Kubokawa (1995) and Kubokawa, Saleh and Konno (1999) from a frequentist-view point. Recently Srivastava and Kubokawa (1999) succeeded in extending the dominance results to estimation of covariance components in a multivariate mixed linear model with equal replications.

The dominance results in estimation of the variance components are essentially equivalent to those in estimation of ordered scale parameters and are related to the result of Stein (1964), who constructed a truncated procedure improving upon an unbiased estimator for a

variance of a normal distribution with an unknown mean. These estimation issues are technically characterized to be resolved without employing the integration-by-parts methods, namely, the Stein and Haff identities, which are heavily exploited to construct improved procedures in estimation of a mean vector and a covariance matrix of a multivariate normal distribution. Recently Kubokawa and Srivastava (1999a,b) extended these identities to an elliptically contoured distribution model, and showed that almost all dominance results in the normal model remain true in the elliptically contoured distribution model by utilizing the extended identities. Then one will have the query of whether such robustness of improvement holds for the estimation of the variance components in which the identities are of no use.

In this paper, we consider the problem of estimating the variance components  $\sigma_1^2$ ,  $\sigma_A^2$  and their function  $\sigma_2^2 = \sigma_1^2 + r\sigma_A^2$  and of estimating multivariate components of covariance in *elliptically contoured distribution models*, hereafter referred to as *ECD models*, in a decision-theoretic setup. Empirical Bayes or generalized Bayes estimators and several other positive or nonnegative (definite) estimators improving upon usual unbiased estimators of the variance components are obtained. These dominance results are discussed here in the ECD models from two aspects of parametric and semiparametric situations, which correspond to the cases where the function  $f(\cdot)$  describing the ECD model is known and unknown, respectively. In the parametric case, it is shown that the structure of the dominance results still holds in the ECD models. In the semiparametric case, however, the superiority of improved estimators in the normal distribution does not hold for every unknown function  $f(\cdot)$ . When the ECD model are restricted to the class of multivariate  $t$ -distributions, the dominance results in estimation of  $\sigma_1^2$  remain true, that is, the robust improvements are guaranteed while the robustness of the dominance never holds for estimation of  $\sigma_2^2$ . This phenomenon demonstrates that two estimation issues of  $\sigma_1^2$  and  $\sigma_2^2$  with  $\sigma_1^2 \leq \sigma_2^2$  possess different stories from the respect of the robust improvement. It is of quite interest to point out that there exists a class of distributions in which the dominance results in the normal distribution remain true. Denote a class of the following distributions by  $C_0$ :

$$C_0 = \left\{ f : f(\|\mathbf{x}\|^2) = \frac{\nu^{\nu/2+1}\Gamma((N+\nu)/2+1)}{\pi^{N/2}\Gamma(\nu/2+1)}(\nu + \|\mathbf{x}\|^2)^{-(N+\nu)/2+1}, \nu > 0 \right\},$$

which is slightly different from the multivariate  $t$ -distributions. Within this class, the robustness of the superiority of all estimators given in the normal model is established.

Section 2 deals with the estimation of the variance components in a one-way random effects model, which is represented in the ECD model with covariance having the intraclass correlation structure. In Subsection 2.2, broad classes of estimators improving on ANOVA (unbiased) estimators of the variance components are derived. Out of these classes, empirical Bayes and generalized Bayes rules are developed. These are extensions of results in Kubokawa *et al.* (1999) to the ECD model. In Subsection 2.3, we discuss the robustness of the dominance results given under the normality, and provide the class  $C_0$  in which the results of Kubokawa *et al.* (1999) hold still for any  $f(\cdot) \in C_0$ . An extension to general mixed linear models with two variance components is shown in Subsection 2.4.

Section 3 is devoted to estimation of multivariate components of variance in an ECD model which corresponds to a one-way multivariate mixed linear model with equal replications. This issue did not receive a much attention primarily due to technical difficulties

encountered in obtaining dominance results. Amemiya (1985) proposed a restricted maximum likelihood (REML) estimator for the ‘between’ component but it is not known whether it is better than the usual unbiased estimator. Also see Calvin and Dykstra (1991) and Mathew, Niyogi and Sinha (1994). Recently Srivastava and Kubokawa (1999) succeeded in establishing a truncation rule for improving up on non-truncated procedures. This rule was applied to show the superiority of REML estimators to unbiased ones as well as to derive truncated estimators dominating Stein type minimax estimators. In Section 3, we extend these dominance results to the ECD model and investigate their robustness.

## 2 Estimation of Variance Components in a Random Effect Model

In this section, we treat an elliptically contoured distribution (ECD) model with covariance having the intraclass correlation structure, which is interpreted as an extension of a one-way random effect model in the normal distribution. In this model, we consider the problem of estimating the variance components from two aspects of decision theory and robustness.

### 2.1 A Random Effect Model and Its Canonical Form

Let us consider the following model with a covariance structure of intraclass correlation:

$$\mathbf{y} = \mu \mathbf{j}_N + \mathbf{u}, \quad (2.1)$$

where  $\mathbf{y}$  is an  $N \times 1$  observed vector,  $\mu \in \mathbf{R}$  is an unknown common mean,  $\mathbf{j}_N = (1, \dots, 1)^t \in \mathbf{R}^N$ , and  $\mathbf{u}$  is an  $N$ -variate error vector. Assume that the error has an elliptical density

$$|\Omega|^{-1/2} f(\mathbf{u}^t \Omega^{-1} \mathbf{u}), \quad (2.2)$$

where  $\Omega$  is a  $N \times N$  matrix with the intraclass correlation structure

$$\Omega = \sigma_1^2 \mathbf{I}_N + \sigma_a^2 \mathbf{J}_r \otimes \mathbf{I}_k \quad (2.3)$$

for  $\mathbf{J}_r = \mathbf{j}_r \mathbf{j}_r^t$ ,  $N = kr$  and  $\otimes$  designates the Kronecker product defined by  $\mathbf{A} \otimes \mathbf{B} = (a_{ij} \mathbf{B})$  for  $\mathbf{A} = (a_{ij})$ . In the case of a normal distribution, this corresponds to the one-way mixed linear model with two variance components:

$$\mathbf{u} = \mathbf{j}_r \otimes \boldsymbol{\alpha} + \boldsymbol{\epsilon}$$

where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\epsilon}$  are independent random variables with  $\boldsymbol{\alpha} \sim \mathcal{N}_k(\mathbf{0}, \sigma_a^2 \mathbf{I}_k)$  and  $\boldsymbol{\epsilon} \sim \mathcal{N}_N(\mathbf{0}, \sigma_1^2 \mathbf{I}_N)$ .

For providing a canonical form, consider  $N \times N$  orthogonal matrix  $\mathbf{H} = (\mathbf{H}_1^t, \mathbf{H}_2^t)^t$  for  $1 \times N$  vector  $\mathbf{H}_1 = N^{-1/2} \mathbf{j}_N^t$  and  $(N-1) \times N$  matrix  $\mathbf{H}_2$  such that  $\mathbf{H}_2 \mathbf{j}_N = \mathbf{0}$  and  $\mathbf{H}_2 \mathbf{H}_2^t = \mathbf{I}_{N-1}$ . Let  $x_0 = \mathbf{H}_1 \mathbf{y}$  and  $\mathbf{z} = \mathbf{H}_2 \mathbf{y}$ . The joint density of  $\mathbf{H} \mathbf{y} = (x_0, \mathbf{z}^t)^t$  is written by

$$|\Omega|^{-1/2} f\left(D(x_0) + \mathbf{z}^t (\sigma_1^2 \mathbf{I}_{N-1} + \sigma_a^2 \mathbf{H}_2 \{\mathbf{I}_r \otimes \mathbf{J}_k\} \mathbf{H}_2^t)^{-1} \mathbf{z}\right),$$

where  $D(x_0) = (x_0 - \mu \sqrt{N})^2 / (\sigma_1^2 + r \sigma_a^2)$ . It can be easily seen that  $\mathbf{H}_2 \{\mathbf{I}_r \otimes \mathbf{J}_k\} \mathbf{H}_2^t = r \mathbf{E}_2$  where  $\mathbf{E}_2$  is an  $(N-1) \times (N-1)$  idempotent matrix with  $\text{rank}(\mathbf{E}_2) = k-1$ . Letting  $\mathbf{E}_1 = \mathbf{I}_{N-1} - \mathbf{E}_2$  with  $\text{rank}(\mathbf{E}_1) = N-k$ , we see that

$$\sigma_1^2 \mathbf{I}_{N-1} + \sigma_a^2 \mathbf{H}_2 \{\mathbf{I}_r \otimes \mathbf{J}_k\} \mathbf{H}_2^t = \sigma_1^2 \mathbf{E}_1 + (\sigma_1^2 + r \sigma_a^2) \mathbf{E}_2.$$

Then the density of  $\mathbf{z}$  is expressed by

$$(\sigma_1^2)^{-n_1/2}(\sigma_2^2)^{-n_2/2}g\left(\sigma_1^{-2}\mathbf{z}^t\mathbf{E}_1\mathbf{z}+\sigma_2^{-2}\mathbf{z}^t\mathbf{E}_2\mathbf{z}\right),$$

where  $n_1 = N - k$ ,  $n_2 = k - 1$ ,  $\sigma_2^2 = \sigma_1^2 + r\sigma_a^2$  and

$$g(a) = \int \sigma_2^{-1}f(D(x_0) + a)dx_0. \quad (2.4)$$

Here we can take variables  $x_1, \dots, x_{n_1}$  such that  $\mathbf{z}^t\mathbf{E}_1\mathbf{z} = \sum_{i=1}^{n_1}x_i^2 = \|\mathbf{x}\|^2$  for  $\mathbf{x} = (x_1, \dots, x_{n_1})^t$ . Let us define  $S_1$  by  $S_1 = \mathbf{z}^t\mathbf{E}_1\mathbf{z}$ . We here make the transformation  $(x_1, \dots, x_{n_1}) \rightarrow (s_1, \theta_1, \dots, \theta_{n_1-1})$  such that

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_1-1} \\ x_{n_1} \end{pmatrix} = \sqrt{s_1} \cdot \begin{pmatrix} \cos \theta_1 \\ \sin \theta_1 \cos \theta_2 \\ \vdots \\ \sin \theta_1 \sin \theta_2 \cdots \cos \theta_{n_1-1} \\ \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{n_1-1} \end{pmatrix},$$

where  $0 \leq \theta_i \leq \pi$  for  $i = 1, \dots, n_1 - 2$ , and  $0 \leq \theta_{n_1-1} \leq 2\pi$ . Since the Jacobian is given by

$$\left| \frac{\partial(x_1, \dots, x_{n_1})}{\partial(s, \theta_1, \dots, \theta_{n_1-1})} \right| = 2^{-1} s_1^{(n_1-2)/2} \sin^{n_1-2} \theta_1 \sin^{n_1-3} \theta_2 \cdots \sin \theta_{n_1-2},$$

the joint density of  $(S_1, \theta_1, \dots, \theta_{n_1-1}, \mathbf{E}_2\mathbf{z})$  is expressed as

$$h(\theta_1, \dots, \theta_{n_1-1}) \times c_{n_1}(\sigma_1^2)^{-n_1/2}(\sigma_2^2)^{-n_2/2}s_1^{n_1/2-1}g\left(\sigma_1^{-2}s_1+\sigma_2^{-2}\mathbf{z}^t\mathbf{E}_2\mathbf{z}\right), \quad (2.5)$$

where

$$c_{n_1} = \frac{\pi^{n_1/2}}{\Gamma(n_1/2)} = \prod_{i=1}^{n_1-1} B\left(\frac{1}{2}, \frac{n_1-i}{2}\right), \quad (2.6)$$

and

$$h(\theta_1, \dots, \theta_{n_1-1}) = \frac{1}{2} \prod_{i=1}^{n_1-1} \frac{\sin^{n_1-i-1} \theta_i}{B(1/2, (n_1-i)/2)},$$

which is a uniform distribution on the hypersphere. Hence the joint density of  $(S_1, \mathbf{E}_2\mathbf{z})$  is given by

$$c_{n_1}(\sigma_1^2)^{-n_1/2}(\sigma_2^2)^{-n_2/2}s_1^{n_1/2-1}g\left(\sigma_1^{-2}s_1+\sigma_2^{-2}\mathbf{z}^t\mathbf{E}_2\mathbf{z}\right). \quad (2.7)$$

Let  $S_2 = \mathbf{z}^t\mathbf{E}_2\mathbf{z}$ , and the same arguments provide the joint density of  $(S_1, S_2)$  of the form

$$c_{n_1}c_{n_2}(\sigma_1^2)^{-n_1/2}(\sigma_2^2)^{-n_2/2}s_1^{n_1/2-1}s_2^{n_2/2-1}g\left(\sigma_1^{-2}s_1+\sigma_2^{-2}s_2\right). \quad (2.8)$$

Here it is noted that  $S_1$  and  $S_2$  are represented by

$$\begin{aligned} S_1 &= \mathbf{z}^t\mathbf{E}_1\mathbf{z} = \mathbf{y}^t\mathbf{H}_2^t\mathbf{E}_1\mathbf{H}_2\mathbf{y} \\ &= \mathbf{y}^t\left(\mathbf{I}_{kr} - \frac{1}{r}\mathbf{I}_k \otimes \mathbf{J}_r\right)\mathbf{y} \\ &= \sum_{i=1}^k \sum_{j=1}^r (y_{ij} - \bar{y}_{i\cdot})^2 \end{aligned} \quad (2.9)$$

and

$$\begin{aligned}
S_2 &= \mathbf{z}^t \mathbf{E}_2 \mathbf{z} = \mathbf{y}^t \mathbf{H}_2^t \mathbf{E}_2 \mathbf{H}_2 \mathbf{y} \\
&= \mathbf{y}^t \left( \frac{1}{r} \mathbf{I}_k \otimes \mathbf{J}_r - \frac{1}{rk} \mathbf{J}_{rk} \right) \mathbf{y} \\
&= r \sum_{i=1}^k (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2,
\end{aligned} \tag{2.10}$$

for  $\bar{y}_{i\cdot} = r^{-1} \sum_{j=1}^r y_{ij}$  and  $\bar{y}_{\cdot\cdot} = (rk)^{-1} \sum_{i=1}^k \sum_{j=1}^r y_{ij}$ . The estimation of the variance components is discussed in the following subsections on the basis of the density (2.8).

## 2.2 Estimation of the Variance Components

Now we address the problems of estimating  $\sigma_1^2$ ,  $\sigma_2^2$  and  $(\sigma_1^2, \sigma_2^2)$  based on the canonical form (2.8) of the ECD model (2.1). For each estimation problem, we provide generalized Bayes and empirical Bayes estimators improving upon ordinary procedures.

The estimation of the ‘within’ component of variance  $\sigma_1^2$  is first discussed with respect to the Stein loss  $L(\hat{\sigma}_1^2, \sigma_1^2)$ , where the function  $L(\cdot, \cdot)$  is defined by

$$L(a, b) = a/b - \log a/b - 1. \tag{2.11}$$

It is noted that the loss  $L(\hat{\sigma}_1^2, \sigma_1^2)$  can be derived from the Kullback-Leibler information loss in the normal or chi-squared distributions. The unbiased estimator based on  $S_1$  is given by

$$\hat{\sigma}_1^{2UB}(g) = n_1^{-1} S_1 (n_1 + n_2) A_{n_1+n_2}(g), \tag{2.12}$$

which is the best among multiples of  $S_1$  where

$$A_k(g) = \frac{\int_0^\infty y^{k/2-1} g(y) dy}{\int_0^\infty y^{k/2} g(y) dy}. \tag{2.13}$$

In fact, the coefficient of  $S_1$  in (2.12) is calculated from (2.8) as

$$\begin{aligned}
E[S_1] &= \sigma_1^2 \int \int c_{n_1} c_{n_2} s_1^{n_1/2} s_2^{n_2/2-1} g(s_1 + s_2) ds_1 ds_2 \\
&= \sigma_1^2 c_{n_1} c_{n_2} \int_0^1 z^{n_1/2} (1-z)^{n_2/2-1} dz \times \int_0^\infty y^{(n_1+n_2)/2} g(y) dy \\
&= \frac{n_1}{n_1 + n_2} \frac{1}{A_{n_1+n_2}(g)},
\end{aligned}$$

where the transformation  $(s_1, s_2) \rightarrow (yz, y(1-z))$  is made in the second equality, and the extreme equality follows from

$$(c_{n_1} c_{n_2})^{-1} = \int_0^1 z^{n_1/2-1} (1-z)^{n_2/2-1} dz \times \int_0^\infty y^{(n_1+n_2)/2-1} g(y) dy.$$

It is noted that the parameters  $\sigma_1^2$  and  $\sigma_2^2$  are restricted by the inequality  $\sigma_1^2 \leq \sigma_2^2$ , which implies that  $S_2$  contains the information up to  $\sigma_1^2$ . This information will be employed to improve upon the unbiased estimator  $\hat{\sigma}_1^{2UB}(g)$ . For the purpose, more generally, we consider the estimators of the form

$$\hat{\sigma}_1^2(\psi) = S_1 \psi(S_2/S_1) \tag{2.14}$$

and obtain the condition on  $\psi(\cdot)$  for  $\hat{\sigma}_1^2(\psi)$  dominating estimator  $a_1 S_1$ , a multiple of  $S_1$ .

**Proposition 2.1.** *Assume that*

- (a)  $\psi(w)$  is nondecreasing and  $\lim_{w \rightarrow \infty} \psi(w) = a_1$ ,
- (b)  $\psi(w) \geq \psi_0(w) A_{n_1+n_2}(g)$  where

$$\psi_0(w) = \frac{\int_0^w x^{n_2/2-1} (1+x)^{-(n_1+n_2)/2} dx}{\int_0^w x^{n_2/2-1} (1+x)^{-(n_1+n_2)/2-1} dx}. \quad (2.15)$$

Then  $S_1 \psi(S_2/S_1)$  is better than  $a_1 S_1$  uniformly for every  $\omega$  relative to the loss  $L(\hat{\sigma}_1^2, \sigma_1^2)$  for  $L(\cdot, \cdot)$  given by (2.11).

The proof is deferred to Section 2.5. From Proposition 2.1, we can get some specific improved estimators of  $\sigma_1^2$ , which will be given after stating the results corresponding to Proposition 2.1 for  $\sigma_2^2$ . For the estimation of  $\sigma_2^2$ , the unbiased estimator is given by

$$\hat{\sigma}_2^{2UB}(g) = n_2^{-1} S_2 (n_1 + n_2) A_{n_1+n_2}(g), \quad (2.16)$$

for  $A_{n_1+n_2}(g)$  defined by (2.13). For improving upon the unbiased estimator or the multiple of  $S_2$ , we consider estimators of the form

$$\hat{\sigma}_2^2(\phi) = S_2 \phi(S_1/S_2). \quad (2.17)$$

**Proposition 2.2.** *Assume that*

- (a)  $\phi(w)$  is nondecreasing and  $\lim_{w \rightarrow 0} \phi(w) = a_2$ ,
- (b)  $\phi(w) \leq \phi_0(w) A_{n_1+n_2}(g)$ , where

$$\phi_0(w) = \frac{\int_0^{1/w} x^{n_2/2-1} (1+x)^{-(n_1+n_2)/2} dx}{\int_0^{1/w} x^{n_2/2-1} (1+x)^{-(n_1+n_2)/2-1} dx}. \quad (2.18)$$

Then  $S_2 \phi(S_1/S_2)$  is better than  $a_2 S_2$  uniformly for every  $\omega$  relative to the loss  $L(\hat{\sigma}_2^2, \sigma_2^2)$  for  $L(\cdot, \cdot)$  given by (2.11).

The proof is given in Section 2.5. In Propositions 2.1 and 2.2, the difference between the normal and elliptically contoured distributions appears in  $A_{n_1+n_2}(g)$ . This means that the same techniques as used in the case of the normal distribution are helpful for providing specific forms of improved procedures. From Lemma 1 of Kubokawa *et al.* (1999), it follows that

- (i)  $\psi_0(w)$  is increasing,  $\lim_{w \rightarrow \infty} \psi_0(w) = (n_1 + n_2)/n_1$  and  $\psi_0(w) \leq 1 + n_2(n_2 + 2)^{-1}w$ ,
- (ii)  $\phi_0(w)$  is increasing,  $\lim_{w \rightarrow 0} \phi_0(w) = (n_1 + n_2)/n_2$  and  $\phi_0(w) \geq 1 + (n_2 + 2)n_2^{-1}w$ .

Letting  $\psi_1(w) = \min\{(n_1 + n_2)n_1^{-1}, 1 + w\} A_{n_1+n_2}(g)$  and  $\psi_2(w) = \min\{(n_1 + n_2)n_1^{-1}, 1 + (n_2 + 2)n_2^{-1}w\} A_{n_1+n_2}(g)$ , we see that  $\psi_0(w)$ ,  $\psi_1(w)$  and  $\psi_2(w)$  satisfy the conditions (a) and (b) of Proposition 2.1 for  $a_1 = n_1^{-1}(n_1 + n_2) A_{n_1+n_2}(g)$ . Then we obtain the improved estimators

$$\begin{aligned} \hat{\sigma}_1^{2GB}(g) &= S_1 \psi_0(S_2/S_1) A_{n_1+n_2}(g), \\ \hat{\sigma}_1^{2REML}(g) &= \min \left\{ \frac{S_1}{n_1}, \frac{S_1 + S_2}{n_1 + n_2} \right\} (n_1 + n_2) A_{n_1+n_2}(g), \\ \hat{\sigma}_1^{2EB}(g) &= \min \left\{ \frac{S_1}{n_1}, \frac{S_1 + n_2(n_2 + 2)^{-1} S_2}{n_1 + n_2} \right\} (n_1 + n_2) A_{n_1+n_2}(g). \end{aligned}$$

Similarly, combining Proposition 2.2 and the above arguments provides the estimators

$$\begin{aligned}\hat{\sigma}_2^{2GB}(g) &= S_2\psi_0(S_1/S_2)A_{n_1+n_2}(g), \\ \hat{\sigma}_2^{2REML}(g) &= \max\left\{\frac{S_2}{n_2}, \frac{S_1+S_2}{n_1+n_2}\right\}(n_1+n_2)A_{n_1+n_2}(g), \\ \hat{\sigma}_2^{2EB}(g) &= \max\left\{\frac{S_2}{n_2}, \frac{(n_2+2)n_2^{-1}S_1+S_2}{n_1+n_2}\right\}(n_1+n_2)A_{n_1+n_2}(g)\end{aligned}$$

having uniformly smaller risks than the unbiased one (2.16). The generalized Bayesness of  $\hat{\sigma}_1^{2GB}$  and  $\hat{\sigma}_2^{2GB}$  and the empirical Bayesness of  $\hat{\sigma}_1^{2REML}$ ,  $\hat{\sigma}_1^{2EB}$ ,  $\hat{\sigma}_2^{2REML}$  and  $\hat{\sigma}_2^{2EB}$  can be demonstrated in the end of this subsection. It is noted that  $\hat{\sigma}_1^{2REML}$  and  $\hat{\sigma}_2^{2REML}$  can be derived as the restricted maximum likelihood (REML) estimators in the case of the normal distribution.

We deal with the third issue of estimating the variance components  $(\sigma_1^2, \sigma_a^2)$  simultaneously relative to the intrinsic loss

$$L_s(\hat{\sigma}_1^2, \hat{\sigma}_a^2, \sigma_1^2, \sigma_a^2) = n_1L(\hat{\sigma}_1^2, \sigma_1^2) + n_2L(\hat{\sigma}_1^2 + r\hat{\sigma}_a^2, \sigma_1^2 + r\sigma_a^2), \quad (2.19)$$

where  $L(\cdot, \cdot)$  is defined in (2.11). The ANOVA (unbiased) estimator of  $(\sigma_1^2, \sigma_a^2)$  is given by  $(\hat{\sigma}_1^{2UB}, r^{-1}(\hat{\sigma}_2^{2UB} - \hat{\sigma}_1^{2UB}))$ . This form suggests to consider the estimators of the form  $(\hat{\sigma}_1^2(\psi), \hat{\sigma}_a^2(\phi, \psi))$  where  $\hat{\sigma}_1^2(\psi) = S_1\psi(S_2/S_1)$  and  $\hat{\sigma}_a^2(\phi, \psi) = r^{-1}\{S_2\phi(S_1/S_2) - S_1\psi(S_2/S_1)\}$ . Then the risk function of this estimator relative to the loss (2.19) is written by

$$R_s(\omega, \hat{\sigma}_1^2, \hat{\sigma}_a^2) = n_1E\left[L(S_1\psi(S_2/S_1), \sigma_1^2)\right] + n_2E\left[L(S_2\phi(S_1/S_2), \sigma_2^2)\right], \quad (2.20)$$

which implies that Propositions 2.1 and 2.2 can be directly applied to get specific estimators superior to  $(\hat{\sigma}_1^{2UB}, \hat{\sigma}_a^{2UB})$ . Especially for  $\sigma_a^2$ , we obtain the improved estimators:

$$\begin{aligned}\hat{\sigma}_a^{2GB}(g) &= r^{-1}\{S_2\phi_0(S_1/S_2) - S_1\psi_0(S_2/S_1)\}A_{n_1+n_2}(g), \\ \hat{\sigma}_a^{2REML}(g) &= r^{-1}\max\left\{\frac{S_2}{n_2} - \frac{S_1}{n_1}, 0\right\}(n_1+n_2)A_{n_1+n_2}(g), \\ \hat{\sigma}_a^{2EB}(g) &= r^{-1}\max\left\{\frac{n}{n_2}\frac{S_2}{S_1} - 1, \frac{2}{n_2}\right\} \\ &\quad \times \min\left\{\frac{S_1}{n_1}, \frac{S_1 + n_2(n_2+2)^{-1}S_2}{n_1+n_2}\right\}(n_1+n_2)A_{n_1+n_2}(g).\end{aligned}$$

It is noted that  $\hat{\sigma}_a^{2REML}(g)$  is nonnegative and that  $\hat{\sigma}_a^{2GB}(g)$  and  $\hat{\sigma}_a^{2EB}(g)$  are positive almost everywhere. Although the unbiased estimator  $\hat{\sigma}_a^{2UB}$  has a drawback of taking negative values with positive probability, the above improved estimators resolve this undesirable property.

We conclude this subsection with stating the Bayesian properties of the above given estimators.

Let  $\eta = 1/\sigma_1^2$  and  $\xi = \sigma_1^2/(\sigma_1^2 + r\sigma_a^2)$  and note that  $\xi$  is constrained by  $0 < \xi < 1$ . When prior distribution  $\pi(\eta, \xi)$  of  $(\eta, \xi)$  is supposed, the Bayes estimators under the corresponding



loss functions are generally given by  $\hat{\sigma}_1^{2B}$ ,  $\hat{\sigma}_2^{2B}$  and  $(\hat{\sigma}_1^{2B}, \hat{\sigma}_a^{2B})$  where

$$\begin{aligned}\hat{\sigma}_1^{2B} &= \frac{1}{E^\pi[\eta|S_1, S_2]}, \\ \hat{\sigma}_2^{2B} &= \frac{1}{E^\pi[\eta\xi|S_1, S_2]}, \\ \hat{\sigma}_a^{2B} &= \frac{1}{r} \left[ \frac{1}{E^\pi[\eta\xi|S_1, S_2]} - \hat{\sigma}_1^{2B} \right],\end{aligned}$$

where  $E^\pi[\cdot|S_1, S_2]$  stands for an expectation with respect to the posterior distribution  $\pi(\eta, \xi|S_1, S_2)$  of  $(\eta, \xi)$  given  $S_1$  and  $S_2$ .

For deriving the generalized Bayes estimators, suppose the improper prior distribution

$$\pi(\eta, \xi)d\eta d\xi = \eta^{-1}\xi^{-1}d\eta d\xi I(0 < \xi < 1)$$

for the indicator function  $I(\cdot)$ . Then the posterior density of  $(\eta, \xi)$  given  $S_1$  and  $S_2$  is proportional to

$$\xi^{n_2/2-1} \eta^{(n_1+n_2)/2-1} g((S_1 + \xi S_2)\eta) d\eta d\xi I(0 < \xi < 1).$$

By using this posterior distribution, it can be easily verified that the estimators  $\hat{\sigma}_1^{2GB}(g)$ ,  $\hat{\sigma}_2^{2GB}(g)$  and  $\hat{\sigma}_a^{2GB}(g)$  are generalized Bayes.

For deriving the empirical Bayes estimators, let  $d$  be a given positive constant less than or equal to one, and suppose the improper prior distribution  $\pi(\eta, \xi) = \eta^{-1}I(\xi = \xi_0)$  where  $P(\xi = \xi_0) = 1$  and  $\xi_0$  is an unknown constant such that  $0 < \xi_0 \leq d$ . Namely the supposed prior information about  $\xi$  is that  $\xi$  is unknown and in the interval  $(0, d]$ . Then the posterior density of  $\eta$  given  $S_1$  and  $S_2$ , and the marginal density of  $S_1$  and  $S_2$  are given by

$$\begin{aligned}(\text{posterior density}) &\propto \eta^{(n_1+n_2)/2-1} g((S_1 + \xi_0 S_2)\eta), \\ (\text{marginal density}) &\propto \xi_0^{n_2/2} [S_1 + \xi_0 S_2]^{-(n_1+n_2)/2} S_1^{n_1/2-1} S_2^{n_2/2-1} \\ &\quad \times \int_0^\infty y^{(n_1+n_2)/2-1} g(y) dy.\end{aligned}$$

Hence the Bayes estimators of  $\sigma_1^2$ ,  $\sigma_2^2$  and  $\sigma_a^2$  are

$$\begin{aligned}\hat{\sigma}_1^{2B}(\xi_0) &= (S_1 + \xi_0 S_2) A_{n_1+n_2}(g), \\ \hat{\sigma}_2^{2B}(\xi_0) &= (\xi_0)^{-1} (S_1 + \xi_0 S_2) A_{n_1+n_2}(g), \\ \hat{\sigma}_a^{2B}(\xi_0) &= \frac{1}{r} [\hat{\sigma}_2^{2B}(\xi_0) - \hat{\sigma}_1^{2B}(\xi_0)] \\ &= \frac{1 - \xi_0}{r \xi_0} (S_1 + \xi_0 S_2) A_{n_1+n_2}(g).\end{aligned}$$

Since  $\xi_0$  is unknown, it should be estimated from the marginal density. Noting that  $0 < \xi_0 < d$ , it is seen that the maximum likelihood estimator of  $\xi_0$  is written by  $\hat{\xi}_0^{ML} = \min\{n_2 S_1 / (n_1 S_2), d\}$ , which is substituted in the above Bayes estimators so as to obtain the empirical Bayes rules:

$$\begin{aligned}\hat{\sigma}_1^{2EB}(d) &= \hat{\sigma}_1^{2B}(\hat{\xi}_0^{ML}) \\ &= \min \left\{ \frac{S_1}{n_1}, \frac{S_1 + d S_2}{n_1 + n_2} \right\} (n_1 + n_2) A_{n_1+n_2}(g),\end{aligned}$$

$$\begin{aligned}\hat{\sigma}_2^{2EB}(d) &= \hat{\sigma}_2^{2B}(\hat{\xi}_0^{ML}) \\ &= \max \left\{ \frac{S_2}{n_2}, \frac{S_1/d + S_2}{n_1 + n_2} \right\} (n_1 + n_2) A_{n_1+n_2}(g)\end{aligned}$$

and

$$\begin{aligned}\hat{\sigma}_a^{2EB}(d) &= \hat{\sigma}_a^{2B}(\hat{\xi}_0^{ML}) \\ &= \frac{1}{r} \max \left\{ \frac{n_1}{n_2} \frac{S_2}{S_1} - 1, d^{-1} - 1 \right\} \\ &\quad \times \min \left\{ \frac{S_1}{n_1}, \frac{S_1 + dS_2}{n_1 + n_2} \right\} (n_1 + n_2) A_{n_1+n_2}(g).\end{aligned}$$

Especially putting  $d = 1$  yields the estimators  $\hat{\sigma}_1^{2REML}(g)$ ,  $\hat{\sigma}_2^{2REML}(g)$  and  $\hat{\sigma}_a^{2REML}(g)$ . Also putting  $d = n_2/(n_2 + 2)$  gives  $\hat{\sigma}_1^{2EB}(g)$ ,  $\hat{\sigma}_2^{2EB}(g)$  and  $\hat{\sigma}_a^{2EB}(g)$ .

### 2.3 Robustness of the Improvements

The results given in the previous subsection depend on the function  $g$  or  $f$  through  $A_{n_1+n_2}(g)$ . Especially for the normal distribution, the corresponding conclusions are provided by letting  $A_{n_1+n_2}(g) = (n_1 + n_2)^{-1}$ . For instance, the unbiased estimators of  $\sigma_1^2$ ,  $\sigma_2^2$  and  $\sigma_a^2$  are given by  $\hat{\sigma}_1^{2UB} = n_1^{-1}S_1$ ,  $\hat{\sigma}_2^{2UB} = n_2^{-1}S_2$  and  $\hat{\sigma}_a^{2UB} = r^{-1}(\hat{\sigma}_2^{2UB} - \hat{\sigma}_1^{2UB})$ , which are improved by several truncated and smooth procedures.

The objective in this subsection is to investigate whether these improvements are robust within a class of functions  $f$ . The conditions for the robust improvements on  $\hat{\sigma}_1^{2UB} = n_1^{-1}S_1$  and  $\hat{\sigma}_2^{2UB} = n_2^{-1}S_2$  can be derived based on Propositions 2.1 and 2.1. Let  $C$  be a class of functions on nonnegative real numbers.

**Proposition 2.3.** *Assume that*

- (a)  $\psi(w)$  is nondecreasing and  $\lim_{w \rightarrow \infty} \psi(w) = n_1^{-1}$ ,
- (b)  $\psi(w) \geq \psi_0(w) \sup_{g \in C} A_{n_1+n_2}(g)$ .

*Then  $S_1\psi(S_2/S_1)$  is better than  $n_1^{-1}S_1$  uniformly for every  $\omega$  and every  $g \in C$ .*

**Proposition 2.4.** *Assume that*

- (a)  $\phi(w)$  is nondecreasing and  $\lim_{w \rightarrow 0} \phi(w) = n_2^{-1}$ ,
- (b)  $\phi(w) \leq \phi_0(w) \inf_{g \in C} A_{n_1+n_2}(g)$ .

*Then  $S_2\phi(S_1/S_2)$  is better than  $n_2^{-1}S_2$  uniformly for every  $\omega$  and every  $g \in C$ .*

Propositions 2.3 and 2.4 imply that the robustness of improvement in estimation of  $\sigma_1^2$  has a quite different story from the case of estimation of  $\sigma_2^2$ : the robust improvements for  $\sigma_1^2$  hold in the case where  $\sup_{g \in C} A_{n_1+n_2}(g) \leq (n_1 + n_2)^{-1}$ , while those for  $\sigma_2^2$  hold when  $\inf_{g \in C} A_{n_1+n_2}(g) \geq (n_1 + n_2)^{-1}$ . This further means that the robustness of improvements in estimation of  $(\sigma_1^2, \sigma_a^2)$  are guaranteed in the only case where  $A_{n_1+n_2}(g) = (n_1 + n_2)^{-1}$  for all  $g$  in  $C$ .

The underlying distributions are hereafter focused on a class of normal-scale-mixed distributions:

$$f_H(\|\mathbf{x}\|^2) = \int_0^\infty (2\pi)^{-N/2} \eta^{N/2} e^{-(\eta/2)\|\mathbf{x}\|^2} dH(\eta) \quad (2.21)$$

where  $\mathbf{x} \in \mathbf{R}^N$  and  $H(\cdot)$  is a probability measure on nonnegative real numbers. Then  $g$  defined in (2.4) is expressed by

$$g_H(y) = \int_0^\infty (2\pi)^{(n_1+n_2)/2} \eta^{(n_1+n_2)/2} e^{-\eta y/2} dH(\eta), \quad (2.22)$$

where  $N - 1 = n_1 + n_2$ . Also  $A_{n_1+n_2}(g)$  defined in (2.13) is written as

$$\begin{aligned} A_{n_1+n_2}(g_H) &= \frac{\int_0^\infty y^{(n_1+n_2)/2-1} \int_0^\infty \eta^{(n_1+n_2)/2} e^{-\eta y/2} dH(\eta) dy}{\int_0^\infty y^{(n_1+n_2)/2} \int_0^\infty \eta^{(n_1+n_2)/2} e^{-\eta y/2} dH(\eta)} \\ &= (n_1 + n_2)^{-1} \cdot \frac{\int_0^\infty dH(\eta)}{\int_0^\infty \eta^{-1} dH(\eta)} \\ &= (n_1 + n_2)^{-1} \cdot \left[ \int_0^\infty \eta^{-1} dH(\eta) \right]^{-1}, \end{aligned}$$

which implies that

$A_{n_1+n_2}(g) \leq$  (resp.  $>$ )  $(n_1 + n_2)^{-1}$  if and only if  $\int_0^\infty \eta^{-1} dH(\eta) \geq$  (resp.  $<$ ) 1. The normal distribution corresponds to taking the probability as  $H(\{\eta = 1\}) = 1$ .

Let us specify the density  $h(\eta) = dH(\eta)/d\eta$  as

$$h(\eta) = \frac{1}{\Gamma(a/2)} \left(\frac{b}{2}\right)^{a/2} \eta^{a/2-1} e^{-b\eta/2} I(\eta > 0) \quad (2.23)$$

for  $a > 2$  and  $b > 0$ . Then

$$\int_0^\infty \eta^{-1} dH(\eta) = \frac{b}{a-2}.$$

We thus consider the following three classes.

(i) Putting  $a = \nu + 2$  and  $b = \nu$  in (2.23) for  $\nu > 0$  gives that  $\int_0^\infty \eta^{-1} dH(\eta) = 1$  and produces a class of densities of the form

$$C_0 = \left\{ f : f(\|\mathbf{x}\|^2) = \frac{\nu^{\nu/2+1} \Gamma((N+\nu)/2+1)}{\pi^{N/2} \Gamma(\nu/2+1)} (\nu + \|\mathbf{x}\|^2)^{-(N+\nu)/2+1}, \nu > 0 \right\}. \quad (2.24)$$

This includes the normal distribution as a limit of  $\nu$ . Since  $A_{n_1+n_2}(g) = (n_1 + n_2)^{-1}$  for all  $f \in C_0$ , all dominance results in the three estimation issues in the normal distribution remain true in  $C_0$ . Especially the robustness of improvements is guaranteed for the simultaneous estimation of  $(\sigma_1^2, \sigma_a^2)$ .

(ii) Putting  $a = b = \nu$  in (2.23) for  $\nu > 2$  gives that  $\int_0^\infty \eta^{-1} dH(\eta) = \nu/(\nu-2) > 1$  and produces a class of multivariate  $t$ -distributions, where the density is of the form

$$C_1 = \left\{ f : f(\|\mathbf{x}\|^2) = \frac{\nu^{\nu/2} \Gamma((N+\nu)/2)}{\pi^{N/2} \Gamma(\nu/2)} (\nu + \|\mathbf{x}\|^2)^{-(N+\nu)/2}, \nu > 0 \right\}. \quad (2.25)$$

The robust improvements still hold for  $\sigma_1^2$  while they are not guaranteed for  $\sigma_2^2$ . When we can impose the restriction on  $\nu$  as  $\nu \geq \nu_0$  for known value  $\nu_0 > 2$ , from Proposition 3.4, it follows that  $\hat{\sigma}_2^{2UB}$  is dominated by  $\max\{n_1^{-1} S_1, \nu_0^{-1} (\nu_0 - 2) (n_1 + n_2)^{-1} (S_1 + S_2)\}$  uniformly within the class  $C_1$  with  $\nu \geq \nu_0$ .

(iii) In the case that  $a = \nu + 4$  and  $b = \nu$  for  $\nu > 0$ , we observe that  $\int_0^\infty \eta^{-1} dH(\eta) = \nu/(\nu + 2) < 1$ , so that the robust improvements are guaranteed for  $\sigma_2^2$  while they never hold for  $\sigma_1^2$ .

As stated in (i), it is quite interesting to point out again that Propositions 2.1 and 2.2 with  $A_{n_1+n_2}(g) = (n_1 + n_2)^{-1}$  remain still established for all  $f$  in  $C_0$ , that is, the robustness of all dominance results in the normal distribution holds.

As one of other classes, we can consider contaminated normal distributions such that the error term  $\mathbf{u}$  in the model (2.1) has

$$(1 - \lambda)\mathcal{N}_N(\mathbf{0}, \mathbf{\Omega}) + \lambda\mathcal{N}_N(\mathbf{0}, \tau\mathbf{\Omega}) \quad (2.26)$$

where  $\lambda$  and  $\tau$  are unknown parameters satisfying  $0 < \lambda < 1$  and  $\tau > 1$ . This model means some data with a larger variance can be taken with probability  $\lambda$ . Since this distribution is derived by putting  $H(\{\eta = 1\}) = \lambda$  and  $H(\{\eta = \tau\}) = 1 - \lambda$ , we see that

$$\int_0^\infty \eta^{-1} dH(\eta) = \lambda + \tau^{-1}(1 - \lambda) < 1,$$

which implies the robustness of improvements in the estimation of  $\sigma_1^2$ .

## 2.4 An Extension to General Mixed Linear Models

We extend the results of the previous subsections to the following models with a more general covariance structure:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (2.27)$$

where  $\mathbf{y}$  is an  $N \times 1$  vector of observations,  $\mathbf{X}$  is an  $N \times p_1$  known matrix with  $\text{rank}(\mathbf{X}) = r$ ,  $\boldsymbol{\beta}$  is a  $p_1 \times 1$  vector of parameters and  $\mathbf{u}$  is an  $N$ -variate random error vector having the elliptical density (2.2) with

$$\mathbf{\Omega} = \sigma_1^2 \mathbf{I}_N + \sigma_a^2 \mathbf{A}\mathbf{A}^t \quad (2.28)$$

for  $N \times p_2$  known matrix  $\mathbf{A}$ . In the case of a normal distribution, this corresponds to the general mixed linear model with two variance components:

$$\mathbf{u} = \mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$$

where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\epsilon}$  are independent random variables with  $\boldsymbol{\alpha} \sim \mathcal{N}_{p_2}(\mathbf{0}, \sigma_a^2 \mathbf{I}_{p_2})$  and  $\boldsymbol{\epsilon} \sim \mathcal{N}_N(\mathbf{0}, \sigma_1^2 \mathbf{I}_N)$ . This includes a one-way random effects model with unbalanced replications and an error component model treated by in Fuller and Battese (1973), Battese *et al.*(1988) and Rao *et al.*(1993).

For giving a canonical form, consider  $N \times N$  orthogonal matrix  $\mathbf{H}^* = (\mathbf{H}_1^{*t}, \mathbf{H}_2^{*t})^t$  such that  $\mathbf{H}_2^* \mathbf{X} = \mathbf{0}$ ,  $\mathbf{H}_2^* \mathbf{H}_2^{*t} = \mathbf{I}_{N-r}$  and  $\mathbf{H}_2^*$  is an  $(N - r) \times N$  matrix. Let  $\mathbf{x}_0 = \mathbf{H}_1^* \mathbf{y}$  and  $\mathbf{z} = \mathbf{H}_2^* \mathbf{y}$ . Define  $\mathbf{\Omega}^*$  by

$$\begin{aligned} \mathbf{\Omega}^* &= \begin{pmatrix} \mathbf{\Omega}_{11}^* & \mathbf{\Omega}_{12}^* \\ \mathbf{\Omega}_{21}^* & \mathbf{\Omega}_{22}^* \end{pmatrix} \\ &= \sigma_1^2 \mathbf{I}_N + \sigma_a^2 \mathbf{H}^* \mathbf{A}\mathbf{A}^t \mathbf{H}^{*t}, \end{aligned}$$

for  $r \times r$  matrix  $\mathbf{\Omega}_{11}^*$ . Then the marginal density of  $\mathbf{z}$  is written by  $|\mathbf{\Omega}_{22}^*|^{-1/2} g(\mathbf{z}^t \mathbf{\Omega}_{22}^{*-1} \mathbf{z})$  where

$$g(a) = \int |\mathbf{\Omega}_{11,2}^*|^{-1/2} f(D(\mathbf{x}_0) + a) d\mathbf{x}_0 \quad (2.29)$$

for  $\Omega_{11.2}^* = \Omega_{11}^* - \Omega_{12}^* \Omega_{22}^{*-1} \Omega_{21}^*$  and

$$D(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{H}_1^* \mathbf{X} \boldsymbol{\beta} - \Omega_{12}^* \Omega_{22}^{*-1})^t \Omega_{11.2}^{*-1} (\mathbf{x}_0 - \mathbf{H}_1^* \mathbf{X} \boldsymbol{\beta} - \Omega_{12}^* \Omega_{22}^{*-1}).$$

Here we consider the spectral decomposition  $\mathbf{H}_2^* \mathbf{A} \mathbf{A}^t \mathbf{H}_2^{*t} = \sum_{i=1}^{\ell} \lambda_i \mathbf{E}_i$ , where  $\text{rank}(\mathbf{E}_i) = m_i$  and  $\sum_{i=1}^{\ell} m_i = \text{rank}(\mathbf{H}^* \mathbf{U} \mathbf{U}^t \mathbf{H}^{*t})$ . Assume that  $N - r - \sum_{i=1}^{\ell} m_i > 0$ , and let

$$n_1 = N - r - \sum_{i=1}^{\ell} m_i \quad \text{and} \quad n_2 = \sum_{i=1}^{\ell} m_i. \quad (2.30)$$

Let  $\mathbf{E}_0 = \mathbf{I}_{N-r} - \sum_{i=1}^{\ell} \mathbf{E}_i$ , whose rank is  $n_1$ . We thus get the quadratic statistics  $S_1 = \mathbf{z}^t \mathbf{E}_0 \mathbf{z}$  and  $T_i = \mathbf{z}^t \mathbf{E}_i \mathbf{z}$  for  $i = 1, \dots, \ell$ . The same argument as used above (2.5) give a joint density of  $(S_1, T_1, \dots, T_\ell)$  of the form

$$\begin{aligned} c_{n_1}(\sigma_1^2)^{-n_1/2} s_1^{n_1/2-1} \prod_{i=1}^{\ell} \left\{ c_{m_i}(\sigma_1^2 + \lambda_i \sigma_a^2)^{-m_i/2} t_i^{m_i/2-1} \right\} \\ \times g \left( \sigma_1^{-2} s_1 + \sum_{i=1}^{\ell} (\sigma_1^2 + \lambda_i \sigma_a^2)^{-1} t_i \right), \end{aligned} \quad (2.31)$$

where the constant  $c_{n_1}$  is defined by (2.6) and  $c_{m_i}$ 's are defined similarly.

Corresponding to the discussions in the previous subsections, we address three problems of estimating the variance components  $\sigma_1$ ,  $\sigma_2$  and  $(\sigma_1, \sigma_a)$  on the basis of  $(S_1, T_1, \dots, T_\ell)$  where  $\sigma_2 = \sigma_1 + \lambda \sigma_a$  for

$$\lambda = \sum_{i=1}^{\ell} \lambda_i m_i / \sum_{i=1}^{\ell} m_i = \sum_{i=1}^{\ell} \lambda_i m_i / n_2. \quad (2.32)$$

Let  $S_2 = \sum_{i=1}^{\ell} T_i$ . Then we can verify that all the results obtained in the previous subsections hold in the model (2.27) by replacing the  $n_1$ ,  $n_2$ ,  $S_1$ ,  $S_2$  and  $r$  in Subsection 2.1 with  $n_1$ ,  $n_2$ ,  $S_1$ ,  $S_2$  and  $\lambda$  in this subsection, respectively. Especially, the results corresponding to Propositions 2.1 and 2.2 are provided as follows:

**Proposition 2.5.** *Under the same conditions as in Proposition 2.1, the estimator  $S_1 \psi(S_2/S_1)$  of the form (2.14) dominates  $a_1 S_1$  relative to the loss  $L(\hat{\sigma}_1^2, \sigma_1^2)$  for  $L(a, b)$  given by (2.11).*

**Proposition 2.6.** *Under the same conditions as in Proposition 2.2, the estimator  $S_1 \phi(S_2/S_1)$  of the form (2.17) dominates  $a_1 S_1$  relative to the loss  $L(\hat{\sigma}_2^2, \sigma_2^2)$ .*

The proofs are given in the next subsection. For the simultaneous estimation of  $(\sigma_1^2, \sigma_a^2)$ , combining Propositions 2.5 and 2.6 and the same arguments as used around (2.20) provides a dominance result relative to the loss  $n_1 L(\hat{\sigma}_1^2, \sigma_1^2) + n_2 L(\hat{\sigma}_1^2 + \lambda \hat{\sigma}_a^2, \sigma_1^2 + \lambda \sigma_a^2)$ . Unbiased estimators and the corresponding improved estimators of  $\sigma_1^2$ ,  $\sigma_2^2$  and  $\sigma_a^2$  have the same forms as given in Subsection 2.2 with replacing  $r$  with  $\lambda$ .

Following the results in Subsection 2.3, we can get the robustness of the above dominance results within the class of densities  $C_0$ . The class  $C_1$  of multivariate  $t$ -distributions guarantees the robustness of the dominance results in estimation of  $\sigma_1^2$ .

## 2.5 Proofs

We here prove Propositions 2.5 and 2.6 in the general model (2.27) since Propositions 2.1 and 2.2 are included by Propositions 2.5 and 2.6.

The key tool for these proofs is the *Integral-Expression-of-Risk-Difference* (IERD) method given by Takeuchi (1991) and Kubokawa (1994, 99) in estimation of a scale parameter. The IERD method is, through the fundamental theorem of calculus, to give an integral-expression for a difference of risks of two estimators. For other instances in which the IERD method was applied, see Kubokawa (1998, 99), Kubokawa and Srivastava (1996) and Kubokawa *et al.* (1993a, 94).

**Proof of Propositions 2.1 and 2.5.** Since  $\lim_{w \rightarrow \infty} \psi(w) = a_1$ , from the IERD method of Kubokawa (1994a,b), we have

$$\begin{aligned} & R_1(\omega; a_1 S_1) - R_1(\omega; S_1 \psi\left(\frac{S_2}{S_1}\right)) \\ &= E \left[ \left\{ \frac{S_1}{\sigma_e^2} \psi\left(\frac{S_2}{S_1} t\right) - \log \frac{S_1}{\sigma_e^2} \psi\left(\frac{S_2}{S_1} t\right) - 1 \right\} \Big|_{t=1}^{\infty} \right] \\ &= E \left[ \int_1^{\infty} \frac{d}{dt} \left\{ \frac{S_1}{\sigma_e^2} \psi\left(\frac{S_2}{S_1} t\right) - \log \frac{S_1}{\sigma_e^2} \psi\left(\frac{S_2}{S_1} t\right) - 1 \right\} dt \right]. \end{aligned} \quad (2.33)$$

Let  $v = S_1/\sigma_e^2$ ,  $u_i = T_i/(\sigma_e^2 + \lambda_i \sigma_v^2)$ , and  $\theta_i = 1 + \lambda_i \sigma_v^2/\sigma_e^2$ , and denote the density functions of  $v$  and  $u_i$  by  $f$  and  $g_i$ , respectively. Carrying out the differentiation in (2.33) gives

$$\begin{aligned} & E \left[ \int_1^{\infty} \left\{ \frac{S_1}{\sigma_e^2} - \frac{1}{\psi(S_2 t/S_1)} \right\} \frac{S_2}{S_1} \psi'\left(\frac{S_2}{S_1} t\right) dt \right] \\ &= \int \cdots \int \int_1^{\infty} \left\{ v - \frac{1}{\psi(\sum \theta_i u_i v/t)} \right\} (\sum \theta_i u_i/v) \psi'(\sum \theta_i u_i t/v) dt \\ &\quad \times c v^{n_1/2-1} \left( \prod_i u_i^{m_i/2-1} \right) g\left(v + \sum_i u_i\right) dv \prod_i du_i, \end{aligned}$$

for  $c = c_{n_1} \prod_{i=1}^{\ell} c_{m_i}$ . Making the transformations  $(t/v)u_i = w_i$  and  $1/t = z$  in order, we observe that the r.h.s. of (2.33) is equal to

$$\begin{aligned} & \int \cdots \int \int_1^{\infty} \left\{ v - \frac{1}{\psi(\sum \theta_i w_i)} \right\} (\sum \theta_i w_i/t) \psi'(\sum \theta_i w_i t) \\ &\quad \times c \left( \prod_i w_i^{m_i/2-1} \right) v^{(n_1+n_2)/2-1} t^{-n_2/2} g((1 + \sum_i w_i/t)v) dt dv \prod_i dw_i \\ &= \int \cdots \int \left\{ v - \frac{1}{\psi(\sum \theta_i w_i)} \right\} (\sum \theta_i w_i) \psi'(\sum \theta_i w_i) \\ &\quad \times c \left( \prod_i w_i^{m_i/2-1} \right) v^{(n_1+n_2)/2-1} \int_0^1 z^{n_2/2-1} g((1 + \sum_i w_i z)v) dz dv \prod_i dw_i. \end{aligned} \quad (2.34)$$

Since  $\psi'(w) \geq 0$ , it is concluded that the r.h.s. of (2.34) is non-negative if

$$\begin{aligned} \psi(\sum \theta_i w_i) &\geq \frac{\int_0^{\infty} v^{(n_1+n_2)/2-1} \int_0^1 z^{n_2/2-1} g((1 + \sum_i w_i z)v) dz dv}{\int_0^{\infty} v^{(n_1+n_2)/2} \int_0^1 z^{n_2/2-1} g((1 + \sum_i w_i z)v) dz dv} \\ &= \frac{\int_0^{\sum w_i} x^{n_2/2-1} (1+x)^{-(n_1+n_2)/2} dx}{\int_0^{\sum w_i} x^{n_2/2-1} (1+x)^{-(n_1+n_2)/2-1} dx} \frac{\int_0^{\infty} y^{(n_1+n_2)/2-1} g(y) dy}{\int_0^{\infty} y^{(n_1+n_2)/2} g(y) dy} \\ &= \psi_0(\sum_i w_i) A_{n_1+n_2}(g). \end{aligned} \quad (2.35)$$

Since  $\theta_i \geq 1$  and  $\psi'(w) \geq 0$ , it follows that  $\psi(\Sigma_i \theta_i w_i) \geq \psi(\Sigma_i w_i)$ , which, from (2.35), gives the sufficient condition that  $\psi(\Sigma w_i)$  is greater than or equal to the r.h.s. of (2.35). This is guaranteed by the condition (b) of Propositions 2.1 and 2.5, which are established.  $\square$

**Proof of Propositions 2.2 and 2.6.** Since  $\lim_{w \rightarrow 0} \phi(w) = a_2$ , observe that

$$\begin{aligned} & R_2(\omega; a_2 S_2) - R_2(\omega; S_2 \phi\left(\frac{S_1}{S_2}\right)) \\ &= -E \left[ \int_0^1 \frac{d}{dt} \left\{ \frac{S_2}{\sigma_e^2 + \lambda \sigma_v^2} \phi\left(\frac{S_1}{S_2} t\right) - \log \frac{S_2}{\sigma_e^2 + \lambda \sigma_v^2} \phi\left(\frac{S_1}{S_2} t\right) - 1 \right\} dt \right] \\ &= \int \cdots \int \int_0^1 \left\{ \frac{1}{\phi(vt/\Sigma \theta_i u_i)} - \frac{\Sigma \theta_i u_i}{1 + \lambda \tau} \right\} \frac{v}{\Sigma \theta_i u_i} \phi'\left(\frac{vt}{\Sigma \theta_i u_i}\right) dt \\ &\quad \times c v^{n_1/2-1} \left( \prod_i u_i^{m_i/2-1} \right) g(v + \Sigma_i u_i) dv \Pi_i du_i \end{aligned} \tag{2.36}$$

for  $\tau = \sigma_v^2/\sigma_e^2$ . Making the transformations  $(t/\Sigma \theta_i u_i)v = w$  and  $w(1/t) = z$  in order, we can rewrite (2.36) as

$$\begin{aligned} & \int \cdots \int \int_0^1 \left\{ \frac{1}{\phi(w)} - \frac{\Sigma \theta_i u_i}{1 + \lambda \tau} \right\} \phi'(w) w \Sigma \theta_i u_i / t^2 \\ &\quad \times c \left( \Sigma_i \theta_i u_i w t^{-1} \right)^{n_1/2-1} \left( \prod_i u_i^{m_i/2-1} \right) g\left(\Sigma_i \theta_i u_i w t^{-1} + \Sigma_i u_i\right) dt \Pi_i du_i dw \\ &= \int \cdots \int \int_w^\infty \left\{ \frac{1}{\phi(w)} - \frac{\Sigma \theta_i u_i}{1 + \lambda \tau} \right\} \phi'(w) \Sigma \theta_i u_i \\ &\quad \times c \left( \Sigma_i \theta_i u_i z \right)^{n_1/2-1} \left( \prod_i u_i^{m_i/2-1} \right) g\left(\Sigma_i \theta_i u_i z + \Sigma_i u_i\right) dz \Pi_i du_i dw \end{aligned} \tag{2.37}$$

so that since  $\phi'(w) \geq 0$ , the l.h.s. of (2.36) is nonnegative if

$$\phi(w) \leq \frac{\int \cdot \int \int_w^\infty (\Sigma \theta_i u_i) (\Sigma \theta_i u_i z)^{n_1/2-1} (\prod_i u_i^{m_i/2-1}) g(\Sigma \theta_i u_i z + \Sigma u_i) dz \Pi_i du_i}{\int \cdot \int \int_w^\infty (\Sigma \theta_i u_i)^2 / (1 + \lambda \tau) (\Sigma \theta_i u_i z)^{n_1/2-1} (\prod_i u_i^{m_i/2-1}) g(\Sigma \theta_i u_i z + \Sigma u_i) dz \Pi_i du_i}. \tag{2.38}$$

Let  $s = \Sigma_{i=1}^\ell u_i$  and  $t_i = u_i/s$  for  $i = 1, \dots, \ell$ . The numerator of the r.h.s. of (2.38) is written as

$$\begin{aligned} & \int \cdot \int \int_w^\infty (\Sigma \theta_i u_i)^{n_1/2} z^{n_1/2-1} (\prod_i u_i^{m_i/2-1}) g(\Sigma \theta_i u_i z + \Sigma u_i) dz \Pi du_i \\ &= \int \cdot \int \int_w^\infty (\Sigma \theta_i t_i)^{n_1/2} z^{n_1/2-1} s^{(n_1+n_2)/2-1} g((\Sigma \theta_i t_i z + 1)s) dz ds \frac{\prod_i t_i^{m_i/2-1}}{t_1 \cdots t_{\ell-1}} \prod_{i=1}^{\ell-1} dt_i \\ &= \int \cdot \int \int_{\Sigma \theta_i t_i w}^\infty x^{n_1/2-1} s^{(n_1+n_2)/2-1} g((1+x)s) dx ds \frac{\prod_i t_i^{m_i/2-1}}{t_1 \cdots t_{\ell-1}} \prod_{i=1}^{\ell-1} dt_i \\ &= \int \cdot \int \int_{\Sigma \theta_i t_i w}^\infty x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2} dx \frac{\prod_i t_i^{m_i/2-1}}{t_1 \cdots t_{\ell-1}} \prod_{i=1}^{\ell-1} dt_i \int y^{(n_1+n_2)/2-1} g(y) dy. \end{aligned}$$

Similarly, we have that

$$\int \cdot \int \int_w^\infty (\Sigma \theta_i u_i)^{n_1/2+1} z^{n_1/2-1} (\prod_i u_i^{m_i/2-1}) g(\Sigma \theta_i u_i z + \Sigma u_i) dz \Pi du_i$$

$$\begin{aligned}
&= \int \cdot \int \int_{\Sigma \theta_i t_i w}^{\infty} (\Sigma \theta_i t_i) x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2-1} dx \frac{\prod t_i^{m_i/2-1}}{t_1 \cdots t_{\ell-1}} \prod_{i=1}^{\ell-1} dt_i \\
&\quad \times \int y^{(n_1+n_2)/2} g(y) dy.
\end{aligned}$$

Let  $Q = \Sigma_i \theta_i t_i$ , and the r.h.s. of (2.38) is expressed by

$$\frac{E^Q \left[ \int_{Q_w}^{\infty} x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2} dx \right]}{E^Q \left[ Q(1+\lambda\tau)^{-1} \int_{Q_w}^{\infty} x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2-1} dx \right]} A_{n_1+n_2}(g). \quad (2.39)$$

Since  $Q$  and  $\int_{Q_w}^{\infty} x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2-1} dx$  are monotone in the opposite directions, we can show the following inequality for the denominator of the r.h.s. of (2.39):

$$\begin{aligned}
&E^Q \left[ \frac{Q}{1+\lambda\tau} \int_{Q_w}^{\infty} x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2-1} dx \right] \\
&\leq E^Q \left[ \frac{Q}{1+\lambda\tau} \right] E^Q \left[ \int_{Q_w}^{\infty} x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2-1} dx \right]. \quad (2.40)
\end{aligned}$$

Here observe that

$$\begin{aligned}
E^Q \left[ \frac{Q}{1+\lambda\tau} \right] &= \frac{1}{1+\lambda\tau} + \frac{1}{1+\lambda\tau} \sum_{i=1}^{\ell} \lambda_i \tau E[z_i] \\
&= \frac{1}{1+\lambda\tau} + \frac{1}{1+\lambda\tau} \frac{\sum \lambda_i \tau m_i}{n_2} = 1,
\end{aligned} \quad (2.41)$$

since  $\lambda = \sum \lambda_i m_i / n_2$ . Combining (2.38), (2.39), (2.40) and (2.41) gives a sufficient condition as

$$\phi(w) \leq \frac{E^Q \left[ \int_{Q_w}^{\infty} x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2} dx \right]}{E^Q \left[ \int_{Q_w}^{\infty} x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2-1} dx \right]} A_{n_1+n_2}(g). \quad (2.42)$$

For the r.h.s. of (2.42), we can easily check that

$$\begin{aligned}
\frac{E^Q \left[ \int_{Q_w}^{\infty} x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2} dx \right]}{E^Q \left[ \int_{Q_w}^{\infty} x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2-1} dx \right]} &\geq \inf_{Q \geq 1} \left\{ \frac{\int_{Q_w}^{\infty} x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2} dx}{\int_{Q_w}^{\infty} x^{n_1/2-1} (1+x)^{-(n_1+n_2)/2-1} dx} \right\} \\
&= \inf_{Q \geq 1} \{ \phi_0(Qw) \},
\end{aligned}$$

which is equal to  $\phi_0(w)$  since  $\phi_0(\cdot)$  is nondecreasing. Hence we get the sufficient condition that  $\phi(w) \leq \phi_0(w) A_{n_1+n_2}(g)$ , which is just the condition (b), and the proof of Propositions 2.2 and 2.6 are complete.  $\square \square$

### 3 Estimation of Covariance Components in a Multivariate Random Effect Model

In this section, we provide a multivariate extension of some results of Section 2. An elliptically contoured distribution model which corresponds to a one-way multivariate random effects model is given and its canonical form is derived in Subsection 3.1. Dominance results for estimation of ‘within’ component of variance  $\Sigma_1$  are developed in Subsections 3.2 and 3.3, and their robustness is discussed in Subsection 3.4. Estimation of ‘between’ component of variance  $\Sigma_A$  and function  $\Sigma_2 = \Sigma_1 + r \Sigma_A$  is studied in Subsection 3.5.



### 3.1 A Multivariate Random Effect Model and Its Canonical Form

Let  $\mathbf{y}_{ij}$ 's be  $p$ -variate random variables for  $i = 1, \dots, k$  and  $j = 1, \dots, r$ , and let  $\mathbf{y} = (\mathbf{y}_1^t, \dots, \mathbf{y}_k^t)$  for  $\mathbf{y}_i = (\mathbf{y}_{i1}^t, \dots, \mathbf{y}_{ir}^t)^t$ . Suppose that  $\mathbf{y}$  has the following model with a covariance structure of the multivariate intraclass correlation:

$$\mathbf{y} = \mathbf{j}_N \otimes \boldsymbol{\mu} + \mathbf{u}, \quad (3.1)$$

where  $N = kr$ ,  $\boldsymbol{\mu} \in \mathbf{R}^p$  is an unknown vector and  $\mathbf{u}$  is a  $pN$ -variate error vector. Assume that the error has an elliptical density

$$|\boldsymbol{\Omega}|^{-1/2} f(\mathbf{u}^t \boldsymbol{\Omega}^{-1} \mathbf{u})$$

where  $\boldsymbol{\Omega}$  is a  $(pN) \times (pN)$  matrix with the multivariate intraclass correlation structure

$$\boldsymbol{\Omega} = \mathbf{I}_N \otimes \boldsymbol{\Sigma}_1 + \mathbf{I}_k \otimes \mathbf{J}_r \otimes \boldsymbol{\Sigma}_a$$

for  $p \times p$  positive definite matrices  $\boldsymbol{\Sigma}_1$  and  $\boldsymbol{\Sigma}_a$ . This structure corresponds to the one-way multivariate mixed linear model under the normality:

$$\mathbf{u} = (\mathbf{I}_k \otimes \mathbf{j}_r \otimes \mathbf{I}_p) \boldsymbol{\alpha} + \boldsymbol{\epsilon},$$

where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\epsilon}$  are independent random variables with  $\boldsymbol{\alpha} \sim \mathcal{N}_{pk}(\mathbf{0}, \mathbf{I}_k \otimes \boldsymbol{\Sigma}_a)$  and  $\boldsymbol{\epsilon} \sim \mathcal{N}_{pN}(\mathbf{0}, \mathbf{I}_N \otimes \boldsymbol{\Sigma}_1)$ . Namely,  $\mathbf{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}$  for  $i = 1, \dots, k$  and  $j = 1, \dots, r$ , where  $\boldsymbol{\alpha}_i \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_a)$  and  $\boldsymbol{\epsilon}_{ij} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_1)$ .

The  $N \times N$  matrix  $\mathbf{H}$  defined in Section 2.1 is employed to get a canonical form. Letting  $\mathbf{Y} = (\mathbf{y}_{11}, \dots, \mathbf{y}_{ir}; \dots; \mathbf{y}_{k1}, \dots, \mathbf{y}_{kr})$ , we see that

$$\begin{aligned} (\mathbf{H} \otimes \mathbf{I}_p) \mathbf{y} &= \text{vec}(\mathbf{Y} \mathbf{H}^t) \\ &= \text{vec}(\sqrt{N} \bar{\mathbf{y}}, \mathbf{Y} \mathbf{H}_2^t) \\ &= \begin{pmatrix} \sqrt{N} \bar{\mathbf{y}} \\ \text{vec}(\mathbf{Y} \mathbf{H}_2^t) \end{pmatrix}. \end{aligned}$$

Also we observe that

$$(\mathbf{H} \otimes \mathbf{I}_p)(\mathbf{j}_N \otimes \boldsymbol{\mu}) = (\sqrt{N}, 0, \dots, 0)^t \otimes \boldsymbol{\mu},$$

and that

$$\begin{aligned} &(\mathbf{H} \otimes \mathbf{I}_p) \boldsymbol{\Omega} (\mathbf{H}^t \otimes \mathbf{I}_p) \\ &= \mathbf{I}_N \otimes \boldsymbol{\Sigma}_1 + \mathbf{H}(\mathbf{I}_k \otimes \mathbf{J}_r) \mathbf{H}^t \otimes \boldsymbol{\Sigma}_a \\ &= \begin{pmatrix} \boldsymbol{\Sigma}_1 + r \boldsymbol{\Sigma}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N-1} \otimes \boldsymbol{\Sigma}_1 + \{\mathbf{H}_2(\mathbf{I}_k \otimes \mathbf{J}_r) \mathbf{H}_2^t\} \otimes \boldsymbol{\Sigma}_a \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\Sigma}_1 + r \boldsymbol{\Sigma}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_1 \otimes \boldsymbol{\Sigma}_1 + \mathbf{E}_2 \otimes (\boldsymbol{\Sigma} + r \boldsymbol{\Sigma}_a) \end{pmatrix}, \end{aligned}$$

where  $\mathbf{E}_1$  and  $\mathbf{E}_2$  are idempotent matrices defined in Section 2.1 such that  $\mathbf{E}_1 + \mathbf{E}_2 = \mathbf{I}_{N-1}$ . We thus write the quadratic form  $\mathbf{u}^t \boldsymbol{\Omega}^{-1} \mathbf{u}$  as

$$\begin{aligned} & (\mathbf{y} - \mathbf{j}_N \otimes \boldsymbol{\mu})^t \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{j}_N \otimes \boldsymbol{\mu}) \\ &= \begin{pmatrix} \sqrt{N} \bar{\mathbf{y}} - \sqrt{N} \boldsymbol{\mu} \\ \text{vec}(\mathbf{YH}_2^t) \end{pmatrix}^t \begin{pmatrix} \boldsymbol{\Sigma}_1 + r \boldsymbol{\Sigma}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_1 \otimes \boldsymbol{\Sigma}_1 + \mathbf{E}_2 \otimes (\boldsymbol{\Sigma} + r \boldsymbol{\Sigma}_a) \end{pmatrix}^{-1} \\ & \quad \times \begin{pmatrix} \sqrt{N} \bar{\mathbf{y}} - \sqrt{N} \boldsymbol{\mu} \\ \text{vec}(\mathbf{YH}_2^t) \end{pmatrix} \\ &= N(\bar{\mathbf{y}} - \boldsymbol{\mu})^t (\boldsymbol{\Sigma}_1 + r \boldsymbol{\Sigma}_a)^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu}) + \\ & \quad \{\text{vec}(\mathbf{YH}_2^t)\}^t \left\{ \mathbf{E}_1 \otimes \boldsymbol{\Sigma}_1^{-1} + \mathbf{E}_2 \otimes (\boldsymbol{\Sigma}_1 + r \boldsymbol{\Sigma}_a)^{-1} \right\} \text{vec}(\mathbf{YH}_2^t). \end{aligned}$$

By integrating out with respect to  $\sqrt{N} \bar{\mathbf{y}}$ , the density of  $\mathbf{YH}_2^t$  can be written by

$$|\boldsymbol{\Sigma}_1|^{-n_1/2} |\boldsymbol{\Sigma}_2|^{-n_2/2} g \left( \text{tr} \boldsymbol{\Sigma}_1^{-1} \mathbf{YH}_2^t \mathbf{E}_1 \mathbf{H}_2 \mathbf{Y}^t + \text{tr} \boldsymbol{\Sigma}_2^{-1} \mathbf{YH}_2^t \mathbf{E}_2 \mathbf{H}_2 \mathbf{Y}^t \right),$$

where  $n_1 = N - k$ ,  $n_2 = k - 1$ ,  $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_1 + r \boldsymbol{\Sigma}_a$ , and

$$g(a) = \int |\boldsymbol{\Sigma}_2|^{-1/2} f \left( N(\bar{\mathbf{y}} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}_2^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu}) + a \right) d(\sqrt{N} \bar{\mathbf{y}}).$$

Let  $\mathbf{S}_1 = \mathbf{YH}_2^t \mathbf{E}_1 \mathbf{H}_2 \mathbf{Y}^t$ . Since  $\mathbf{E}_1$  is an idempotent matrix with rank  $n_1$ , we can take  $p \times n_1$  matrix  $\mathbf{x}$  such that  $\mathbf{S}_1 = \mathbf{x} \mathbf{x}^t$ . The matrix  $\mathbf{x}$  can be written as  $\mathbf{x} = \mathbf{TK}$  where  $\mathbf{T}$  is a  $p \times p$  low-triangular matrix with positive diagonal elements and  $\mathbf{K}$  is a  $p \times n_1$  matrix with  $\mathbf{KK}^t = \mathbf{I}_p$ . The usual technique used in the theory of multivariate analysis gives that

$$d\mathbf{x}^t = 2^{-p} |\mathbf{S}_1|^{(n_1-p-1)/2} (d\mathbf{S}_1) \mathbf{K} d(\mathbf{K}^t)$$

and

$$\int_{V_{p,n_1}} \mathbf{K} d(\mathbf{K}^t) = 2^p \pi^{pn_1/2} / \Gamma_p(n_1/2),$$

where  $\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma(a - 2^{-1}(i-1))$  for  $a > (p-1)/2$ , and  $V_{p,n_1}$  designates the Stiefel manifold of  $n_1 \times p$  matrix with orthonormal columns (see Srivastava and Khatri (1979)). Hence the joint density of  $(\mathbf{S}_1, \mathbf{YH}_2^t \mathbf{E}_2)$  is expressed by

$$c_{p,n_1} |\boldsymbol{\Sigma}_1|^{-n_1/2} |\boldsymbol{\Sigma}_2|^{-n_2/2} |\mathbf{S}_1|^{(n_1-p-1)/2} g \left( \text{tr} \boldsymbol{\Sigma}_1^{-1} \mathbf{S}_1 + \text{tr} \boldsymbol{\Sigma}_2^{-1} \mathbf{YH}_2^t \mathbf{E}_2 \mathbf{H}_2 \mathbf{Y}^t \right),$$

where

$$c_{p,n_1} = \pi^{pn_1/2} / \Gamma_p(n_1/2). \quad (3.2)$$

Also let  $\mathbf{S}_2 = \mathbf{YH}_2^t \mathbf{E}_2 \mathbf{H}_2 \mathbf{Y}^t$ . Then the same arguments can be used to provide the joint density of  $(\mathbf{S}_1, \mathbf{S}_2)$  as

$$c_{p,n_1} c_{p,n_2} |\boldsymbol{\Sigma}_1|^{-n_1/2} |\boldsymbol{\Sigma}_2|^{-n_2/2} |\mathbf{S}_1|^{(n_1-p-1)/2} |\mathbf{S}_2|^{(n_2-p-1)/2} g \left( \text{tr} \boldsymbol{\Sigma}_1^{-1} \mathbf{S}_1 + \text{tr} \boldsymbol{\Sigma}_2^{-1} \mathbf{S}_2 \right). \quad (3.3)$$

Similar to (2.9) and (2.10),  $\mathbf{S}_1$  and  $\mathbf{S}_2$  are represented by

$$\mathbf{S}_1 = \mathbf{Y} \left( \mathbf{I}_{kr} - \frac{1}{r} \mathbf{I}_k \otimes \mathbf{J}_r \right) \mathbf{Y}^t$$

$$\begin{aligned}
&= \sum_{i=1}^k \sum_{j=1}^r (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i\cdot})(\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i\cdot})^t, \\
\mathbf{S}_2 &= \mathbf{Y} \left( \frac{1}{r} \mathbf{I}_k \otimes \mathbf{J}_r - \frac{1}{rk} \mathbf{J}_{rk} \right) \mathbf{Y}^t \\
&= r \sum_{i=1}^k (\bar{\mathbf{y}}_{i\cdot} - \bar{\mathbf{y}}_{\cdot\cdot})(\bar{\mathbf{y}}_{i\cdot} - \bar{\mathbf{y}}_{\cdot\cdot})^t,
\end{aligned}$$

where  $\bar{\mathbf{y}}_{i\cdot} = r^{-1} \sum_{j=1}^r \mathbf{y}_{ij}$  and  $\bar{\mathbf{y}}_{\cdot\cdot} = (rk)^{-1} \sum_{i=1}^k \sum_{j=1}^r \mathbf{y}_{ij}$ . It is noted that  $\boldsymbol{\Sigma}_1 \leq \boldsymbol{\Sigma}_2$  where  $\boldsymbol{\Sigma}_1 \leq \boldsymbol{\Sigma}_2$  denotes that  $\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1$  is nonnegative definite. Based on  $\mathbf{S}_1$ ,  $\mathbf{S}_2$  and their joint density (3.3), we shall address the problems of estimating  $\boldsymbol{\Sigma}_1$ ,  $\boldsymbol{\Sigma}_2$  and  $\boldsymbol{\Sigma}_a$  in the next section.

### 3.2 Improvement on an Unbiased Estimator of ‘Within’ Component of Variance

We now consider the problem of estimating  $\boldsymbol{\Sigma}_1$  relative to the Stein (or entropy) loss function

$$L(\widehat{\boldsymbol{\Sigma}}_1, \boldsymbol{\Sigma}_1) = \text{tr} \widehat{\boldsymbol{\Sigma}}_1 \boldsymbol{\Sigma}_1^{-1} - \log |\widehat{\boldsymbol{\Sigma}}_1 \boldsymbol{\Sigma}_1^{-1}| - p, \quad (3.4)$$

where every estimator  $\widehat{\boldsymbol{\Sigma}}_1$  is evaluated by the risk function  $R_1(\omega; \widehat{\boldsymbol{\Sigma}}_1) = E_\omega[L(\widehat{\boldsymbol{\Sigma}}_1, \boldsymbol{\Sigma}_1)]$  for  $\omega \in \Omega = \{(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) \mid \boldsymbol{\Sigma}_1 \leq \boldsymbol{\Sigma}_2\}$ .

We begin with deriving an unbiased estimator of  $\boldsymbol{\Sigma}_1$  based on  $\mathbf{S}_1$  only.

**Proposition 3.1.** For  $A_k(g)$  defined in (2.13), let

$$A_g^* = p(n_1 + n_2) A_{p(n_1+n_2)}(g). \quad (3.5)$$

The an unbiased estimator of  $\boldsymbol{\Sigma}_1$  is given by

$$\widehat{\boldsymbol{\Sigma}}_1^{UB}(g) = \frac{1}{n_1} \mathbf{S}_1 \cdot A_g^*, \quad (3.6)$$

which has the risk function

$$R_1(\omega; \widehat{\boldsymbol{\Sigma}}_1^{UB}(g)) = E \left[ p \log n_1 - \log |\mathbf{S}_1 A_g^*| \right]. \quad (3.7)$$

**Proof.** Let  $\mathbf{T}_1 = (t_{ij})$  be a lower triangular matrix with positive diagonal elements such that  $\boldsymbol{\Sigma}_1^{-1/2} \mathbf{S}_1 \boldsymbol{\Sigma}_1^{-1/2} = \mathbf{T}_1 \mathbf{T}_1^t$ . Since  $d\mathbf{S}_1 = 2^p \prod_{i=1}^p t_{ii}^{p+1-i} d\mathbf{T}_1$ , for  $c = c_{p,n_1} c_{p,n_2}$ , we have

$$\begin{aligned}
&\boldsymbol{\Sigma}_1^{-1/2} E[\mathbf{S}_1] \boldsymbol{\Sigma}_1^{-1/2} \\
&= \int \int c \mathbf{S}_1 |\mathbf{S}_1|^{(n_1-p-1)/2} |\mathbf{S}_2|^{(n_2-p-1)/2} g(\text{tr} \mathbf{S}_1 + \text{tr} \mathbf{S}_2) d\mathbf{S}_1 d\mathbf{S}_2 \\
&= 2c \int \int \mathbf{T}_1 \mathbf{T}_1^t t_{11}^{n_1-1} t_{22}^{n_1-2} \dots t_{pp}^{n_1-p} |\mathbf{S}_2|^{(n_2-p-1)/2} g \left( \sum_{i=1}^p (t_{i1}^2 + \dots + t_{ii}^2) + \text{tr} \mathbf{S}_2 \right) d\mathbf{T}_1 d\mathbf{S}_2 \\
&= \frac{\int \cdot \int \text{diag}(x_1, \dots, x_p) \prod_{i=1}^p x_i^{n_1/2-1} |\mathbf{S}_2|^{(n_2-p-1)/2} g \left( \sum_{j=1}^p x_j + \text{tr} \mathbf{S}_2 \right) \prod_i dx_i d\mathbf{S}_2}{\int \cdot \int \prod_{i=1}^p x_i^{n_1/2-1} |\mathbf{S}_2|^{(n_2-p-1)/2} g \left( \sum_{j=1}^p x_j + \text{tr} \mathbf{S}_2 \right) \prod_i dx_i d\mathbf{S}_2},
\end{aligned} \quad (3.8)$$

where the transformation  $x_i = t_{i1}^2 + \dots + t_{ii}^2$  are made. The same arguments are applied to the random matrix  $\mathbf{S}_2$ , and the r.h.s. of (3.8) is rewritten by

$$\frac{\int \cdot \int \text{diag}(x_1, \dots, x_p) \prod_{i=1}^p x_i^{n_1/2-1} y^{pn_2/2-1} g\left(\sum_{j=1}^p x_j + y\right) \prod_i dx_i dy}{\int \cdot \int \prod_{i=1}^p x_i^{n_1/2-1} y^{pn_2/2-1} g\left(\sum_{j=1}^p x_j + y\right) \prod_i dx_i dy}.$$

For  $j = 1, \dots, p$ , making the transformation  $z = \sum_{i \neq j} x_i + y$  again gives

$$\begin{aligned} & \frac{\int \cdot \int x_j \prod_{i=1}^p x_i^{n_1/2-1} y^{pn_2/2-1} g\left(\sum_{j=1}^p x_j + y\right) \prod_i dx_i dy}{\int \cdot \int \prod_{i=1}^p x_i^{n_1/2-1} y^{pn_2/2-1} g\left(\sum_{j=1}^p x_j + y\right) \prod_i dx_i dy} \\ &= \frac{\int \cdot \int x_j^{n_1/2} z^{\{(p-1)n_1+pn_2\}/2-1} g(x_j + z) \prod_i dx_j dz}{\int \cdot \int x_j^{n_1/2-1} z^{\{(p-1)n_1+pn_2\}/2-1} g(x_j + z) \prod_i dx_j dz}. \end{aligned} \quad (3.9)$$

Making the transformations  $v = x_j + z$  and  $w = z/(x_j + z)$  with  $J((x_j, z) \rightarrow (v, w)) = v$ , the r.h.s. of (3.9) is expressed by

$$\begin{aligned} & \frac{\int \cdot \int w^{\{(p-1)n_1+pn_2\}/2-1} (1-w)^{n_1/2} v^{p(n_1+n_2)/2} g(v) \prod_i dw dz}{\int \cdot \int w^{\{(p-1)n_1+pn_2\}/2-1} (1-w)^{n_1/2-1} v^{p(n_1+n_2)/2-1} g(v) \prod_i dw dz} \\ &= n_1 \left\{ p(n_1 + n_2) A_{p(n_1+n_2)}(g) \right\}^{-1}, \end{aligned}$$

which yields the unbiased estimator (3.6). The risk function (3.7) can be easily derived.  $\square$

We now provide a general truncation rule for improving upon estimators based on  $\mathbf{S}_1$  like  $\widehat{\Sigma}_1^{UB}(g)$  by employing the information on the order restriction that  $\Sigma_1 \leq \Sigma_2$ . The estimators which we treat are of the general form

$$\widehat{\Sigma}_1(\Psi, g) = \mathbf{S}_2^{1/2} \mathbf{P} \Psi(\mathbf{A}) \mathbf{P}^t \mathbf{S}_2^{1/2} \mathbf{A}_g^* \quad (3.10)$$

for

$$\Psi(\mathbf{A}) = \text{diag}(\psi_1(\mathbf{A}), \dots, \psi_p(\mathbf{A})),$$

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

$$\mathbf{P}^t \mathbf{S}_2^{-1/2} \mathbf{S}_1 \mathbf{S}_2^{-1/2} \mathbf{P} = \mathbf{A} = \text{diag}(\lambda_1, \dots, \lambda_p)$$

with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ , and  $\psi_i(\mathbf{A})$ 's are non-negative functions of  $\mathbf{A}$ . For given estimator  $\widehat{\Sigma}_1(\Psi, g)$ , we define a truncation rule  $[\Psi(\mathbf{A})]^{TR}$  by

$$\begin{aligned} [\Psi(\mathbf{A})]^{TR} &= \text{diag}(\psi_1^{TR}(\mathbf{A}), \dots, \psi_p^{TR}(\mathbf{A})), \\ \psi_i^{TR}(\mathbf{A}) &= \min \left\{ \psi_i(\mathbf{A}), \frac{\lambda_i + 1}{n_1 + n_2} \right\}, \quad i = 1, \dots, p, \end{aligned} \quad (3.11)$$

which gives the corresponding truncated estimator of the form

$$\widehat{\Sigma}_1([\Psi]^{TR}, g) = \mathbf{S}_2^{1/2} \mathbf{P} \text{diag}(\psi_1^{TR}(\mathbf{A}), \dots, \psi_p^{TR}(\mathbf{A})) \mathbf{P}^t \mathbf{S}_2^{1/2} \mathbf{A}_g^*. \quad (3.12)$$

Then we get the following general dominance result.

**Theorem 3.1.** *The estimator  $\widehat{\Sigma}_1([\Psi]^{TR}, g)$  dominates  $\widehat{\Sigma}_1(\Psi, g)$  relative to the Stein loss (3.4) if  $P \left[ [\Psi(\Lambda)]^{TR} \neq \Psi(\Lambda) \right] > 0$  at some  $\omega \in \Omega$ .*

**Proof.** Without any loss of generality, let  $\Sigma_1 = \mathbf{I}$  and  $\Sigma_2 = \Theta = \text{diag}(\theta_1, \dots, \theta_p)$  with  $\theta_1 \geq 1, \dots, \theta_p \geq 1$ . The joint density of  $\mathbf{S}_1$  and  $\mathbf{S}_2$  is

$$\text{const.} |\mathbf{S}_1|^{(n_1-p-1)/2} |\mathbf{S}_2|^{(n_2-p-1)/2} |\Theta|^{-n_2/2} g(\text{tr}(\mathbf{S}_1 + \Theta^{-1}\mathbf{S}_2)).$$

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

$$f_{\mathbf{F}, \mathbf{S}_2}(\mathbf{F}, \mathbf{S}_2) = \text{const.} |\mathbf{F}|^{(n_1-p-1)/2} |\mathbf{S}_2|^{(n_1+n_2-p-1)/2} |\Theta|^{-n_2/2} g(\text{tr}(\mathbf{F} + \Theta^{-1}\mathbf{S}_2)). \quad (3.13)$$

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

$$f_{\mathbf{A}, \mathbf{P}, \mathbf{S}_2}(\mathbf{A}, \mathbf{P}, \mathbf{S}_2) = \text{const.} f_p(\mathbf{P}) h(\mathbf{A}) |\mathbf{S}_2|^{(n_1+n_2-p-1)/2} |\Theta|^{-n_2/2} g((\mathbf{P}\mathbf{A}\mathbf{P}^t + \Theta^{-1})\mathbf{S}_2),$$

where  $f_p(\mathbf{P}) = J(\mathbf{P}^t d\mathbf{P} \rightarrow d\mathbf{P})$  and  $h(\mathbf{A})$  is a function of  $\mathbf{A}$  (see Srivastava and Khatri (1979, p.31-32)). Hence the conditional expectation of  $\mathbf{S}_2$  given  $(\mathbf{A}, \mathbf{P})$  is written by

$$\begin{aligned} & (\mathbf{P}\mathbf{A}\mathbf{P}^t + \Theta^{-1})^{1/2} E[\mathbf{S}_2 | \mathbf{A}, \mathbf{P}] (\mathbf{P}\mathbf{A}\mathbf{P}^t + \Theta^{-1})^{1/2} \\ &= \frac{\int \mathbf{S}_2 |\mathbf{S}_2|^{(n_1+n_2-p-1)/2} g(\text{tr} \mathbf{S}_2) d\mathbf{S}_2}{\int |\mathbf{S}_2|^{(n_1+n_2-p-1)/2} g(\text{tr} \mathbf{S}_2) d\mathbf{S}_2}. \end{aligned} \quad (3.14)$$

From the same arguments as in the proof of Proposition 3.1, we see that the r.h.s. of (3.14) is equal to

$$\mathbf{I}_p \left\{ p A_{p(n_1+n_2)}(g) \right\}^{-1},$$

so that

$$\begin{aligned} E[\mathbf{S}_2 | \mathbf{A}, \mathbf{P}] &= (\mathbf{P}\mathbf{A}\mathbf{P}^t + \Theta^{-1})^{-1} \left\{ p A_{p(n_1+n_2)}(g) \right\}^{-1} \\ &\geq (\mathbf{P}\mathbf{A}\mathbf{P}^t + \mathbf{I}_p)^{-1} \left\{ p A_{p(n_1+n_2)}(g) \right\}^{-1}. \end{aligned} \quad (3.15)$$

The difference of the risk functions of  $\widehat{\Sigma}_1(\Psi)$  and  $\widehat{\Sigma}_1([\Psi]^{TR})$  is written as

$$\begin{aligned} & R_1(\omega, \widehat{\Sigma}_1(\Psi, g)) - R_1(\omega, \widehat{\Sigma}_1([\Psi]^{TR}, g)) \\ &= E \left[ \text{tr}(\mathbf{P}\Psi(\Lambda)\mathbf{P}^t - \mathbf{P}[\Psi(\Lambda)]^{TR}\mathbf{P}^t) \mathbf{S}_2 \mathbf{A}_g^* - \log |\Psi(\Lambda)| \{ [\Psi(\Lambda)]^{TR} \}^{-1} \right] \\ &= E^{\mathbf{A}, \mathbf{P}} \left[ \text{tr}(\Psi(\Lambda) - [\Psi(\Lambda)]^{TR}) \mathbf{P}^t E[\mathbf{S}_2 | \mathbf{A}, \mathbf{P}] \mathbf{P} \mathbf{A}_g^* - \log |\Psi(\Lambda)| \{ [\Psi(\Lambda)]^{TR} \}^{-1} \right]. \end{aligned} \quad (3.16)$$

From (3.15) and the fact that  $\Psi(\Lambda) \geq [\Psi(\Lambda)]^{TR}$ , it follows that the r.h.s. in (3.16) is greater than or equal to

$$E \left[ \text{tr} \{ \Psi(\Lambda) - [\Psi(\Lambda)]^{TR} \} (n_1 + n_2) (\Lambda + \mathbf{I})^{-1} - \log |\Psi(\Lambda)| \{ [\Psi(\Lambda)]^{TR} \}^{-1} \right]$$

$$\begin{aligned}
&= \sum_{i=1}^p E \left[ \left\{ \left( \psi_i(\mathbf{A}) - \frac{\lambda_i + 1}{n_1 + n_2} \right) \frac{n_1 + n_2}{\lambda_i + 1} - \log \psi_i(\mathbf{A}) \frac{n_1 + n_2}{\lambda_i + 1} \right\} \right. \\
&\quad \left. \times I \left( \psi_i(\mathbf{A}) > \frac{\lambda_i + 1}{n_1 + n_2} \right) \right] \tag{3.17} \\
&= \sum_{i=1}^p E \left[ \left\{ \psi_i \frac{n_1 + n_2}{\lambda_i + 1} - \log \psi_i \frac{n_1 + n_2}{\lambda_i + 1} - 1 \right\} I \left( \psi_i(\mathbf{A}) > \frac{\lambda_i + 1}{n_1 + n_2} \right) \right] \\
&\geq 0,
\end{aligned}$$

which proves Theorem 3.1.  $\square$

Since  $\mathbf{S}_1 = \mathbf{S}_2^{1/2} \mathbf{S}_2^{-1/2} \mathbf{S}_1 \mathbf{S}_2^{-1/2} \mathbf{S}_2^{1/2} = \mathbf{S}_2^{1/2} \mathbf{P} \mathbf{A} \mathbf{P}' \mathbf{S}_2^{1/2}$ , the unbiased estimator  $\widehat{\Sigma}_1^{UB}(g)$  given by (3.6) can be expressed in the same manner as (3.10) by

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

where

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

The truncation rule (3.11) produces the estimator

$$\widehat{\Sigma}_1^{REML}(g) = \widehat{\Sigma}_1([\Psi^{UB}]^{TR}, g), \tag{3.18}$$

where

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

From Theorem 3.1, we get

**Corollary 3.1.** *The estimator  $\widehat{\Sigma}_1^{REML}(g)$  dominates the unbiased estimator  $\Sigma_1^{UB}(g)$  relative to the Stein loss (3.4).*

The estimator  $\widehat{\Sigma}_1^{REML}(g)$  is known to be the restricted (or residual) maximum likelihood (REML) estimator of  $\Sigma_1$  under the constraint  $\Sigma_1 \leq \Sigma_2$ .

It is interesting to note that the REML estimator  $\widehat{\Sigma}_1^{REML}(g)$  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  $(\boldsymbol{\eta}, \mathbf{S}_1, \mathbf{S}_2)$  has the form

$$\text{const.} |\boldsymbol{\eta}|^{(n_1+n_2-p-1)/2} |\mathbf{S}_1|^{(n_1-p-1)/2} |\mathbf{S}_2|^{(n_2-p-1)/2} |\boldsymbol{\xi}|^{n_1/2} g \left( \text{tr} \left( \boldsymbol{\xi}^{1/2} \mathbf{S}_1 \boldsymbol{\xi}^{1/2} + \mathbf{S}_2 \right) \boldsymbol{\eta} \right),$$

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

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

From (3.15), we thus get the Bayes estimator of  $\boldsymbol{\Sigma}_1$  under the Stein loss (3.4)

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

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 maximising  $|\boldsymbol{\beta}|^{n_2/2}|\mathbf{I} + \boldsymbol{\beta}|^{-(n_1+n_2)/2}$  subject to the order restriction  $\boldsymbol{\beta} \leq \mathbf{S}_1^{-1/2}\mathbf{S}_2\mathbf{S}_1^{-1/2}$  since  $\boldsymbol{\xi} \geq \mathbf{I}$ . The resulting MLE of  $\boldsymbol{\beta}$  is

$$\widehat{\boldsymbol{\beta}} = \mathbf{Q}\text{diag}\left(\min\left\{\frac{n_2}{n_1}, \frac{1}{\lambda_i}\right\}, i = 1, \dots, p\right)\mathbf{Q}^t$$

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

$$\mathbf{Q}^t\mathbf{S}_1^{-1/2}\mathbf{S}_2\mathbf{S}_1^{-1/2}\mathbf{Q} = \text{diag}\left(\lambda_1^{-1}, \dots, \lambda_p^{-1}\right).$$

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

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_1^B(\widehat{\boldsymbol{\xi}}, g) &= \frac{1}{n_1 + n_2}\mathbf{S}_1^{1/2}\mathbf{Q}\left\{\text{diag}\left[\min\left\{\frac{n_2}{n_1}, \frac{1}{\lambda_i}\right\}\right]_i + \mathbf{I}\right\}\mathbf{Q}^t\mathbf{S}_1^{1/2}A_g^* \\ &= \mathbf{S}_1^{1/2}\mathbf{Q}\text{diag}\left[\min\left\{\frac{1}{n_1}, \frac{\lambda_i + 1}{(n_1 + n_2)\lambda_i}\right\}\right]_i\mathbf{Q}^t\mathbf{S}_1^{1/2}A_g^*,\end{aligned}$$

where  $\text{diag}[a_i]_i = \text{diag}(a_1, \dots, a_p)$ . Here note that orthogonal matrices  $\mathbf{P}$  and  $\mathbf{Q}$  satisfy

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

Then we have that

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

Hence  $\widehat{\boldsymbol{\Sigma}}_1^B(\widehat{\boldsymbol{\xi}}, g)$  is rewritten as

$$\widehat{\boldsymbol{\Sigma}}_1^B(\widehat{\boldsymbol{\xi}}, g) = \mathbf{S}_2^{1/2}\mathbf{P}\text{diag}\left[\min\left\{\frac{\lambda_i}{n_1}, \frac{\lambda_i + 1}{n_1 + n_2}\right\}\right]_i\mathbf{P}^t\mathbf{S}_2^{1/2}A_g^*,$$

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

### 3.3 An Extension of Stein's Dominance Results

In estimation of a covariance matrix of a multivariate normal distribution, James and Stein (1961) established a surprising decision-theoretical result that the unbiased estimator  $n_1^{-1}\mathbf{S}_1$  is not admissible nor minimax. Let  $G_T^+$  be the triangular group consisting of lower

triangular matrices with positive diagonal elements. Let  $\mathbf{T}_1$  be a matrix in  $G_T^*$  with  $\mathbf{T}_1 \mathbf{T}_1^t = \mathbf{S}_1$ . Then their minimax estimator dominating  $n_1^{-1} \mathbf{S}_1$  is written in our setup as

$$\widehat{\boldsymbol{\Sigma}}_1^{JS}(g) = \mathbf{T}_1 \mathbf{D}_m \mathbf{T}_1^t A_g^*, \quad (3.19)$$

where  $A_g^*$  is given by (3.5) and  $\mathbf{D}_m$  is the diagonal matrix given by  $\mathbf{D}_m = \text{diag}(d_1, \dots, d_p)$  for

$$d_i = k_i^{-1} = (n_1 + p + 1 - 2i)^{-1}, \quad i = 1, \dots, p. \quad (3.20)$$

This dominance result is verified to remain true in the ECD model.

**Proposition 3.2.** *The estimator  $\widehat{\boldsymbol{\Sigma}}_1^{JS}(g)$  dominates the unbiased estimator  $\widehat{\boldsymbol{\Sigma}}_1^{UB}(g)$  relative to the Stein loss (3.4) and the risk function of  $\widehat{\boldsymbol{\Sigma}}_1^{JS}(g)$  is given by*

$$R_1(\omega, \widehat{\boldsymbol{\Sigma}}_1^{JS}(g)) = E \left[ -\log |\mathbf{D}_m| - \log |\mathbf{S}_1 A_g^*| \right]. \quad (3.21)$$

**Proof.** After making the transformations, we can set  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \mathbf{I}_p$  and write the risk function of  $\widehat{\boldsymbol{\Sigma}}_1^{JS}(g)$  as

$$\begin{aligned} R_1(\omega, \widehat{\boldsymbol{\Sigma}}_1^{JS}(g)) &= E \left[ L \left( \widehat{\boldsymbol{\Sigma}}_1^{JS}(g) \boldsymbol{\Sigma}_1^{-1} \right) \right] \\ &= E \left[ \text{tr} \mathbf{D}_m \mathbf{T}_1^t \mathbf{T}_1 A_g^* - \log |\mathbf{D}_m| - \log |\mathbf{S}_1 A_g^*| - p \right]. \end{aligned} \quad (3.22)$$

Note that  $\sum_{i=1}^p \sum_{j=1}^i t_{ij}^2 = \sum_{j=1}^p \sum_{i=j}^p t_{ij}^2$  for  $\mathbf{T}_1 = (t_{ij})$ . Let  $v_j = \sum_{i=j}^p t_{ij}^2$ . Then the same arguments as used in the proof of Proposition 3.1 can show that

$$\begin{aligned} E[\mathbf{T}_1^t \mathbf{T}_1] &= 2c \int \int \mathbf{T}_1^t \mathbf{T}_1 t_{11}^{n_1-1} \dots t_{pp}^{n_1-p} |\mathbf{S}_2|^{(n_2-p-1)/2} g \left( \sum_{i=1}^p \sum_{j=1}^i t_{ij}^2 + \text{tr} \mathbf{S}_2 \right) d\mathbf{T}_1 d\mathbf{S}_2 \\ &= \frac{\int \cdot \int \text{diag}(v_1, \dots, v_p) \prod_{j=1}^p v_j^{k_j/2-1} |\mathbf{S}_2|^{(n_2-p-1)/2} g \left( \sum_{j=1}^p v_j + \text{tr} \mathbf{S}_2 \right) \Pi_j dv_j d\mathbf{S}_2}{\int \cdot \int \prod_{j=1}^p v_j^{k_j/2-1} |\mathbf{S}_2|^{(n_2-p-1)/2} g \left( \sum_{j=1}^p v_j + \text{tr} \mathbf{S}_2 \right) \Pi_j dv_j d\mathbf{S}_2}, \end{aligned} \quad (3.23)$$

for  $k_i$  defined by (3.20). It can be also shown that for  $i = 1, \dots, p$ ,

$$\begin{aligned} &\frac{\int \cdot \int v_i \prod_{j=1}^p v_j^{k_j/2-1} |\mathbf{S}_2|^{(n_2-p-1)/2} g \left( \sum_{j=1}^p v_j + \text{tr} \mathbf{S}_2 \right) \Pi_j dv_j d\mathbf{S}_2}{\int \cdot \int \prod_{j=1}^p v_j^{k_j/2-1} |\mathbf{S}_2|^{(n_2-p-1)/2} g \left( \sum_{j=1}^p v_j + \text{tr} \mathbf{S}_2 \right) \Pi_j dv_j d\mathbf{S}_2} \\ &= (n_1 + p + 1 - 2i) \left\{ p(n_1 + n_2) A_{p(n_1+n_2)}(g) \right\}^{-1} \\ &= k_i (A_g^*)^{-1}. \end{aligned} \quad (3.24)$$

Combining (3.22), (3.23) and (3.24) gives the risk function (3.21). Comparing the risk functions (3.7) and (3.21), we can get Proposition 3.2 by noting the inequality that  $\sum_{i=1}^p \log(n_1 + p + 1 - 2i) \leq p \log n_1$ , which can be easily checked.  $\square \square$



It is known that the estimator  $\widehat{\Sigma}_1^{JS}(g)$  has a drawback that it depends on the coordinate system. Thus it will be desirable to construct orthogonally invariant estimators improving upon  $\widehat{\Sigma}_1^{JS}(g)$ . Stein (1977) and Dey and Srinivasan (1985) obtained an orthogonally invariant and improved estimator in a multivariate normal distribution. Their estimator is written in our situation as

$$\widehat{\Sigma}_1^S(g) = \widehat{\Sigma}_1(\Psi^S, g) = \mathbf{S}_2^{1/2} \mathbf{P} \Psi^S(\Lambda) \mathbf{P}^t \mathbf{S}_2^{1/2} A_g^* \quad (3.25)$$

where

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

On the other hand, Stein (1956), Eaton (1970) and Takemura (1984) gave an orthogonally invariant and improved estimator, which is written in our problem as

$$\widehat{\Sigma}_1^T(g) = \mathbf{S}_2^{1/2} \left\{ \int_{O(p)} \Gamma \mathbf{U} \Gamma \mathbf{D}_m \mathbf{U}^t \Gamma^t d\Gamma \right\} \mathbf{S}_2^{1/2} A_g^*, \quad (3.26)$$

where  $O(p)$  is the group of  $p \times p$  orthogonal matrices and  $\mathbf{U} \Gamma \in G_T^+$  with  $\mathbf{U} \Gamma \mathbf{U}^t \Gamma = \Gamma^t \mathbf{F} \Gamma$  for  $\mathbf{F} = \mathbf{S}_2^{-1/2} \mathbf{S}_1 \mathbf{S}_2^{-1/2} = \mathbf{P} \Lambda \mathbf{P}^t$ . Takemura (1984) provided another expression of the form

$$\widehat{\Sigma}_1^T(g) = \mathbf{S}_2^{1/2} \mathbf{P} \Psi^T(\Lambda) \mathbf{P}^t \mathbf{S}_2^{1/2} A_g^* \quad (3.27)$$

where

$$\begin{aligned} \Psi^T(\Lambda) &= \text{diag}(\psi_1^T(\Lambda), \dots, \psi_p^T(\Lambda)), \\ (\psi_1^T(\Lambda), \dots, \psi_p^T(\Lambda))^t &= \text{diag}(\lambda_1, \dots, \lambda_p) \mathbf{W}(\Lambda) (d_1, \dots, d_p)^t, \end{aligned}$$

for  $p \times p$  doubly stochastic matrix  $\mathbf{W}(\Lambda)$ . In fact this dominance result still holds in the ECD model.

**Proposition 3.3.** *The estimator  $\widehat{\Sigma}_1^{JS}(g)$  is further dominated by  $\widehat{\Sigma}_1^S(g)$  and  $\widehat{\Sigma}_1^T(g)$ , which has the risk functions given by*

$$R_1(\omega, \widehat{\Sigma}_1^S(g)) = E \left[ 2 \sum_{i>j} \frac{d_i - d_j}{\lambda_i - \lambda_j} \lambda_i - \log |\mathbf{D}_m| - \log |\mathbf{S}_1 A_g^*| \right], \quad (3.28)$$

$$R_1(\omega, \widehat{\Sigma}_1^T(g)) = E \left[ - \sum_{i=1}^p \log \left( \sum_{j=1}^p w_{ij}(\Lambda) d_j \right) - \log |\mathbf{S}_1 A_g^*| \right], \quad (3.29)$$

where  $\mathbf{W}(\Lambda) = (w_{ij}(\Lambda))$ .

**Proof.** We first treat estimators of the general form  $\widehat{\Sigma}_1^*(\Phi, g) = \mathbf{S}_2^{1/2} \Phi(\mathbf{F}) \mathbf{S}_2^{1/2} A_g^*$  for  $p \times p$  positive definite matrix  $\Phi(\mathbf{F})$ . Let  $\Sigma_1 = \mathbf{I}$  and  $\Sigma_2 = \Delta^{-1}$  without loss of generality. The risk function of  $\widehat{\Sigma}_1^*(\Phi, g)$  is expressed as

$$\begin{aligned} R_1(\omega, \widehat{\Sigma}_1^*(\Phi, g)) &= E \left[ \text{tr} \Phi(\mathbf{F}) E[\mathbf{S}_2 | \mathbf{F}] A_g^* - \log |\Phi(\mathbf{F})| - \log |\mathbf{S}_2 A_g^*| - p \right] \\ &= E \left[ \text{tr} \Phi(\mathbf{F}) (\mathbf{F} + \Delta)^{-1} (n_1 + n_2) - \log |\Phi(\mathbf{F})| - \log |\mathbf{S}_2 A_g^*| - p \right], \end{aligned} \quad (3.30)$$

where the second equality follows from (3.15). From (3.13), it is seen that  $\mathbf{F}$  has a multivariate  $F$ -distribution with the density

$$(\text{const.})|\mathbf{F}|^{(n_1-p-1)/2}/|\mathbf{F} + \mathbf{\Delta}|^{(n_1+n_2)/2}, \quad (3.31)$$

which does not depend on  $g$ . For evaluating the term  $E[\text{tr} \mathbf{\Phi}(\mathbf{F})(\mathbf{F} + \mathbf{\Delta})^{-1}]$ , we employ the  $F$ -identity of Muirhead and Verathaworn(1985):

$$E \left[ \text{tr}(\mathbf{F} + \mathbf{\Delta})^{-1} \mathbf{V} \right] = (n_1 + n_2)^{-1} E \left[ 2 \text{tr}(\mathbf{D}\mathbf{V}) + (n_1 - p - 1) \text{tr}(\mathbf{F}^{-1} \mathbf{V}) \right], \quad (3.32)$$

where  $\mathbf{V} = (v_{ij}(\mathbf{F}))$  is a  $p \times p$  matrix and  $\mathbf{D} = (d_{ij})$  is a  $p \times p$  matrix of partial derivative operators with  $d_{ij} = 2^{-1}(1 + \delta_{ij})\partial/\partial f_{ij}$  for  $\mathbf{F} = (f_{ij})$  and the Kronecker delta  $\delta_{ij}$ .

When we consider the special form  $\mathbf{\Phi}(\mathbf{F}) = \mathbf{P}\mathbf{\Psi}(\mathbf{\Lambda})\mathbf{P}^t$  for  $\mathbf{\Psi}(\mathbf{\Lambda}) = \text{diag}(\psi_1(\mathbf{\Lambda}), \dots, \psi_p(\mathbf{\Lambda}))$ , it follows from Stein(1975) and Bilodeau and Srivastava(1992) that

$$\mathbf{D}(\mathbf{P}\mathbf{\Psi}(\mathbf{\Lambda})\mathbf{P}^t) = \mathbf{P}\mathbf{\Psi}^{(1)}(\mathbf{\Lambda})\mathbf{P}^t, \quad (3.33)$$

in which

$$\begin{aligned} \mathbf{\Psi}^{(1)}(\mathbf{\Lambda}) &= \text{diag}(\psi_1^{(1)}(\mathbf{\Lambda}), \dots, \psi_p^{(1)}(\mathbf{\Lambda})), \\ \psi_i^{(1)}(\mathbf{\Lambda}) &= \frac{1}{2} \sum_{j=1, j \neq i}^p \frac{\psi_i(\mathbf{\Lambda}) - \psi_j(\mathbf{\Lambda})}{\lambda_i - \lambda_j} + \frac{\partial \psi_i(\mathbf{\Lambda})}{\partial \lambda_i}. \end{aligned}$$

Hence from (3.32) and (3.33), we get that

$$\begin{aligned} & E \left[ \text{tr}(\mathbf{F} + \mathbf{\Delta})^{-1} \mathbf{P}\mathbf{\Psi}^S(\mathbf{\Lambda})\mathbf{P}^t \right] (n_1 + n_2) \\ &= E \left[ 2 \sum_{i>j} \frac{d_i \lambda_i - d_j \lambda_j}{\lambda_i - \lambda_j} + (n_1 - p + 1) \sum_{i=1}^p d_i \right] \\ &= E \left[ 2 \sum_{i>j} \frac{d_i - d_j}{\lambda_i - \lambda_j} \lambda_i + \sum_{i=1}^p (n_1 + p + 1 - 2i) d_i \right] \\ &= E \left[ 2 \sum_{i>j} \frac{d_i - d_j}{\lambda_i - \lambda_j} \lambda_i + p \right], \end{aligned} \quad (3.34)$$

where the second equality can be verified by the equations

$$\frac{d_i \lambda_i - d_j \lambda_j}{\lambda_i - \lambda_j} = \frac{d_i - d_j}{\lambda_i - \lambda_j} \lambda_i + d_j$$

and  $\sum_{i>j} d_j = \sum_{i=1}^p \sum_{j=1}^{i-1} d_j = \sum_{j=1}^p \sum_{i=j+1}^p d_j = \sum_{j=1}^p (p-j)d_j$ . Combining (3.30) and (3.34) gives the risk function (3.28), which can be easily checked to be smaller than the risk of  $\widehat{\Sigma}_1^{JS}(g)$  given by (3.21), and the first part of Proposition 3.3 is proved.

For the second part, we first verify that the risk of  $\widehat{\Sigma}_1^{JS}(g)$  is equal to that of the estimator

$$\widetilde{\Sigma}_1^{JS}(g) = \mathbf{S}_2^{1/2} \mathbf{U} \mathbf{D}_m \mathbf{U}^t \mathbf{S}_2^{1/2} \mathbf{A}_g^*, \quad (3.35)$$

where  $\mathbf{U} \in G_T^+$  with  $\mathbf{U}\mathbf{U}^t = \mathbf{F}$ . In fact, from (3.30), the risk of  $\widetilde{\Sigma}_1^{JS}(g)$  is written by

$$R_1(\omega, \widetilde{\Sigma}_1^{JS}(g)) = E_{\Delta} \left[ \text{tr } \mathbf{D}_m \mathbf{U}^t (\mathbf{F} + \Delta)^{-1} \mathbf{U} (n_1 + n_2) - \log |\mathbf{D}_m| - \log |\mathbf{S}_1 \mathbf{A}_g^*| - p \right].$$

From Lemma 2 of Bilodeau and Srivastava (1992), we observe that

$$\begin{aligned} E_{\Delta} \left[ \text{tr } \mathbf{D}_m \mathbf{U}^t (\mathbf{F} + \Delta)^{-1} \mathbf{U} \right] &= E_{\mathbf{I}} \left[ \text{tr } \mathbf{D}_m \mathbf{U}^t (\mathbf{F} + \mathbf{I})^{-1} \mathbf{U} \right] \\ &= p(n_1 + n_2)^{-1}, \end{aligned} \quad (3.36)$$

which implies that  $R_1(\omega, \widehat{\Sigma}_1^{JS}(g)) \equiv R_1(\omega, \widetilde{\Sigma}_1^{JS}(g))$ . Hence from the convexity of the loss (3.4), it follows that  $\widehat{\Sigma}_1^T(g)$  dominates  $\widehat{\Sigma}_1^{JS}(g)$ . This dominance result is also verified by using the risk expression (3.29), which is derived as follows: From (3.36), we observe that

$$\begin{aligned} E \left[ \text{tr } \Sigma_1^{-1} \widehat{\Sigma}_1^T(g) \right] &= E_{\Delta} \left[ \text{tr } (\mathbf{F} + \Delta)^{-1} \int_{O(p)} \mathbf{\Gamma} \mathbf{U} \mathbf{\Gamma}^t \mathbf{D}_m \mathbf{U}^t \mathbf{\Gamma}^t d\mathbf{\Gamma} \right] (n_1 + n_2) \\ &= \int_{O(p)} E_{\Delta^*} \left[ \text{tr } (\mathbf{F} + \Delta^*)^{-1} \mathbf{U} \mathbf{D}_m \mathbf{U}^t \right] d\mathbf{\Gamma} (n_1 + n_2) \\ &= p, \end{aligned}$$

where  $\Delta^* = \mathbf{\Gamma}^t \Delta \mathbf{\Gamma}$ . From this equation, we get the risk expression (3.29). Since  $\mathbf{W}(\mathbf{A})$  is doubly stochastic, the concavity of  $\log(\cdot)$  implies that

$$\begin{aligned} - \sum_{i=1}^p \log \left( \sum_{j=1}^p w_{ij}(\mathbf{A}) d_j \right) &\leq - \sum_{j=1}^p \sum_{i=1}^p w_{ij}(\mathbf{A}) \log d_j \\ &= - \log |\mathbf{D}_m|, \end{aligned}$$

proving that  $\widehat{\Sigma}_1^T(g)$  dominates  $\widehat{\Sigma}_1^{JS}(g)$ . Therefore the proof of Proposition 3.3 is complete.  $\square$

Takemura (1984) gave exact expressions for  $\Psi^T(\mathbf{A})$  for  $p = 2$  and  $3$ . However, the explicit calculation of  $\mathbf{W}(\mathbf{A})$  for  $p > 3$  remains an intractable problem. Perron (1992) gave an approximation to  $\mathbf{W}(\mathbf{A})$ , say  $\widetilde{\mathbf{W}}(\mathbf{A})$ , with a doubly stochastic property, and showed the minimaxity of the approximated estimator in the normal distribution. In the multivariate  $F$ -distribution, a similar dominance result was obtained by Bilodeau and Srivastava (1992), where the matrix  $\widetilde{\mathbf{W}}(\mathbf{A})$  is given componentwise by

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

for

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

and

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

The same arguments as in Bilodeau and Srivastava (1992) demonstrates that  $\widehat{\Sigma}_1^{JS}(g)$  is improved on by

$$\widehat{\Sigma}_1^P(g) = \mathbf{S}_2^{1/2} \mathbf{P} \Psi^P(\mathbf{A}) \mathbf{P}^t \mathbf{S}_2^{1/2} A_g^*,$$

where

$$\begin{aligned} \Psi^P(\mathbf{A}) &= \text{diag}(\psi_1^P(\mathbf{A}), \dots, \psi_p^P(\mathbf{A})), \\ (\psi_1^P(\mathbf{A}), \dots, \psi_p^P(\mathbf{A}))^t &= \text{diag}(\lambda_1, \dots, \lambda_p) \widetilde{\mathbf{W}}(\mathbf{A})(d_1, \dots, d_p)^t. \end{aligned}$$

Now, applying the truncation rule (3.11) to the estimator  $\widehat{\Sigma}_1^S(g)$  yields the estimator

$$\widehat{\Sigma}_1^{STR}(g) = \widehat{\Sigma}_1([\Psi^S]^{TR}, g), \quad (3.37)$$

where

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

Similarly we obtain the truncated estimators  $\widehat{\Sigma}_1([\Psi^T]^{TR}, g)$  and  $\widehat{\Sigma}_1([\Psi^P]^{TR}, g)$ . From Theorem 2.1, we can get

**Corollary 3.2.** *The estimators  $\widehat{\Sigma}_1^{STR}(g)$ ,  $\widehat{\Sigma}_1([\Psi^T]^{TR}, g)$  and  $\widehat{\Sigma}_1([\Psi^P]^{TR}, g)$  dominate  $\widehat{\Sigma}_1^S(g)$ ,  $\widehat{\Sigma}_1^T(g)$  and  $\widehat{\Sigma}_1^P(g)$ , respectively, relative to the Stein loss (3.4).*

We conclude this subsection with demonstrating further improvement on  $\widehat{\Sigma}_1^{STR}(g)$ . When we consider estimators of the general form

$$\widehat{\Sigma}_1(\Psi) = \mathbf{S}_2^{1/2} \mathbf{P} \Psi(\mathbf{A}) \mathbf{P}^t \mathbf{S}_2^{1/2} A_g^*, \quad \Psi(\mathbf{A}) = \text{diag}(\psi_1(\mathbf{A}), \dots, \psi_p(\mathbf{A})),$$

it is quite natural to satisfy the condition

$$\psi_1(\mathbf{A}) \geq \psi_2(\mathbf{A}) \geq \dots \geq \psi_p(\mathbf{A}) \quad \text{for any } \mathbf{A},$$

which is called *order-preserving* in Sheena and Takemura (1992). The estimator  $\widehat{\Sigma}_1^{STR}$  do not satisfy the order-preserving condition.

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

$$\sum_{i=1}^j \psi_i^O \geq \sum_{i=1}^j \psi_i \quad \text{for } 1 \leq j \leq p-1 \quad \text{and} \quad \sum_{i=1}^p \psi_i^O = \sum_{i=1}^p \psi_i. \quad (3.38)$$

Let  $\widehat{\Sigma}_1(\Psi^O, g) = \mathbf{S}_2^{1/2} \mathbf{P} \Psi^O(\mathbf{A}) \mathbf{P}^t \mathbf{S}_2^{1/2} A_g^*$  for  $\Psi^O(\mathbf{A}) = \text{diag}(\psi_1^O(\mathbf{A}), \dots, \psi_p^O(\mathbf{A}))$ . Then we get

**Proposition 3.4.** *If  $P_\omega[\Psi(\mathbf{A}) \neq \Psi^O(\mathbf{A})] > 0$  for some  $\omega \in \Omega$ , then  $\widehat{\Sigma}_1(\Psi, g)$  is dominated by the order-preserving estimator  $\widehat{\Sigma}_1(\Psi^O, g)$  relative to the Stein loss (3.4).*

**Proof.** From (3.30), the risk difference can be written as

$$\begin{aligned} & R_1(\omega, \widehat{\Sigma}_1(\Psi^O, g)) - R_1(\omega, \widehat{\Sigma}_1(\Psi, g)) \\ &= (n_1 + n_2)E[\text{tr}(\Psi^O(\mathbf{A}) - \Psi(\mathbf{A}))\mathbf{B}^{-1}] \\ &= (n_1 + n_2)E^{\mathbf{A}} \left[ \sum_{i=1}^p (\psi_i^O(\mathbf{A}) - \psi_i(\mathbf{A}))E[\mathbf{B}^{ii}|\mathbf{A}] \right], \end{aligned} \quad (3.39)$$

where  $\mathbf{B} = \mathbf{A} + \mathbf{P}^t\Theta^{-1}\mathbf{P}$ ,  $\mathbf{B}^{ii}$  denotes the  $(i, i)$ -diagonal element of  $\mathbf{B}^{-1}$  and  $E[\cdot|\mathbf{A}]$  is a conditional expectation with respect to  $\mathbf{P}$  given  $\mathbf{A}$ . Following Sheena and Takemura (1992), we use the Abel's identity to get the equation

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

From Lemma 1 of Srivastava and Kubokawa(1999), we have the inequality

$$E[\mathbf{B}^{ii}|\mathbf{A}] \leq E[\mathbf{B}^{jj}|\mathbf{A}]$$

for  $i < j$ . This implies that the r.h.s. of (3.39) is negative, and Proposition 3.4 is proved.  $\square\square$

Applying Proposition 3.4 to  $\widehat{\Sigma}_1^{STR}(g)$ , we obtain the following order-preserving estimator improving on it, given by

$$\widehat{\Sigma}_1^{STRO}(g) = \mathbf{S}_2^{1/2} \mathbf{P} [[\Psi^S]^{TR}]^O \mathbf{P}^t \mathbf{S}_2^{1/2} A_g^*,$$

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

### 3.4 Robustness of the dominance results

We discuss here the robustness of the dominance results given in Subsections 3.2 and 3.3. The results of Theorem 3.1 and Corollaries 3.1 and 3.2 employ the truncation rule (3.11), so that the same arguments as in Subsection 2.3 can be applied to establish their robustness within the classes  $C_0$  and  $C_1$  given by (2.24) and (2.25), where  $N$  in (2.24) is replaced with  $pk_r$ .

The other type of the improvements includes Propositions 3.2, 3.3 and 3.4, which do not use the truncation rule. Since  $A_g^* = 1$  in the multivariate normal distribution, we denote the estimators  $\widehat{\Sigma}_1^{UB}(g)$ ,  $\widehat{\Sigma}_1^{JS}(g)$ ,  $\widehat{\Sigma}_1^S(g)$  and  $\widehat{\Sigma}_1^T(g)$  with  $A_g^* = 1$  by  $\widehat{\Sigma}_1^{UB}$ ,  $\widehat{\Sigma}_1^{JS}$ ,  $\widehat{\Sigma}_1^S$

and  $\widehat{\Sigma}_1^T$ , respectively. From Propositions 3.1, 3.2 and 3.3, their risk functions in the ECD model are provided by

$$\begin{aligned} R_1(\omega, \widehat{\Sigma}^{UB}) &= E \left[ p \left\{ (A_g^*)^{-1} - 1 \right\} + p \log n_1 - \log |\mathbf{S}_1| \right], \\ R_1(\omega, \widehat{\Sigma}^{JS}) &= E \left[ p \left\{ (A_g^*)^{-1} - 1 \right\} - \log |\mathbf{D}_m| - \log |\mathbf{S}_1| \right], \\ R_1(\omega, \widehat{\Sigma}^S) &= E \left[ p \left\{ (A_g^*)^{-1} - 1 \right\} + \frac{2}{A_g^*} \sum_{i>j} \frac{d_i - d_j}{\lambda_i - \lambda_j} \lambda_i - \log |\mathbf{D}_m| - \log |\mathbf{S}_1| \right], \\ R_1(\omega, \widehat{\Sigma}^T) &= E \left[ p \left\{ (A_g^*)^{-1} - 1 \right\} - \sum_{i=1}^p \log \left( \sum_{j=1}^p w_{ij}(\mathbf{A}) d_j \right) - \log |\mathbf{S}_1| \right]. \end{aligned}$$

These expressions imply that the ordering in domination of the estimators is preserved in the ECD model, that is, the robust improvements hold for every  $f(\cdot)$ . This is related to the results of Kubokawa and Srivastava (1999) who established the robustness through the extended Stein-Haff identity. Instead of the extended Stein-Haff identity, we utilized the multivariate  $F$ -identity in Proposition 3.3 since the two statistics  $\mathbf{S}_1$  and  $\mathbf{S}_2$  are available. Also it is seen that the improvement by the order-preserving rule is robust since Proposition 3.4 was shown based on the multivariate  $F$ -distribution.

### 3.5 Estimation of the ‘Between’ Multivariate Component of Variance

We here address the problem of estimating  $\Sigma_2 = \Sigma_1 + r \Sigma_A$  and the ‘between’ component  $\Sigma_A$  based on the canonical form (3.3) of the ECD model along the same line as in Subsection 2.2.

For the estimation of  $\Sigma_2$ , similar results as in the previous subsections can be derived by exchanging the indices 1 and 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  $\mathbf{Q}^t \mathbf{S}_1^{-1/2} \mathbf{S}_2 \mathbf{S}_1^{-1/2} \mathbf{Q} = \mathbf{A}^{-1} = \text{diag}(\lambda_1^{-1}, \dots, \lambda_p^{-1})$ , where  $\lambda_1^{-1} \leq \dots \leq \lambda_p^{-1}$ . The diagonal matrix  $\mathbf{A}$  is also defined in Subsection 3.2 as  $\mathbf{P}^t \mathbf{S}_2^{-1/2} \mathbf{S}_1 \mathbf{S}_2^{-1/2} \mathbf{P} = \mathbf{A}$ , so that we note that the following relation holds:

$$\mathbf{S}_2^{1/2} \mathbf{P} = \mathbf{S}_1^{1/2} \mathbf{Q} \mathbf{A}^{-1/2}. \quad (3.40)$$

$\Phi(\mathbf{A}) = \text{diag}(\phi_1(\mathbf{A}), \dots, \phi_p(\mathbf{A}))$  and consider estimators of the form

$$\widehat{\Sigma}_2(\Phi, g) = \mathbf{S}_1^{1/2} \mathbf{Q} \Phi(\mathbf{A}) \mathbf{Q}^t \mathbf{S}_1^{1/2} A_g^*, \quad (3.41)$$

which is, from (3.40), represented as

$$\widehat{\Sigma}_2(\Phi, g) = \mathbf{S}_2^{1/2} \mathbf{P} \mathbf{A}^{1/2} \Phi(\mathbf{A}) \mathbf{A}^{1/2} \mathbf{P}^t \mathbf{S}_2^{1/2} A_g^*. \quad (3.42)$$

When estimator  $\widehat{\Sigma}_2$  of  $\Sigma_2$  is evaluated in terms of the risk  $R_2(\omega, \widehat{\Sigma}_2) = E[L(\widehat{\Sigma}_2, \Sigma_2)]$  for the function  $L(\cdot, \cdot)$  given by (3.4), we can suppose that  $\Sigma_2 = \mathbf{I}$  and  $\Sigma_1 = \Theta^{-1} = \text{diag}(\theta_1^{-1}, \dots, \theta_p^{-1})$  for  $\theta_1^{-1} \leq 1, \dots, \theta_p^{-1} \leq 1$ . Therefore we can apply the results directly to get the improvements on  $\widehat{\Sigma}_2(\Phi, g)$ . The corresponding truncation rule is described as

$$\begin{aligned} \{\Phi(\mathbf{A})\}^{TR} &= \text{diag} \left( \phi_1^{TR}(\mathbf{A}), \dots, \phi_p^{TR}(\mathbf{A}) \right), \\ \phi_i^{TR}(\mathbf{A}) &= \max \left\{ \phi_i(\mathbf{A}), \frac{\lambda_i^{-1} + 1}{n_1 + n_2} \right\}, \quad i = 1, \dots, p, \end{aligned} \quad (3.43)$$

which yields the corresponding truncated estimator

$$\widehat{\Sigma}_2(\{\Phi\}^{TR}, g) = \mathbf{S}_1^{1/2} \mathbf{Q} \text{diag} \left( \phi_1^{TR}(\mathbf{A}), \dots, \phi_p^{TR}(\mathbf{A}) \right) \mathbf{Q}^t \mathbf{S}_1^{1/2} \mathbf{A}_g^*. \quad (3.44)$$

Similar to Theorem 2.1, we can verify that  $\widehat{\Sigma}_2(\{\Phi\}^{TR}, g)$  dominates  $\widehat{\Sigma}_2(\Phi, g)$  in terms of the risk  $R_2(\omega; \widehat{\Sigma}_2)$ .

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

$$\widehat{\Sigma}_2^{UB}(g) = n_2^{-1} \mathbf{S}_2 = \mathbf{S}_1^{1/2} \mathbf{Q} \Phi^{UB} \mathbf{Q}^t \mathbf{S}_1^{1/2} \mathbf{A}_g^*$$

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

$$\widehat{\Sigma}_2^{REML}(g) = \widehat{\Sigma}_2(\{\Phi^{UB}\}^{TR}, g), \quad (3.45)$$

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

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

Also the Stein type improved estimator corresponded to (3.25) for  $\Sigma_2$  is given by

$$\widehat{\Sigma}_2^S(g) = \widehat{\Sigma}_2(\Phi^S, g) = \mathbf{S}_1^{1/2} \mathbf{Q} \Phi^S(\mathbf{A}) \mathbf{Q}^t \mathbf{S}_1^{1/2} \mathbf{A}_g^*,$$

where

$$\Phi^S(\mathbf{A}) = \text{diag} \left( \frac{e_p}{\lambda_1}, \dots, \frac{e_1}{\lambda_p} \right),$$

for  $e_i = (n_2 + p + 1 - 2i)^{-1}$ . It should be noted that the order of  $e_1, \dots, e_p$  in  $\Phi^S(\mathbf{A})$  is reversed to the case of  $\Psi^S(\mathbf{A})$  in (3.25) because  $\lambda_p^{-1} \geq \dots \geq \lambda_1^{-1}$ . Applying the truncation rule yields

$$\widehat{\Sigma}_2^{STR}(g) = \widehat{\Sigma}_2(\{\Phi^S\}^{TR}, g), \quad (3.46)$$

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

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

We next consider the problem of estimating  $\Sigma_A$  through simultaneous estimation of  $(\Sigma_1, \Sigma_A)$  simultaneously relative to the loss

$$L_s(\widehat{\Sigma}_1, \widehat{\Sigma}_A; \Sigma_1, \Sigma_A) = n_1 L(\widehat{\Sigma}_1, \Sigma_1) + n_2 L(\widehat{\Sigma}_1 + r \widehat{\Sigma}_A, \Sigma_1 + r \Sigma_A) \quad (3.47)$$

for  $L(\cdot, \cdot)$  given by (3.4). When  $\Sigma_1$  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  $\Sigma_A$ . 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 (3.47) is written as

$$\begin{aligned} R_s(\omega; \widehat{\Sigma}_1, \widehat{\Sigma}_A) &= E_\omega \left[ L_s(\widehat{\Sigma}_1, \widehat{\Sigma}_A; \Sigma_1, \Sigma_A) \right] \\ &= n_1 R_1(\omega; \widehat{\Sigma}_1) + n_2 R_2(\omega; \widehat{\Sigma}_2), \end{aligned}$$

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

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

Hence we can obtain improved estimators of  $(\Sigma_1, \Sigma_A)$  by combining dominance results in estimation of  $\Sigma_1$  and  $\Sigma_A$ .

Combining  $\widehat{\Sigma}_1([\Psi]^{TR}, g)$  given by (3.12) and  $\widehat{\Sigma}_2(\{\Phi\}^{TR}, g)$  given by (3.44), and noting the expression (3.42), we get the estimator of  $\Sigma_A$  of the form

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

where

$$\begin{aligned} &\Lambda^{1/2} \{\Phi(\Lambda)\}^{TR} \Lambda^{1/2} - [\Psi(\Lambda)]^{TR} \\ &= \text{diag} \left[ \max \left\{ \phi_i(\Lambda) \lambda_i, \frac{\lambda_i + 1}{n_1 + n_2} \right\} - \min \left\{ \psi_i(\Lambda), \frac{\lambda_i + 1}{n_1 + n_2} \right\} \right]_i. \end{aligned} \quad (3.48)$$

In the case of combining the REML estimators  $\widehat{\Sigma}_1^{REML}$  and  $\widehat{\Sigma}_2^{REML}$ , the  $i$ -th diagonal element in (3.48) 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

$$\begin{aligned} \widehat{\Sigma}_A^{REML}(g) &= r^{-1} (\widehat{\Sigma}_2^{REML}(g) - \widehat{\Sigma}_1^{REML}(g)) \\ &= r^{-1} \mathbf{S}_2^{1/2} \mathbf{P} \text{diag} \left[ \max \left\{ \frac{1}{n_2} - \frac{\lambda_i}{n_1}, 0 \right\} \right]_i \mathbf{P}' \mathbf{S}_2^{1/2} A_g^*, \end{aligned}$$

which is n.n.d. This REML estimator of  $\Sigma_A$  is similar to the one proposed by Amemiya (1985). We thus get n.n.d. estimators  $(\widehat{\Sigma}_1^{REML}(g), \widehat{\Sigma}_A^{REML}(g))$  improving on  $(\widehat{\Sigma}_1^{UB}(g), \widehat{\Sigma}_A^{UB}(g))$  relative to the loss (3.47).

In the case of combining Stein type improved estimators  $\widehat{\Sigma}_1^{STR}(g)$  and  $\widehat{\Sigma}_2^{STR}(g)$ , the  $i$ -th diagonal element in (3.48) 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

$$\begin{aligned}\widehat{\Sigma}_A^{STR}(g) &= r^{-1} \left( \widehat{\Sigma}_2^{STR}(g) - \widehat{\Sigma}_1^{STR}(g) \right) \\ &= r^{-1} \mathbf{S}_2^{1/2} \mathbf{P} \text{diag} \left[ \max \left\{ \frac{1}{n_2 - (p+1-2i)} - \frac{\lambda_i}{n_1 + p + 1 - 2i}, 0 \right\} \right]_i \mathbf{P}^t \mathbf{S}_2^{1/2} A_g^*\end{aligned}$$

which is also n.n.d. In the sequel we get n.n.d. estimators  $(\widehat{\Sigma}_1^{STR}(g), \widehat{\Sigma}_A^{STR}(g))$  improving on  $(\widehat{\Sigma}_1^S(g), \widehat{\Sigma}_A^S(g))$  in terms of the risk  $R_s(\omega; \widehat{\Sigma}_1, \widehat{\Sigma}_A)$  where  $\widehat{\Sigma}_A^S(g) = r^{-1}(\widehat{\Sigma}_2^S(g) - \widehat{\Sigma}_1^S(g))$ . Comparing two n.n.d. estimators  $\widehat{\Sigma}_A^{REML}(g)$  and  $\widehat{\Sigma}_A^{STR}(g)$ , 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.

For the robustness of the improvements, we can get similar conclusions by combining Subsections 2.3 and 3.4. In particular, we recall that  $A_g^* \equiv 1$  for every  $f(\cdot)$  in  $C_0$ , and see that the dominance results given in this subsection are robust in  $C_0$ .

### Acknowledgments

The research of the first author was supported in part by a grant from the Shirakawa Institute of Animal Genetics and the Japan Livestock Technology Association, by the Ministry of Education, Japan, Grant No. 08780216, No. 09780214 and by a grant from the Research Institute for the Japanese Economy, University of Tokyo. The research of the second author was supported by Natural Science and Engineering Research Council of Canada.

## REFERENCES

- Amemiya, Y. (1985). What should be done when an estimated between-group covariance matrix is not nonnegative definite? *Amer. Statist.*, **39**, 112-117.
- Bilodeau, M. and Srivastava, M.S. (1992). Estimation of the eigenvalues of  $\Sigma_1 \Sigma_2^{-1}$ . *J. Multivariate Anal.*, **41**, 1-13.
- 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.
- Cellier, D., Fourdrinier, D. and Robert, C. (1989). Robust shrinkage estimators of the location parameter for elliptically symmetric distributions. *J. Multivariate Anal.*, **29**, 39-52.
- 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. Tech. Report. No. 49, Stanford University.
- 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.
- Klotz, J.H., Milton, R.C. and Zacks, S.(1969). Mean square efficiency of estimators of variance components. *J. Amer. Statist. Assoc.* **64**, 1383-1402.
- Kubokawa, T.(1994). A unified approach to improving equivariant estimators. *Ann. Statist.* **22**, 290-299.
- Kubokawa, T. (1995). Estimation of variance components in mixed linear models. *J. Multivariate Anal.*, **53**, 210-236.
- Kubokawa, T.(1998). The Stein phenomenon in simultaneous estimation: A review. *Applied Statistical Science III* (eds. S.E. Ahmed, M. Ahsanullah and B.K. Sinha), 143-173, NOVA Science Publishers, Inc., New York.
- Kubokawa, T. (1999). Shrinkage and modification techniques in estimation of variance and the related problems: A review. *Communication in Statistics - Theory and Methods* **28**, 613-650.
- Kubokawa, T., Saleh, A.K.Md.E. and Makita, S.(1993). On improved positive estimators of variance components. *Statistics Decisions*, Supplement Issue 3, 1-16.
- Kubokawa, T, Saleh, A.K.Md.E and Konno, Y. (1999). Bayes, Minimax and nonnegative estimators of variance components under Kullback-Leibler loss. *J. Statist. Plann. Inf.*, to appear.
- Kubokawa, T. and Srivastava, M.S. (1996). Double shrinkage estimators of ratio of variances. In *Multidimensional Statistical Analysis and Theory of Random Matrices* (Proceedings of the Sixth Eugene Lukacs Symposium, Bowling Green, OH, USA, 29-30 March 1996, Editors: A.K. Gupta and V.L. Girko), 139-154, VSP, Netherlands.
- Kubokawa, T. and Srivastava, M. (1999). Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution. *Ann. Statist.*, **27**, to appear.
- Kubokawa, T. and Srivastava, M. (1999). Robust improvement in estimation of a mean matrix in an elliptically contoured distribution. *J. Multivariate Anal.*, to appear.
- Mathew, T., Sinha, B.K. and Sutradhar, B.C.(1991). Nonnegative estimation of random effects variance components in balanced mixed models. *Proceedings of the International Symposium on Nonparametric Statistics and Related Topics*(ed. A.K.Md.E. Saleh).
- 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.
- Muirhead, R.J. and Verathaworn, T. (1985). On estimating the latent roots of  $\Sigma_1 \Sigma_2^{-1}$ . In *Multivariate Analysis VI* (P.R. Krishnaiah, Ed.), pp. 431-447. Academic Press, New York/London.

- 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.
- Portnoy, S.(1971). Formal Bayes estimation with application to a random effect model. *Ann. Math. Statist.* **42**, 1379-1402.
- 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.
- Srivastava, M. and Kubokawa, T. (1999). Improved nonnegative estimation of multivariate components of variance. *Ann. Statist.*, **27**, to appear.
- Stein, C.(1956). Some problems in multivariate analysis, Part I. Tech. Report. No. 6, Stanford University.
- Stein, C.(1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.*, **16**, 155-160.
- Stein, C. (1977). Lectures on multivariate estimation theory. In *Investigation on Statistical Estimation Theory I*, 4-65. Zapiski Nauchnykh Seminarov LOMI im. V.A. Steklova AN SSSR vol. 74, Leningrad (in Russian).
- Takeuchi, K.(1991). Personal Communications.
- Thompson, W.A., Jr. (1962). The problem of negative estimates of variance components. *Ann. Math. Statist.*, **33**, 273-289.