Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 97 (2006) 2242–2261
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Estimation of covariance matrices in fixed and mixed effects linear models

Tatsuya Kubokawa[a,*], Ming-Tien Tsai[b]

[a] *Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[b] *Institute of Statistical Science, Academia Sinica, 128, Academia Rd. Sec. 2, Taipei 115, Taiwan*

Received 27 January 2005
Available online 19 January 2006

Abstract

The estimation of the covariance matrix or the multivariate components of variance is considered in the multivariate linear regression models with effects being fixed or random. In this paper, we propose a new method to show that usual unbiased estimators are improved on by the truncated estimators. The method is based on the Stein–Haff identity, namely the integration by parts in the Wishart distribution, and it allows us to handle the general types of scale-equivariant estimators as well as the general fixed or mixed effects linear models.

AMS 2000 subject classification: Primary 62F11; 62J12; secondary 62C15; 62C20

Keywords: Covariance matrix; Decision theory; Estimation; Haff identity; Improvement; James–Stein estimator; Linear regression model; Minimaxity; Mixed effects model; Multivariate normal distribution; Stein identity; Variance component; Wishart distribution

1. Introduction

The problems of estimating the covariance matrix or the multivariate components of variance in the fixed or mixed effects linear models are addressed in a decision-theoretic framework. The dominance properties of truncated estimators over non-truncated and unbiased estimators have been studied in the literature. Most of the dominance results have been shown based on the conventional method of using conditional distributions given pivot or test statistics. However, it seems difficult to apply the method to general types of estimators in complicated models. To

* Corresponding author. Fax: +81 3 5841 5521.
E-mail addresses: tatsuya@e.u-tokyo.ac.jp (T. Kubokawa), mttsai@stat.sinica.edu.tw (M.-T. Tsai).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2005.11.004

show the dominance results in such a general setup, in this paper, we propose a new method based on the so-called Stein–Haff identity, namely the integration by parts in the Wishart distribution developed by Stein [21] and Haff [4].

The model we treat here is the multivariate linear regression model with effects being fixed or random, described by

$$\boldsymbol{y} = \boldsymbol{\beta}\boldsymbol{B} + \boldsymbol{\alpha}\boldsymbol{A} + \boldsymbol{\epsilon}, \tag{1.1}$$

where $\boldsymbol{y}$ is a $p \times N$ matrix of observations, $\boldsymbol{B}$ is a $q \times N$ matrix of explanatory variables, $\boldsymbol{\beta}$ is a $p \times q$ matrix of unknown coefficients, $\boldsymbol{A}$ is a $k \times N$ design matrix, $\boldsymbol{\alpha}$ is a $p \times k$ matrix of coefficients and $\boldsymbol{\epsilon}$ is a $p \times N$ matrix of random error variables. It is assumed that $\boldsymbol{\epsilon}$ has the multivariate normal distribution $\mathcal{N}_{p,N}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_N)$ for a $p \times p$ unknown positive definite matrix $\boldsymbol{\Sigma}_A$ and the $N \times N$ identity matrix $\boldsymbol{I}_N$, where we follow the notation of Srivastava and Khatri [17, p. 54, 76]. Two cases are considered for $\boldsymbol{\alpha}$: $\boldsymbol{\alpha}$ is assumed as a matrix of unknown parameters or as a random matrix. The model (1.1) is called a fixed effects model for $\boldsymbol{\alpha}$ unknown, and a mixed effects model for $\boldsymbol{\alpha}$ random.

Our primary interest is in the estimation of the covariance matrix $\boldsymbol{\Sigma}$ in the fixed effects linear model (1.1) in a decision-theoretic framework. Estimator $\widehat{\boldsymbol{\Sigma}}$ is evaluated in terms of the risk function $R(\boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \widehat{\boldsymbol{\Sigma}}) = E[L_{\mathrm{S}}(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})]$, where $L_{\mathrm{S}}(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$ is the Stein loss function defined by

$$L_{\mathrm{S}}(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \operatorname{tr} \widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - |\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p. \tag{1.2}$$

As demonstrated in Section 2, there exists a sufficient statistic $(\boldsymbol{S}, \boldsymbol{X})$ such that $\boldsymbol{S}$ is a $p \times p$ symmetric matrix having a Wishart distribution, $\boldsymbol{X}$ is a $p \times m$ matrix having a multivariate normal distribution with an unknown mean matrix, and $\boldsymbol{S}$ and $\boldsymbol{X}$ are mutually independently distributed. Ordinary estimators of $\boldsymbol{\Sigma}$ such as the maximum likelihood and unbiased estimators are constructed based on $\boldsymbol{S}$, and the statistic $\boldsymbol{X}$ is not employed for estimating $\boldsymbol{\Sigma}$. Sinha and Ghosh [16] showed that the unbiased estimator can be improved on by a truncated estimator by utilizing the information contained in $\boldsymbol{X}$. Perron [15] and Kubokawa et al. [11] provided some dominance results for $m = 1$. These are multivariate extensions of the well-known inadmissibility result of Stein [19]. For a good account and review in the univariate case, see Kubokawa [10]. Hara [6] showed that Sinha and Ghosh's estimator is further dominated by another truncated estimator, and Kubokawa and Srivastava [12] gave minimax and truncated estimators improving on the James–Stein estimator. These dominance results can be derived based on the method of using conditional distributions given test statistics. For example, Kubokawa and Srivastava [12] used a conditional distribution of $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{X}^{\mathrm{t}}$ given $\boldsymbol{W}^{-1/2}\boldsymbol{S}\boldsymbol{W}^{-1/2}$. This method in their paper requests not only that $\boldsymbol{W}$ has full rank, namely, $m \geqslant p$, but also that treated estimators are limited to a specific class of scale-equivariant estimators.

Instead of using the conventional method based on the conditional argument, in this paper, we propose a new method of using the Stein–Haff identity developed by Stein [21] and Haff [5]. In Section 2, we first handle the case of $m \geqslant p$ and use the new method to extend the results of Kubokawa and Srivastava [12] to the general class of scale-equivariant estimators, which satisfy $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^{\mathrm{t}}, \boldsymbol{A}\boldsymbol{W}\boldsymbol{A}^{\mathrm{t}}) = \boldsymbol{A}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{S}, \boldsymbol{W})\boldsymbol{A}^{\mathrm{t}}$ for any $p \times p$ nonsingular matrix $\boldsymbol{A}$. That is, the general scale-equivariant estimators are shown to be improved on by their truncated estimators under the Stein loss (1.2). From this dominance result, we derive several truncated estimators improving on the unbiased and/or the James–Stein minimax estimators. We next handle the case of $m < p$, which means that the matrix $\boldsymbol{X}\boldsymbol{X}^{\mathrm{t}}$ is singular. The method based on the Stein–Haff identity allows us not only to obtain scale-equivariant estimators dominating the unbiased estimator, but also to show that the scale-equivariant estimators are improved on by their truncated ones. All the proofs of the results in Section 2 are given in the Appendix. Numerical studies are also given in Section 2 to investigate the risk behaviors of the proposed estimators.

In Section 3, we treat the mixed linear model (1.1) with $\boldsymbol{\alpha}$ assumed to be a random variable having the distribution $\mathcal{N}_{p,k}(\mathbf{0}, \boldsymbol{\Sigma}_A, \boldsymbol{C})$ for a $p \times p$ positive definite unknown matrix $\boldsymbol{\Sigma}_A$ and a $k \times k$ positive definite known matrix $\boldsymbol{C}$. In the mixed effects linear model, the covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_A$ are referred to as the 'within' and between' multivariate components of variance, respectively, and we treat the estimation of the components $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_A$. Although the estimation of variance components in univariate mixed linear models have been considered extensively in the literature, the estimation of multivariate components of variance has not been well studied primarily due to technical difficulties. For some explanations on this topic, see Amemiya [1], Calvin and Dykstra [3], Mathew et al. [14] and Srivastava and Kubokawa [18]. In the estimation of the 'within' component of variance $\boldsymbol{\Sigma}$, in Section 3, we derive general types of truncated estimators improving on the unbiased and/or the James–Stein minimax estimators in the general mixed effects model (1.1). These dominance results can be derived from the results given in Section 2 by considering a conditional argument given $\boldsymbol{\alpha}_i$'s. In the estimation of the 'between' component of variance $\boldsymbol{\Sigma}_A$, nonnegative definite estimators improving on the usual non-truncated ones can be developed by using the Stein–Haff identity.

2. Estimation in fixed effects models

2.1. Canonical forms

To derive a canonical form of the fixed effects model (1.1) for unknown $\boldsymbol{\alpha}$, let us decompose $\boldsymbol{A}$ as $\boldsymbol{A} = (\mathbf{0}; \boldsymbol{A}^*)\boldsymbol{H}_A$ for a $k \times k$ nonsingular matrix $\boldsymbol{A}^*$ and an $N \times N$ orthogonal matrix $\boldsymbol{H}_A$. Let us decompose $\boldsymbol{y}\boldsymbol{H}_A^{\mathrm{t}}$ and $\boldsymbol{B}\boldsymbol{H}_A^{\mathrm{t}}$ as $\boldsymbol{y}\boldsymbol{H}_A^{\mathrm{t}} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$ and $\boldsymbol{B}\boldsymbol{H}_A^{\mathrm{t}} = (\boldsymbol{B}_1, \boldsymbol{B}_2)$ where $\boldsymbol{H}_A^{\mathrm{t}}$ is the transpose of $\boldsymbol{H}_A$, and $\boldsymbol{y}_1$ and $\boldsymbol{B}_1$ are, respectively, $p \times (N-k)$ and $q \times (N-k)$ matrices. Then the exponent in the joint density of $\boldsymbol{y}$ is proportional to

$$\begin{aligned}
&\operatorname{tr} \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\beta}\boldsymbol{B} - \boldsymbol{\alpha}\boldsymbol{A})(\boldsymbol{y} - \boldsymbol{\beta}\boldsymbol{B} - \boldsymbol{\alpha}\boldsymbol{A})^{\mathrm{t}} \\
&\quad = \operatorname{tr} \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_1 - \boldsymbol{\beta}\boldsymbol{B}_1)(\boldsymbol{y}_1 - \boldsymbol{\beta}\boldsymbol{B}_1)^{\mathrm{t}} \\
&\qquad + \operatorname{tr} \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\beta}\boldsymbol{B}_2 - \boldsymbol{\alpha}\boldsymbol{A}^*)(\boldsymbol{y}_2 - \boldsymbol{\beta}\boldsymbol{B}_2 - \boldsymbol{\alpha}\boldsymbol{A}^*)^{\mathrm{t}}.
\end{aligned}$$

The least squares estimator of $\boldsymbol{\beta}$ in terms of minimizing the quadratic form $\operatorname{tr} \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_1 - \boldsymbol{\beta}\boldsymbol{B}_1)(\boldsymbol{y}_1 - \boldsymbol{\beta}\boldsymbol{B}_1)^{\mathrm{t}}$ is given by

$$\widehat{\boldsymbol{\beta}}_1 = \boldsymbol{y}_1\boldsymbol{B}_1^{\mathrm{t}}(\boldsymbol{B}_1\boldsymbol{B}_1^{\mathrm{t}})^{-},$$

where $\boldsymbol{B}^{-}$ is a generalized inverse of a matrix $\boldsymbol{B}$. Using the equation $\boldsymbol{B}_1^{\mathrm{t}}(\boldsymbol{B}_1\boldsymbol{B}_1^{\mathrm{t}})^{-}\boldsymbol{B}_1\boldsymbol{B}_1^{\mathrm{t}} = \boldsymbol{B}_1^{\mathrm{t}}$, we see that the quadratic form $\operatorname{tr} \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_1 - \boldsymbol{\beta}\boldsymbol{B}_1)(\boldsymbol{y}_1 - \boldsymbol{\beta}\boldsymbol{B}_1)^{\mathrm{t}}$ is rewritten as

$$\operatorname{tr} \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_1 - \widehat{\boldsymbol{\beta}}_1\boldsymbol{B}_1)(\boldsymbol{y}_1 - \widehat{\boldsymbol{\beta}}_1\boldsymbol{B}_1)^{\mathrm{t}} + \operatorname{tr} \boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})\boldsymbol{B}_1\boldsymbol{B}_1^{\mathrm{t}}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})^{\mathrm{t}}$$

Then, the term $\operatorname{tr} \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\beta}\boldsymbol{B} - \boldsymbol{\alpha}\boldsymbol{A})(\boldsymbol{y} - \boldsymbol{\beta}\boldsymbol{B} - \boldsymbol{\alpha}\boldsymbol{A})^{\mathrm{t}}$ is decomposed as

$$\begin{aligned}
&\operatorname{tr} \boldsymbol{\Sigma}^{-1}\boldsymbol{S} + \operatorname{tr} \boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})\boldsymbol{B}_1\boldsymbol{B}_1^{\mathrm{t}}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})^{\mathrm{t}} \\
&\quad + \operatorname{tr} \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\beta}\boldsymbol{B}_2 - \boldsymbol{\alpha}\boldsymbol{A}^*)(\boldsymbol{y}_2 - \boldsymbol{\beta}\boldsymbol{B}_2 - \boldsymbol{\alpha}\boldsymbol{A}^*)^{\mathrm{t}},
\end{aligned} \tag{2.1}$$

where $\boldsymbol{S} = (\boldsymbol{y}_1 - \widehat{\boldsymbol{\beta}}_1\boldsymbol{B}_1)(\boldsymbol{y}_1 - \widehat{\boldsymbol{\beta}}_1\boldsymbol{B}_1)^{\mathrm{t}}$. Let q_1 denote the rank of $\boldsymbol{B}_1$, and it is noted that $q_1 \leqslant \min(q, N-k)$. Assumed that $q_1 < N - k$. Then we get the following canonical form: $\boldsymbol{S}$, $\widehat{\boldsymbol{\beta}}_1$ and $\boldsymbol{y}_2$

are mutually independently distributed as

$$\begin{aligned} \boldsymbol{S} &\sim \mathcal{W}_p(\boldsymbol{\Sigma}, n), \quad n = N - k - q_1, \\ \widehat{\boldsymbol{\beta}}_1 \boldsymbol{B}_1 &\sim \mathcal{N}_{p,N-k}(\boldsymbol{\beta}\boldsymbol{B}_1, \boldsymbol{\Sigma}, \boldsymbol{B}_1^{\mathrm{t}}(\boldsymbol{B}_1\boldsymbol{B}_1^{\mathrm{t}})^{-}\boldsymbol{B}_1), \\ \boldsymbol{y}_2 &\sim \mathcal{N}_{p,k}(\boldsymbol{\beta}\boldsymbol{B}_2 + \boldsymbol{\alpha}\boldsymbol{A}^*, \boldsymbol{\Sigma}, \boldsymbol{I}_k). \end{aligned} \tag{2.2}$$

The canonical form (2.2) is further simplified by decomposing $\boldsymbol{B}_1$ as $\boldsymbol{B}_1 = (\boldsymbol{B}^*; \boldsymbol{0})\boldsymbol{H}_B$, where $\boldsymbol{B}^*$ is a $p \times q_1$ matrix with full rank and $\boldsymbol{H}_B$ is an $(N-k) \times (N-k)$ orthogonal matrix. Letting $\boldsymbol{X} = (\widehat{\boldsymbol{\beta}}_1\boldsymbol{B}^*, \boldsymbol{y}_2)$, from (2.2), we have the following canonical form: $\boldsymbol{S}$ and $\boldsymbol{X}$ are mutually independently distributed as

$$\begin{aligned} &\boldsymbol{S} \sim \mathcal{W}_p(\boldsymbol{\Sigma}, n), \\ &\boldsymbol{X} \sim \mathcal{N}_{p,m}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{I}_m), \end{aligned} \tag{2.3}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}\boldsymbol{B}^*, \boldsymbol{\beta}\boldsymbol{B}_2 + \boldsymbol{\alpha}\boldsymbol{A}^*)$ and $m = q_1 + k$. The set of unknown parameters $(\boldsymbol{\Sigma}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ or $(\boldsymbol{\Sigma}, \boldsymbol{\theta})$ is denoted by ω through this paper.

2.2. *Dominance results*

We now consider the estimation of the covariance matrix $\boldsymbol{\Sigma}$ based on $\boldsymbol{S}$ and $\boldsymbol{X}$ in the canonical form (2.3) under the Stein loss function (1.2). One of well known estimators is the unbiased estimator, denoted by $\widehat{\boldsymbol{\Sigma}}^0 = n^{-1}\boldsymbol{S}$, with the risk $R(\omega, \widehat{\boldsymbol{\Sigma}}^0) = -E[\log|\boldsymbol{S}\boldsymbol{\Sigma}^{-1}|] + p\log n$ for $\omega = (\boldsymbol{\Sigma}, \boldsymbol{\theta})$. The non-minimaxity of $\widehat{\boldsymbol{\Sigma}}^0$ was established by James and Stein [8], who derived the minimax estimator of the form $\widehat{\boldsymbol{\Sigma}}^{\mathrm{JS}} = \boldsymbol{T}\boldsymbol{D}\boldsymbol{T}^{\mathrm{t}}$, where $\boldsymbol{S} = \boldsymbol{T}\boldsymbol{T}^{\mathrm{t}}$, $\boldsymbol{T}$ is a lower triangular matrix with positive diagonal elements (and hence unique), and $\boldsymbol{D} = \mathrm{diag}\,(d_1, \ldots, d_p)$ for $d_i = (n + p + 1 - 2i)^{-1}$. The risk of $\widehat{\boldsymbol{\Sigma}}^{\mathrm{JS}}$ is given by

$$R(\omega, \widehat{\boldsymbol{\Sigma}}^{\mathrm{JS}}) = -E[\log|\boldsymbol{S}\boldsymbol{\Sigma}^{-1}|] - \sum_{i=1}^{p} \log d_i, \tag{2.4}$$

which is smaller than $R(\omega, \widehat{\boldsymbol{\Sigma}}^0)$ for any ω and $p \geqslant 2$ since $-\sum_{i=1}^{p} \log d_i = \sum_{i=1}^{p} \log(n + p + 1 - 2i) \leqslant p\log n$.

The estimators $\widehat{\boldsymbol{\Sigma}}^0$ and $\widehat{\boldsymbol{\Sigma}}^{\mathrm{JS}}$ are based on the only statistic $\boldsymbol{S}$, and we are interested in developing estimators improving on them by using the information contained in the statistic $\boldsymbol{X}$. Some dominance results are provided below in the two cases of the rank $m = \mathrm{rank}\,(\boldsymbol{X})$: $m \geqslant p$ and $m < p$. A couple of estimators dominating $\widehat{\boldsymbol{\Sigma}}^0$ and/or $\widehat{\boldsymbol{\Sigma}}^{\mathrm{JS}}$ will be derived for $m \geqslant p$, while we can find out estimators dominating $\widehat{\boldsymbol{\Sigma}}^0$ in the case of $m < p$.

[1] *Case of* $m \geqslant p$: Let $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{X}^{\mathrm{t}}$, and $\boldsymbol{W}$ has full rank. Let $\boldsymbol{Q}$ be a $p \times p$ nonsingular matrix such that

$$\boldsymbol{S} = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{t}} \quad \text{and} \quad \boldsymbol{W} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{t}} \tag{2.5}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_p)$, $\lambda_1 \geqslant \cdots \geqslant \lambda_p$, namely, $\{\lambda_1, \ldots, \lambda_p\}$ is a set of eigenvalues of $\boldsymbol{S}^{-1}\boldsymbol{W}$. Then we consider estimators of the form

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}) = \boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{Q}^{\mathrm{t}}, \tag{2.6}$$

where $\boldsymbol{\Psi}(\boldsymbol{\Lambda}) = \mathrm{diag}\,(\psi_1(\boldsymbol{\Lambda}), \ldots, \psi_p(\boldsymbol{\Lambda}))$ is a diagonal matrix of absolutely continuous functions of $\boldsymbol{\Lambda}$. This is the general form of estimators satisfying the equivariance under the scale transformation $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{ASA}^{\mathrm{t}}, \boldsymbol{AWA}^{\mathrm{t}}) = \boldsymbol{A}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{S}, \boldsymbol{W})\boldsymbol{A}^{\mathrm{t}}$ for any $p \times p$ nonsingular matrix $\boldsymbol{A}$.

Kubokawa and Srivastava [12] treated a special class of (2.6), and $\boldsymbol{Q}$ in (2.5) corresponds to $\boldsymbol{W}^{1/2}\boldsymbol{P}\boldsymbol{\Lambda}^{-1/2}$ in their paper, where $\boldsymbol{W}^{1/2}$ is a symmetric half matrix of $\boldsymbol{W}$ and $\boldsymbol{P}$ is an orthogonal matrix such that $\boldsymbol{W}^{-1/2}\boldsymbol{S}\boldsymbol{W}^{-1/2} = \boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{P}^{\mathrm{t}}$. They provided some dominance results by using the conventional method based on a conditional distribution. We here obtain similar dominance results for the more general class of scale equivariant estimators $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ given by (2.6) by using a new method based on the Stein–Haff identity.

For any estimator $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ in (2.6), we define the truncation rule $[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{\mathrm{TR}}$ by

$$[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{\mathrm{TR}} = \min\left\{\boldsymbol{\Psi}(\boldsymbol{\Lambda}),\ (n+m)^{-1}(\boldsymbol{I}_p + \boldsymbol{\Lambda})\right\}, \tag{2.7}$$

where for two diagonal matrices $\boldsymbol{A} = \mathrm{diag}\,(a_1, \ldots, a_p)$ and $\boldsymbol{B} = \mathrm{diag}\,(b_1, \ldots, b_p)$, the notation $\min\{\boldsymbol{A}, \boldsymbol{B}\}$ denotes that $\min\{\boldsymbol{A}, \boldsymbol{B}\} = \mathrm{diag}\,\big(\min\{a_1, b_1\}, \ldots, \min\{a_p, b_p\}\big)$, which will be used throughout the paper. That is, $[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{\mathrm{TR}} = \mathrm{diag}(\psi_1^{\mathrm{TR}}(\boldsymbol{\Lambda}), \ldots, \psi_p^{\mathrm{TR}}(\boldsymbol{\Lambda}))$ for $\psi_i^{\mathrm{TR}}(\boldsymbol{\Lambda}) = \min\{\psi_i(\boldsymbol{\Lambda}),\ (n+m)^{-1}(1+\lambda_i)\}$, and the corresponding truncated estimator is of the form

$$\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{\mathrm{TR}}) = \boldsymbol{Q}\mathrm{diag}\,(\psi_1^{\mathrm{TR}}(\boldsymbol{\Lambda}), \ldots, \psi_p^{\mathrm{TR}}(\boldsymbol{\Lambda}))\boldsymbol{Q}^{\mathrm{t}}. \tag{2.8}$$

Then we get the following general dominance result which will be proved in the Appendix.

Theorem 2.1. *Assume that $m \geqslant p$. Then the truncated estimator $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{\mathrm{TR}})$ given by (2.8) dominates the scale-equivariant estimator $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ relative to the Stein loss function $L_{\mathrm{S}}(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$ if $P_\omega\big[[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{\mathrm{TR}} \neq \boldsymbol{\Psi}(\boldsymbol{\Lambda})\big] > 0$ at some ω.*

Before applying Theorem 2.1, we provide a couple of estimators within the class $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ which improve on the unbiased estimator $\widehat{\boldsymbol{\Sigma}}^{0}$ or the James–Stein estimator $\widehat{\boldsymbol{\Sigma}}^{\mathrm{JS}}$. The following expression of the risk is useful for this purpose.

Proposition 2.1. *For the risk function $R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}))$ of the scale-equivariant estimator $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ given by (2.6), the term $R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})) + E[\log|\boldsymbol{S}\boldsymbol{\Sigma}^{-1}]$ is expressed by*

$$\sum_{i=1}^{p} E\left[(n-p+2i-1)\psi_i - 2\lambda_i\frac{\partial\psi_i}{\partial\lambda_i} - 2\sum_{j>i}\frac{\lambda_i(\psi_i-\psi_j)}{\lambda_i-\lambda_j} - \log\psi_i\right] - p. \tag{2.9}$$

From Proposition 2.1, the unbiased estimator $\widehat{\boldsymbol{\Sigma}}^{0} = \boldsymbol{Q}(n^{-1}\boldsymbol{I}_p)\boldsymbol{Q}^{\mathrm{t}}$ has the risk that $R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}^{0}) = p\log n - E[\log|\boldsymbol{S}\boldsymbol{\Sigma}^{-1}]$, and applying Theorem 2.1 to the unbiased estimator yields the improved truncated one $\widehat{\boldsymbol{\Sigma}}^{\mathrm{UTR}} = \widehat{\boldsymbol{\Sigma}}([n^{-1}\boldsymbol{I}_p]^{\mathrm{TR}}) = \boldsymbol{Q}\min\{n^{-1}\boldsymbol{I}_p,\ (n+m)^{-1}(\boldsymbol{I}_p+\boldsymbol{\Lambda})\}\boldsymbol{Q}^{\mathrm{t}}$, which was given by Hara [6]. Putting $\boldsymbol{\Psi} = \boldsymbol{D}^*$ where

$$\boldsymbol{D}^* = \mathrm{diag}\,(d_1^*, \ldots, d_p^*) \quad \text{for } d_i^* = (n-p+2i-1)^{-1}, \tag{2.10}$$

we get the Stein type estimator

$$\widehat{\boldsymbol{\Sigma}}^{\mathrm{S}} = \boldsymbol{Q}\boldsymbol{D}^*\boldsymbol{Q}^{\mathrm{t}},$$

which has the risk $R(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}}^{\mathrm{S}}) = -E[\log|\boldsymbol{S}\mathbf{\Sigma}^{-1}|] - \sum_{i=1}^{p} E[2\sum_{j>i} \lambda_i(d_i^* - d_j^*)/(\lambda_i - \lambda_j) + \log d_i^*]$, being smaller than the risk (2.4) of the James–Stein estimator $\widehat{\mathbf{\Sigma}}^{\mathrm{JS}}$. Applying the truncation rule to $\widehat{\mathbf{\Sigma}}^{\mathrm{S}}$, from Theorem 2.1, we see that the Stein type estimator $\widehat{\mathbf{\Sigma}}^{\mathrm{S}}$ is minimax and improved on by the truncated estimator

$$\widehat{\mathbf{\Sigma}}^{\mathrm{STR}} = \widehat{\mathbf{\Sigma}}([\boldsymbol{D}^*]^{\mathrm{TR}}) = \boldsymbol{Q} \min\left\{\boldsymbol{D}^*, (n+m)^{-1}(\boldsymbol{I}_p + \mathbf{\Lambda})\right\} \boldsymbol{Q}^{\mathrm{t}}.$$

Letting $\mathbf{\Psi}^{\mathrm{H}} = (\phi_1^{\mathrm{H}}, \ldots, \psi_p^{\mathrm{H}})$ for $\psi_i^{\mathrm{H}} = n^{-1}(1 + \lambda_i\phi(\mathbf{\Lambda})/\mathrm{tr}\,\mathbf{\Lambda})$, we have the Haff type estimator given by

$$\widehat{\mathbf{\Sigma}}^{\mathrm{H}}(\phi) = \boldsymbol{Q}\mathbf{\Psi}^{\mathrm{H}}\boldsymbol{Q}^{\mathrm{t}} = \frac{1}{n}\left(\boldsymbol{S} + \frac{\phi(\mathbf{\Lambda})}{\mathrm{tr}\,\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{X}^{\mathrm{t}}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{t}}\right).$$

From Proposition 2.1, we can write the risk difference $R(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}}^{\mathrm{H}}(\phi)) - R(\mathbf{\Sigma}, \widehat{\mathbf{\Sigma}}^{0})$ as

$$E\left[\frac{\phi(\mathbf{\Lambda})}{n}\left\{n - p - 1 + 2\frac{\sum_{i=1}^{p}\lambda_i^2}{(\sum_{j=1}^{p}\lambda_j)^2}\right\} - \sum_{i=1}^{p}\log\left(1 + \phi(\mathbf{\Lambda})\frac{\lambda_i}{\sum_{j=1}^{p}\lambda_j}\right)\right]$$
$$-E\left[\frac{2}{n\sum_{j=1}^{p}\lambda_j}\sum_{i=1}^{p}\lambda_i^2\frac{\partial\phi(\mathbf{\Lambda})}{\partial\lambda_i}\right]. \tag{2.11}$$

Since $\log(1+x) \geqslant x - x^2/2$ for $x > 0$, it is seen that the risk difference (2.11) is bounded above by

$$E\left[\phi(\mathbf{\Lambda})\left\{-\frac{p-1}{n} + \frac{\phi(\mathbf{\Lambda})}{2}\right\} - \frac{2}{n\sum_{i=1}^{p}\lambda_i}\sum_{i=1}^{p}\lambda_i^2\frac{\partial\phi(\mathbf{\Lambda})}{\partial\lambda_i}\right],$$

which is smaller than or equal to zero under the conditions: (a) $0 \leqslant \phi(\mathbf{\Lambda}) \leqslant 2(p-1)/n$ and (b) $\phi(\mathbf{\Lambda})$ is nondecreasing in λ_i for $i = 1, \ldots, p$. This means that $\widehat{\mathbf{\Sigma}}^{\mathrm{H}}(\phi)$ dominates the unbiased estimator $\widehat{\mathbf{\Sigma}}^{0}$ under the two conditions. Applying the truncation rule to $\widehat{\mathbf{\Sigma}}^{\mathrm{H}}(\phi)$ yields the estimator $\widehat{\mathbf{\Sigma}}^{\mathrm{HTR}}(\phi) = \boldsymbol{Q}[\mathbf{\Psi}^{H}]^{\mathrm{TR}}\boldsymbol{Q}^{\mathrm{t}}$ which improves on $\widehat{\mathbf{\Sigma}}^{\mathrm{H}}(\phi)$, where

$$[\mathbf{\Psi}^{\mathrm{H}}]^{\mathrm{TR}} = \min\left\{n^{-1}\left(\boldsymbol{I}_p + [\phi(\mathbf{\Lambda})/\mathrm{tr}\,\mathbf{\Lambda}]\mathbf{\Lambda}\right),\ (n+m)^{-1}(\boldsymbol{I}_p + \mathbf{\Lambda})\right\}.$$

[2] *Case of* $m < p$: Let $\boldsymbol{F} = \boldsymbol{X}^{\mathrm{t}}\boldsymbol{S}^{-1}\boldsymbol{X}$ and write $\boldsymbol{F} = \boldsymbol{R}\mathbf{\Lambda}\boldsymbol{R}^{\mathrm{t}}$ for an $m \times m$ orthogonal matrix and a diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_m)$, $\lambda_1 \geqslant \cdots \geqslant \lambda_m$. Let us consider estimators of the form

$$\widehat{\mathbf{\Sigma}}(a, \mathbf{\Psi}) = a\boldsymbol{S} + \boldsymbol{X}\boldsymbol{R}\mathbf{\Psi}(\mathbf{\Lambda})\boldsymbol{R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}, \tag{2.12}$$

where a is a positive constant and $\mathbf{\Psi}(\mathbf{\Lambda}) = \mathbf{\Psi} = \mathrm{diag}\,(\psi_1(\mathbf{\Lambda}), \ldots, \psi_m(\mathbf{\Lambda}))$ for absolutely continuous functions $\psi_i(\mathbf{\Lambda})$'s. To improve on the estimator $\widehat{\mathbf{\Sigma}}(a, \mathbf{\Psi})$, we consider the truncation rule defined by

$$\widehat{\mathbf{\Sigma}}(a, [\mathbf{\Psi}]^{\mathrm{TR}}) = a\boldsymbol{S} + \boldsymbol{X}\boldsymbol{R}[\mathbf{\Psi}(\mathbf{\Lambda})]^{\mathrm{TR}}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}, \tag{2.13}$$

where $[\mathbf{\Psi}(\mathbf{\Lambda})]^{\mathrm{TR}} = [\mathbf{\Psi}]^{\mathrm{TR}} = \mathrm{diag}\,(\psi_1^{\mathrm{TR}}(\mathbf{\Lambda}), \ldots, \psi_m^{\mathrm{TR}}(\mathbf{\Lambda}))$ and $\psi_i^{\mathrm{TR}}(\mathbf{\Lambda})$ is given by

$$\psi_i^{\mathrm{TR}}(\mathbf{\Lambda}) = \min\left\{\psi_i,\ (n+m)^{-1} + \left[(n+m)^{-1} - a\right]\lambda_i^{-1}\right\}. \tag{2.14}$$

Then we get the following general dominance result which will be proved in the Appendix.

Theorem 2.2. *Assume that* $m < p$. *Then the truncated estimator* $\widehat{\boldsymbol{\Sigma}}(a, [\boldsymbol{\Psi}]^{\mathrm{TR}})$ *given by* (2.13) *dominates the scale equivariant estimator* $\widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{\Psi})$ *given by* (2.12) *relative to the Stein loss* $L_{\mathrm{S}}(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$ *if* $P_\omega\left[[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{\mathrm{TR}} \neq \boldsymbol{\Psi}(\boldsymbol{\Lambda})\right] > 0$ *at some* ω.

Before applying Theorem 2.2, we want to find superior estimators within the class $\widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{\Psi})$ given by (2.12). However, it does not seem possible to find a minimax estimator within the class. We thus derive a couple of estimators improving on the unbiased one $\widehat{\boldsymbol{\Sigma}}^0$. For this purpose, we obtain an expression of the risk of the estimator $\widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{\Psi})$.

Proposition 2.2. *For the risk function* $R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{\Psi}))$ *of the scale-equivariant estimator* $\widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{\Psi})$ *given by* (2.12), *the term* $h(\boldsymbol{\Sigma}, a, \boldsymbol{\Psi}) \equiv R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{\Psi})) + E[\log|\boldsymbol{S}\boldsymbol{\Sigma}^{-1}] - \{pna - (p-m)\log a - p\}$ *is expressed by*

$$\sum_{i=1}^{m} E\left[(n-p-1+2i)\lambda_i\psi_i - \log(a+\lambda_i\psi_i) -2\sum_{j>i}\frac{(\lambda_i\psi_i - \lambda_j\psi_j)\lambda_i}{\lambda_i-\lambda_j} - 2\,\lambda_i\frac{\partial(\lambda_i\psi_i)}{\partial\lambda_i}\right].$$

Let $\boldsymbol{\Psi} = \boldsymbol{C}\boldsymbol{\Lambda}^{-1}$ for $\boldsymbol{C} = \mathrm{diag}\,(c_1, \ldots, c_m)$. Then the term $h(\boldsymbol{\Sigma}, a, \boldsymbol{C}\boldsymbol{\Lambda}^{-1})$ of the resulting estimator

$$\widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{C}\boldsymbol{\Lambda}^{-1}) = a\boldsymbol{S} + \boldsymbol{X}\boldsymbol{R}\boldsymbol{C}\boldsymbol{\Lambda}^{-1}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}$$

has the form $\sum_{i=1}^{m}\{(n-p-1+2i)c_i - \log(a+c_i) - 2E[\sum_{j>i}(c_i-c_j)\lambda_i/(\lambda_i-\lambda_j)]\}$. If c_i's have the ordered relation that $c_1 \geqslant \cdots \geqslant c_m$, it is seen that $R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{C}\boldsymbol{\Lambda}^{-1})) + E[\log|\boldsymbol{S}\boldsymbol{\Sigma}^{-1}]$ is smaller than or equal to

$$pna - (p-m)\log a - p + \sum_{i=1}^{m}\{(n-p-1+2i)c_i - \log(a+c_i)\}, \tag{2.15}$$

which can be minimized by $(p-m)\log(n+m) + \sum_{i=1}^{m}\log(n-p-1+2i)$ at

$$a = \frac{1}{n+m} \equiv a_0 \quad \text{and} \quad c_i = \frac{m+p+1-2i}{(n+m)(n-p-1+2i)} \equiv c_{0i}. \tag{2.16}$$

Note that c_{0i}'s satisfy the ordering $c_{01} \geqslant \cdots \geqslant c_{0m}$. When $c_1 = \cdots = c_m = b$, it can be verified that the best constants a and b are given by $a = a_0$ and $b = p/\{(n+m)(n+m-p)\} \equiv b_0$. Since $\sum_{i=1}^{m}\log(n-p-1+2i) \leqslant m\log(n+m-p)$, it is seen that the estimator

$$\widehat{\boldsymbol{\Sigma}}(a_0, \boldsymbol{C}_0\boldsymbol{\Lambda}^{-1}) = a_0\boldsymbol{S} + \boldsymbol{X}\boldsymbol{R}\boldsymbol{C}_0\boldsymbol{\Lambda}^{-1}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}$$

for $\boldsymbol{C}_0 = \mathrm{diag}\,(c_{01}, \ldots, c_{0m})$ dominates the estimator $\widehat{\boldsymbol{\Sigma}}(a_0, b_0\boldsymbol{\Lambda}^{-1})$. Applying the truncation rule in Theorem 2.2, we get the truncated estimator

$$\widehat{\boldsymbol{\Sigma}}(a_0, [\boldsymbol{C}_0\boldsymbol{\Lambda}^{-1}]^{\mathrm{TR}}) = a_0\boldsymbol{S} + \boldsymbol{X}\boldsymbol{R}\min\left\{\boldsymbol{C}_0\boldsymbol{\Lambda}^{-1}, \boldsymbol{I}_m\right\}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}, \tag{2.17}$$

dominating $\widehat{\boldsymbol{\Sigma}}(a_0, \boldsymbol{C}_0\boldsymbol{\Lambda}^{-1})$.

Table 1
Risks of the estimators in the case of $m \geqslant p$ (UB and JS have the risks 1.183 and 1.079, respectively.)

	γ	UTR	S	STR	DTR	H	HTR
	0	0.951	0.806	0.725	0.730	1.002	0.730
$p = 4$	1	1.001	0.959	0.904	0.912	1.094	0.907
$m = 20$	2	1.025	0.981	0.937	0.946	1.127	0.967
	3	1.048	1.018	0.983	0.991	1.131	0.995
	4	1.055	1.039	1.005	1.012	1.132	1.003
	0	1.016	0.864	0.815	0.828	1.011	0.814
$p = 4$	1	1.043	0.964	0.930	0.942	1.060	0.912
$m = 10$	2	1.054	0.996	0.967	0.980	1.115	0.985
	3	1.072	1.021	0.999	1.010	1.125	1.014
	4	1.079	1.042	1.022	1.032	1.128	1.023
	0	1.100	0.970	0.947	0.958	1.037	0.937
$p = 4$	1	1.105	0.991	0.976	0.977	1.049	0.962
$m = 4$	2	1.109	1.036	1.025	1.028	1.100	1.025
	3	1.113	1.045	1.036	1.039	1.119	1.048
	4	1.118	1.054	1.048	1.051	1.125	1.059

2.3. Simulation studies

We here investigate the risk behaviors of several proposed estimators through Monte Carlo simulation. The reported risks are the averages of the loss functions based on 1,000,000 replications of $\boldsymbol{S}$ and $\boldsymbol{X}$ defined by (2.3) for $n = 10$, $p = 4$ and $m = 1, 3, 4, 10$ and 20. We set $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ and $\theta_{ij} = \gamma \times \{(i-1)/3 + (j-1)/5\}$ for $\gamma = 0, 1, 2, 3$ and 4 where θ_{ij} is the (i, j)th element of $\boldsymbol{\theta}$.

In the case of $m \geqslant p$, the estimators treated here are $\widehat{\boldsymbol{\Sigma}}^0 = n^{-1}\boldsymbol{S}$, the Stein type minimax estimator $\widehat{\boldsymbol{\Sigma}}^{\mathrm{S}} = \boldsymbol{Q}\boldsymbol{D}^*\boldsymbol{Q}^{\mathrm{t}}$, the Haff type estimator $\widehat{\boldsymbol{\Sigma}}^{\mathrm{H}}$ with $\phi(\boldsymbol{\Lambda}) = (p-1)/n$ and their truncated procedures $\widehat{\boldsymbol{\Sigma}}^{\mathrm{UTR}}$, $\widehat{\boldsymbol{\Sigma}}^{\mathrm{STR}}$ and $\widehat{\boldsymbol{\Sigma}}^{\mathrm{HTR}}$, which are denoted by UB, S, H, UTR, STR and HTR, respectively, for the sake of simplicity. Another estimator we want to inspect is of the form

$$\widehat{\boldsymbol{\Sigma}}^{\mathrm{DTR}} = \boldsymbol{Q} \min\left\{\boldsymbol{D}^*,\ \boldsymbol{D}^{**}(\boldsymbol{I} + \boldsymbol{\Lambda})\right\} \boldsymbol{Q}^{\mathrm{t}},$$

called DTR, which uses $\boldsymbol{D}^{**} = \mathrm{diag}\,((n+m-p+2i-1)^{-1}, i = 1, \ldots, p)$ instead of $(n+m)^{-1}$. Table 1 reports the values of the risks of the estimators UB, JS, UTR, S, STR, DTR, H and HTR for the three cases $m = 4$, 10 and 20, where JS means the James–Stein minimax estimator.

Table 1 reveals that the estimators S, STR, DTR and HTR are much better than JS and UTR, and that they get more improvements for $m = 20$ than for $m = 4$ and 10. Although DTR employs the two diagonal matrices $\boldsymbol{D}^*$ and $\boldsymbol{D}^{**}$, it does not provide more improvement than we expected. The risk differences of S and STR are smaller than those of UB and UTR, and we have a question about whether STR actually gives estimates closer to the true parameter $\boldsymbol{\Sigma}$ than S. To examine this issue, we shall compute the probabilities that STR (resp. S) is closer to the true parameter than S (resp. STR) and the probability that STR is identical to S, where a distance between an estimate $\boldsymbol{\delta}$ and the parameter $\boldsymbol{\Sigma}$ is measured by the Stein loss function $L(\boldsymbol{\delta}) = L_{\mathrm{S}}(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ defined by (1.2). We thus investigate the probabilities $P[L(\mathrm{STR}) < L(S)]$, $P[L(\mathrm{STR}) > L(S)]$ and $P[L(\mathrm{STR}) = L(S)]$. Similar kinds of probabilities for the estimators UTR and UB are also examined. The probabilities are computed through the above simulation experiments for $m = 4$, 20 and $\gamma = 0$, 1, 2, 3 and 4 and reported in the following table in percentage terms:

	$m = 4$					$m = 20$				
	0	1	2	3	4	0	1	2	3	4
$P[L(\mathrm{STR}) < L(S)]$	54.1	51.8	49.3	47.0	44.1	67.2	58.7	52.3	46.4	44.9
$P[L(\mathrm{STR}) > L(S)]$	43.7	42.8	42.8	41.6	41.5	27.0	23.0	19.3	18.9	18.9
$P[L(\mathrm{STR}) = L(S)]$	2.2	5.3	8.0	11.4	14.4	5.9	18.2	28.4	34.7	36.2
$P[L(\mathrm{UTR}) < L(\mathrm{UB})]$	65.7	65.4	64.9	64.6	63.6	83.5	81.1	79.5	75.7	74.4
$P[L(\mathrm{UTR}) > L(\mathrm{UB})]$	34.2	34.5	35.0	35.3	36.2	16.3	17.2	16.9	18.7	19.4
$P[L(\mathrm{UTR}) = L(\mathrm{UB})]$	0.0	0.0	0.0	0.1	0.2	0.0	1.6	3.6	5.6	6.2

In all the cases investigated, it is observed that $P[L(\mathrm{STR}) < L(S)] > P[L(\mathrm{STR}) > L(S)]$ and $P[L(\mathrm{UTR}) < L(\mathrm{UB})] > P[L(\mathrm{UTR}) > L(\mathrm{UB})]$, which means that the truncated estimators STR and UTR are closer than the untruncated ones S and UB, respectively. Although the difference between the probabilities $P[L(\mathrm{STR}) < L(S)]$ and $P[L(\mathrm{STR}) > L(S)]$ is small for $m = 4$, it is significant for $m = 20$. As expected, the probability $P[L(\mathrm{STR}) = L(S)]$ is higher than $P[L(\mathrm{UTR}) = L(\mathrm{UB})]$, and these probabilities get higher as γ is larger. The chance to take that $L(\mathrm{STR}) < L(S)$ is lower than $P[L(\mathrm{UTR}) < L(\mathrm{UB})]$. It is also revealed that $P[L(\mathrm{STR}) < L(S)] > P[L(\mathrm{STR}) = L(S)]$, namely, STR actually gives estimates closer to the true parameter $\mathbf{\Sigma}$ than S in the cases investigated here.

In the case of $m < p$, we handle the estimators $\widehat{\mathbf{\Sigma}}(a_0, b_0\mathbf{\Lambda}^{-1})$, $\widehat{\mathbf{\Sigma}}(a_0, \boldsymbol{C}_0\mathbf{\Lambda}^{-1})$ and their truncated procedures $\widehat{\mathbf{\Sigma}}(a_0, [b_0\mathbf{\Lambda}^{-1}]^{\mathrm{TR}})$ and $\widehat{\mathbf{\Sigma}}(a_0, [\boldsymbol{C}_0\mathbf{\Lambda}^{-1}]^{\mathrm{TR}})$ given by (2.17), which are denoted by B*, S*, BTR* and STR*, respectively. We also treat another estimator of the form

$$\widehat{\mathbf{\Sigma}}^{\mathrm{DTR}*} = \frac{n}{n+m}\boldsymbol{TDT}^{\mathrm{t}} + \boldsymbol{XR}\min\left\{\boldsymbol{C}_0\mathbf{\Lambda}^{-1}, \boldsymbol{I}_m\right\}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}},$$

called DTR*, which employs the James–Stein estimator instead of $n^{-1}\boldsymbol{S}$ where $\boldsymbol{T}$ and $\boldsymbol{D}$ are defined above (2.4). Table 2 reports the values of the risks of the estimators UB, JS, B*, BTR*, S*, STR* and DTR* for $m = 1$ and 3.

Table 2
Risks of the estimators in the case of $m < p$ (The risk of UB is 1.183.)

	γ	JS	$B*$	BTR*	$S*$	STR*	DTR*
	0	1.080	1.130	1.069	1.008	0.987	0.962
$p = 4$	1	1.080	1.130	1.074	1.018	1.002	1.015
$m = 3$	2	1.080	1.130	1.081	1.055	1.044	1.077
	3	1.080	1.130	1.083	1.062	1.053	1.087
	4	1.080	1.130	1.089	1.066	1.000	1.092
	0	1.079	1.112	1.096	1.112	1.096	1.015
$p = 4$	1	1.079	1.112	1.098	1.112	1.098	1.031
$m = 1$	2	1.079	1.112	1.111	1.112	1.111	1.082
	3	1.079	1.112	1.112	1.112	1.112	1.095
	4	1.079	1.112	1.112	1.112	1.112	1.098

From Table 2, it is revealed that S^*, STR* and DTR* are better than the other competitors for $m = 3$, but for $m = 1$, S^* and STR* are worse than JS. This demonstrates that the risk behaviors of S* and STR* are superior for large m close to p, but inferior for small m. Compared to S* and STR*, the estimator DTR* has the nice risk performances for $m = 1$ and 3, although the dominance property could not be shown analytically.

3. Estimation in mixed effects models

3.1. Canonical form

We begin with deriving a canonical form for the multivariate mixed effects linear model (1.1), where $\boldsymbol{\alpha}$ is assumed to be a random variable having the distribution $\mathcal{N}_{p,k}(\mathbf{0}, \boldsymbol{\Sigma}_A, \boldsymbol{C})$ for a $p \times p$ positive definite unknown matrix $\boldsymbol{\Sigma}_A$ and a $k \times k$ positive definite known matrix $\boldsymbol{C}$. From (2.1), it follows that the exponent in the joint density of $\boldsymbol{y}$ and $\boldsymbol{\alpha}$ is proportional to $\mathrm{tr}\, \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\beta}\boldsymbol{B} - \boldsymbol{\alpha}\boldsymbol{A})(\boldsymbol{y} - \boldsymbol{\beta}\boldsymbol{B} - \boldsymbol{\alpha}\boldsymbol{A})^{\mathrm{t}} + \mathrm{tr}\, \boldsymbol{\Sigma}_A^{-1}\boldsymbol{\alpha}\boldsymbol{C}\boldsymbol{\alpha}^{\mathrm{t}} = \mathrm{tr}\, \boldsymbol{\Sigma}^{-1}\boldsymbol{S}_1 + \mathrm{tr}\, \boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})\boldsymbol{B}_1\boldsymbol{B}_1^{\mathrm{t}}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})^{\mathrm{t}} + J_*$, where $J_* = \mathrm{tr}\, \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\beta}\boldsymbol{B}_2 - \boldsymbol{\alpha}\boldsymbol{A}^*)(\boldsymbol{y}_2 - \boldsymbol{\beta}\boldsymbol{B}_2 - \boldsymbol{\alpha}\boldsymbol{A}^*)^{\mathrm{t}} + \mathrm{tr}\, \boldsymbol{\Sigma}_A^{-1}\boldsymbol{\alpha}\boldsymbol{C}\boldsymbol{\alpha}^{\mathrm{t}}$. Let $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}\boldsymbol{C}^{1/2}$ for a symmetric half matrix $\boldsymbol{C}^{1/2}$ of $\boldsymbol{C}$, namely $\boldsymbol{C} = (\boldsymbol{C}^{1/2})^2$. Consider the spectral decomposition $\boldsymbol{C}^{-1/2}\boldsymbol{A}^*\boldsymbol{A}^{*\mathrm{t}}\boldsymbol{C}^{-1/2} = \sum_{i=1}^{\ell} \tau_i \boldsymbol{E}_i$, where $\tau_1, \dots, \tau_\ell$ are positive eigenvalues of $\boldsymbol{C}^{-1/2}\boldsymbol{A}^*\boldsymbol{A}^{*\mathrm{t}}\boldsymbol{C}^{-1/2}$, and $\boldsymbol{E}_1, \dots, \boldsymbol{E}_\ell$ are idempotent matrices such that $\sum_{i=1}^{\ell} \mathrm{rank}\,(\boldsymbol{E}_i) = k$ and $\sum_{i=1}^{\ell} \boldsymbol{E}_i = \boldsymbol{I}_k$. Then,

$$
\begin{aligned}
J_* = & \sum_{i=1}^{\ell} \mathrm{tr}\,(\boldsymbol{\Sigma} + \tau_i\boldsymbol{\Sigma}_A)^{-1}(\boldsymbol{y}_2 - \boldsymbol{\beta}\boldsymbol{B}_2)\boldsymbol{E}_i(\boldsymbol{y}_2 - \boldsymbol{\beta}\boldsymbol{B}_2)^{\mathrm{t}} \\
& + \sum_{i=1}^{\ell} \mathrm{tr}\,(\tau_i\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_A^{-1})[\boldsymbol{\alpha}^* - (\tau_i\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_A^{-1})^{-1}\sqrt{\tau_i}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\beta}\boldsymbol{B}_2)] \\
& \times \boldsymbol{E}_i[\boldsymbol{\alpha}^* - (\tau_i\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_A^{-1})^{-1}\sqrt{\tau_i}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\beta}\boldsymbol{B}_2)]^{\mathrm{t}}.
\end{aligned} \tag{3.1}
$$

Hence, $\boldsymbol{y}_2$ in (2.1) is decomposed as follows: $\boldsymbol{y}_2\boldsymbol{E}_1, \dots, \boldsymbol{y}_2\boldsymbol{E}_\ell$ are mutually independently distributed as $\boldsymbol{y}_2\boldsymbol{E}_i \sim \mathcal{N}_{p,k}(\boldsymbol{\beta}\boldsymbol{B}_2\boldsymbol{E}_i, \boldsymbol{\Sigma} + \tau_i\boldsymbol{\Sigma}_A, \boldsymbol{E}_i)$ for $i = 1, \dots, \ell$. Let $\widehat{\boldsymbol{\beta}}_2$ be the least squares estimator of $\boldsymbol{\beta}$ in terms of minimizing the quadratic form $(\boldsymbol{y}_2 - \boldsymbol{\beta}\boldsymbol{B}_2)(\boldsymbol{y}_2 - \boldsymbol{\beta}\boldsymbol{B}_2)^{\mathrm{t}}$, and $\widehat{\boldsymbol{\beta}}_2$ is given by $\widehat{\boldsymbol{\beta}}_2 = \boldsymbol{y}_2\boldsymbol{B}_2^{\mathrm{t}}(\boldsymbol{B}_2\boldsymbol{B}_2^{\mathrm{t}})^-$. Then the first term in the r.h.s. of the Eq. (3.1) can be rewritten as

$$
\begin{aligned}
& \sum_{i=1}^{\ell} \mathrm{tr}\,(\boldsymbol{\Sigma} + \tau_i\boldsymbol{\Sigma}_A)^{-1}(\boldsymbol{y}_2 - \widehat{\boldsymbol{\beta}}_2\boldsymbol{B}_2)\boldsymbol{E}_i(\boldsymbol{y}_2 - \widehat{\boldsymbol{\beta}}_2\boldsymbol{B}_2)^{\mathrm{t}} \\
& \quad + \sum_{i=1}^{\ell} \mathrm{tr}\,(\boldsymbol{\Sigma} + \tau_i\boldsymbol{\Sigma}_A)^{-1}(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta})\boldsymbol{B}_2\boldsymbol{E}_i\boldsymbol{B}_2^{\mathrm{t}}(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta})^{\mathrm{t}} \\
& \quad + 2\sum_{i=1}^{\ell} \mathrm{tr}\,(\boldsymbol{\Sigma} + \tau_i\boldsymbol{\Sigma}_A)^{-1}(\boldsymbol{y}_2 - \widehat{\boldsymbol{\beta}}_2\boldsymbol{B}_2)\boldsymbol{E}_i\boldsymbol{B}_2^{\mathrm{t}}(\widehat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta})^{\mathrm{t}}.
\end{aligned} \tag{3.2}
$$

Let $\boldsymbol{T}_i = (\boldsymbol{y}_2 - \widehat{\boldsymbol{\beta}}_2\boldsymbol{B}_2)\boldsymbol{E}_i(\boldsymbol{y}_2 - \widehat{\boldsymbol{\beta}}_2\boldsymbol{B}_2)^{\mathrm{t}}$ and $m_i = \mathrm{rank}\,[(\boldsymbol{I}_k - \boldsymbol{B}_2^{\mathrm{t}}(\boldsymbol{B}_2\boldsymbol{B}_2^{\mathrm{t}})^-\boldsymbol{B}_2)\boldsymbol{E}_i]$ for $i = 1, \dots, \ell$. Note that $\sum_{i=1}^{\ell} m_i = k - q_2$ for $q_2 = \mathrm{rank}\,(\boldsymbol{B}_2\boldsymbol{B}_2^{\mathrm{t}})$. It is also noted that there exists a $p \times m_i$ matrix $\boldsymbol{Z}_i$ such that $\boldsymbol{T}_i = \boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{t}}$ and

$$
\boldsymbol{Z}_i \sim \mathcal{N}_{p,m_i}(\mathbf{0}, \boldsymbol{\Sigma} + \tau_i\boldsymbol{\Sigma}_A, \boldsymbol{I}_{m_i}).
$$

From (3.2), it is seen that $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_\ell$ are mutually independent. Let us define $\boldsymbol{S}_2$ and n_2 by

$$\boldsymbol{S}_2 = \sum_{i=1}^{\ell} \boldsymbol{T}_i = (\boldsymbol{y}_2 - \widehat{\boldsymbol{\beta}}_2\boldsymbol{B}_2)(\boldsymbol{y}_2 - \widehat{\boldsymbol{\beta}}_2\boldsymbol{B}_2)^{\mathrm{t}} \quad \text{and} \quad n_2 = \sum_{i=1}^{\ell} m_i = k - q_2. \tag{3.3}$$

Then, $\boldsymbol{S}_2$ is expressed by $\boldsymbol{S}_2 = \sum_{i=1}^{\ell} \boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{t}} = \boldsymbol{Z}\boldsymbol{Z}^{\mathrm{t}}$ for the $p \times n_2$ matrix $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_\ell)$. In general, it is noted that $\boldsymbol{S}_2$ is not independent of $\widehat{\boldsymbol{\beta}}_2$. In the balanced case that $\tau_1 = \cdots = \tau_\ell = \tau$, however, $\boldsymbol{S}_2$ and $\widehat{\boldsymbol{\beta}}_2$ are mutually independent and

$$\begin{aligned} &\boldsymbol{S}_2 \sim \mathcal{W}_p(n_2, \tau\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}), \\ &\widehat{\boldsymbol{\beta}}_2\boldsymbol{B}_2 \sim \mathcal{N}_{p,k}(\boldsymbol{\beta}\boldsymbol{B}_2, \boldsymbol{\Sigma} + \tau\boldsymbol{\Sigma}_A, \boldsymbol{B}_2^{\mathrm{t}}(\boldsymbol{B}_2\boldsymbol{B}_2^{\mathrm{t}})^{-}\boldsymbol{B}_2). \end{aligned} \tag{3.4}$$

Using the statistics $\boldsymbol{S}$ and $\boldsymbol{S}_2$, we want to construct improved estimators of the multivariate 'within' and 'between' components of variance $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_A$.

3.2. Dominance results

[1] *Estimation of* $\boldsymbol{\Sigma}$: We first consider the estimation of the 'within' component of variance $\boldsymbol{\Sigma}$ in the mixed linear model relative to the Stein loss (1.2). When $\boldsymbol{\Sigma}$ is estimated based on the statistics $\boldsymbol{S}$ and $\boldsymbol{S}_2$, the two cases are treated, namely, $n_2 \geqslant p$ and $n_2 < p$ for n_2 given in (3.3). When $n_2 \geqslant p$, $\boldsymbol{S}_2$ has full rank, and the same arguments as in the case of $m \geqslant p$ in Section 2 are used to develop the corresponding improved estimators. Let $\boldsymbol{Q}$ be a $p \times p$ matrix such that

$$\boldsymbol{S} = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{t}} \quad \text{and} \quad \boldsymbol{S}_2 = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{t}}, \tag{3.5}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_p)$, $\lambda_1 \geqslant \cdots \geqslant \lambda_p$, and we consider the estimator $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}) = \boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{Q}^{\mathrm{t}}$ and the truncated one

$$\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{\mathrm{TR}}) = \boldsymbol{Q}[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{\mathrm{TR}}\boldsymbol{Q}^{\mathrm{t}}\boldsymbol{Q} \min\left\{\boldsymbol{\Psi}(\boldsymbol{\Lambda}),\ (n+n_2)^{-1}(\boldsymbol{I}_p + \boldsymbol{\Lambda})\right\}\boldsymbol{Q}^{\mathrm{t}}, \tag{3.6}$$

for $\boldsymbol{\Psi}(\boldsymbol{\Lambda})$ and $[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{\mathrm{TR}}$ given by (2.6) and (2.8) where $\boldsymbol{W}$ and m in (2.5) and (2.7) are replaced by $\boldsymbol{S}_2$ and n_2 in (3.3). Then the dominance property of $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{\mathrm{TR}})$ over $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ can be verified from Theorem 2.1. In fact, note that for $i = 1, \ldots, \ell$, $\boldsymbol{Z}_i$ can be expressed as using the mixture model

$$\boldsymbol{Z}_i|\boldsymbol{\xi}_i \sim \mathcal{N}_{p,m_i}(\boldsymbol{\xi}_i, \boldsymbol{\Sigma}, \boldsymbol{I}_{m_i}) \quad \text{and} \quad \boldsymbol{\xi}_i \sim \mathcal{N}_{p,m_i}(\boldsymbol{0}, \tau_i\boldsymbol{\Sigma}_A, \boldsymbol{I}_{m_i}),$$

where $\boldsymbol{\xi}_i$ is a $p \times m_i$ matrix. Let $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_\ell)$, and given $\boldsymbol{\xi}$, $\boldsymbol{Z}$ is conditionally distributed as $\boldsymbol{Z}|\boldsymbol{\xi} \sim \mathcal{N}_{p,n_2}(\boldsymbol{\xi}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n_2})$. This conditional model given $\boldsymbol{\xi}$ corresponds to the fixed effects model (2.3), and the dominance result follows from Theorem 2.1.

Corollary 3.1. *Assume that* $n_2 \geqslant p$. *Then the general scale-equivariant estimator* $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ *is improved on by the truncated one* $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{\mathrm{TR}})$ *relative to the Stein loss function* $L_{\mathrm{S}}(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$ *if* $P\left[[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{\mathrm{TR}} \neq \boldsymbol{\Psi}(\boldsymbol{\Lambda})\right] > 0$ *at some* ω *for* $\omega = (\boldsymbol{\Sigma}, \boldsymbol{\Sigma}_A, \boldsymbol{\beta})$.

Srivastava and Kubokawa (1999) derived similar dominance results in the special case that $\tau_1 = \cdots = \tau_\ell = \tau$ and $\boldsymbol{Q} = \boldsymbol{S}_2^{1/2}\boldsymbol{P}\boldsymbol{\Lambda}^{-1/2}$ for an orthogonal matrix $\boldsymbol{P}$ such that $\boldsymbol{S}_2^{-1/2}\boldsymbol{S}\boldsymbol{S}_2^{-1/2} = \boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{P}^{\mathrm{t}}$. It shows that the results of Srivastava and Kubokawa [18] can be extended, not only to the general mixed effects models, but also to the general class of scale-equivariant estimators.

All the dominance results which follow from Theorem 2.1 still hold in the mixed effects model by replacing $\boldsymbol{W}$ and m with $\boldsymbol{S}_2$ and n_2. For example, applying Corollary 3.1 to the unbiased estimator $\widehat{\boldsymbol{\Sigma}}^0 = n^{-1}\boldsymbol{S}$ and the Stein type minimax estimator $\widehat{\boldsymbol{\Sigma}}^{\mathrm{S}} = \boldsymbol{Q}\boldsymbol{D}^*\boldsymbol{Q}^{\mathrm{t}}$ gives the REML (restricted maximum likelihood) estimator and the truncated Stein type estimators, respectively, given by

$$
\begin{aligned}
\widehat{\boldsymbol{\Sigma}}^{\mathrm{REML}} &= \boldsymbol{Q}\min\left\{n^{-1}\boldsymbol{I}_p,\ (n+n_2)^{-1}(\boldsymbol{I}_p+\boldsymbol{\Lambda})\right\}\boldsymbol{Q}^{\mathrm{t}},\\
\widehat{\boldsymbol{\Sigma}}^{\mathrm{STR}} &= \boldsymbol{Q}\min\left\{\boldsymbol{D}^*,\ (n+n_2)^{-1}(\boldsymbol{I}_p+\boldsymbol{\Lambda})\right\}\boldsymbol{Q}^{\mathrm{t}},
\end{aligned}
\tag{3.7}
$$

where $\boldsymbol{D}^*$ is defined by (2.10) for $m = n_2$.

The method and the results of Srivastava and Kubokawa [18] requests that n_2, the rank of $\boldsymbol{S}_2$, is greater than or equal to the dimension p. To handle the case of $n_2 < p$, we here suggest the following two approaches. One approach to resolving the problem is to use the results given in the case of $m < p$ in Section 2. Let $\boldsymbol{F} = \boldsymbol{Z}^{\mathrm{t}}\boldsymbol{S}^{-1}\boldsymbol{Z}$ and write $\boldsymbol{F} = \boldsymbol{R}\boldsymbol{\Lambda}\boldsymbol{R}^{\mathrm{t}}$ for an $n_2 \times n_2$ orthogonal matrix and a diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_{n_2})$, $\lambda_1 \geqslant \cdots \geqslant \lambda_{n_2}$. Let us consider estimators of the form

$$
\begin{aligned}
&\widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{\Psi}) = a\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{R}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{R}^{\mathrm{t}}\boldsymbol{Z}^{\mathrm{t}},\\
&\widehat{\boldsymbol{\Sigma}}(a, [\boldsymbol{\Psi}]^{\mathrm{TR}}) = a\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{R}[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{\mathrm{TR}}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{Z}^{\mathrm{t}},
\end{aligned}
$$

where a is a positive constant, and $\boldsymbol{\Psi}(\boldsymbol{\Lambda})$ and $[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{\mathrm{TR}}$ are given around (2.12), (2.13) and (2.14) by replacing $\boldsymbol{X}$ and m with $\boldsymbol{Z}$ and n_2. Using the conditional argument given $\boldsymbol{\xi}$, we can get the following dominance result from Theorem 2.2.

Corollary 3.2. *Assume that $n_2 < p$. Then the scale-equivariant estimator $\widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{\Psi})$ is dominated by the truncated one $\widehat{\boldsymbol{\Sigma}}(a, [\boldsymbol{\Psi}]^{\mathrm{TR}})$ relative to the Stein loss function $L_{\mathrm{S}}(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$ if $P_\omega\left[[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{\mathrm{TR}} \neq \boldsymbol{\Psi}(\boldsymbol{\Lambda})\right] > 0$ at some ω.*

Using the arguments below Theorem 2.2 with replacing m by n_2, we can apply Corollary 3.2 to $\widehat{\boldsymbol{\Sigma}}(a_0, b_0\boldsymbol{\Lambda}^{-1})$ for $a_0 = 1/(n+n_2)$ and $b_0 = p/\{(n+n_2)(n+n_2-p)\}$ and $\widehat{\boldsymbol{\Sigma}}(a_0, \boldsymbol{C}_0\boldsymbol{\Lambda}^{-1})$ for $\boldsymbol{C}_0 = \mathrm{diag}\,(c_{01}, \ldots, c_{0n_2})$ given by (2.16), and obtain their truncated estimators

$$
\begin{aligned}
&\widehat{\boldsymbol{\Sigma}}(a_0, [b_0\boldsymbol{\Lambda}^{-1}]^{\mathrm{TR}}) = a_0\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{R}\min\left\{b_0\boldsymbol{\Lambda}^{-1}, \boldsymbol{I}_{n_2}\right\}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{Z}^{\mathrm{t}},\\
&\widehat{\boldsymbol{\Sigma}}(a_0, [\boldsymbol{C}_0\boldsymbol{\Lambda}^{-1}]^{\mathrm{TR}}) = a_0\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{R}\min\left\{\boldsymbol{C}_0\boldsymbol{\Lambda}^{-1}, \boldsymbol{I}_{n_2}\right\}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{Z}^{\mathrm{t}}.
\end{aligned}
$$

Another approach to handling the case of $n_2 < p$ is to combine the quadratic statistics. As seen in (3.1) and (3.2), the quadratic statistic $\boldsymbol{y}_2\boldsymbol{y}_2^{\mathrm{t}}$ is decomposed into $\boldsymbol{S}_2$ and $\widehat{\boldsymbol{\beta}}_2\boldsymbol{B}_2\boldsymbol{B}_2^{\mathrm{t}}\widehat{\boldsymbol{\beta}}_2^{\mathrm{t}}$, namely,

$$
\boldsymbol{y}_2\boldsymbol{y}_2^{\mathrm{t}} = \boldsymbol{S}_2 + \widehat{\boldsymbol{\beta}}_2\boldsymbol{B}_2\boldsymbol{B}_2^{\mathrm{t}}\widehat{\boldsymbol{\beta}}_2^{\mathrm{t}},
$$

which has the rank k. When $k \geqslant p$, we can apply Theorem 2.1 to provide the result in Corollary 3.1 by replacing $\boldsymbol{S}_2$ and n_2 with $\boldsymbol{y}_2\boldsymbol{y}_2^{\mathrm{t}}$ and k. If $k < p$, then combining $\boldsymbol{y}_2\boldsymbol{y}_2^{\mathrm{t}}$ and $\widehat{\boldsymbol{\beta}}_1\boldsymbol{B}_1\boldsymbol{B}_1^{\mathrm{t}}\widehat{\boldsymbol{\beta}}_1^{\mathrm{t}}$ in (2.2) yields

$$
\boldsymbol{X}\boldsymbol{X}^{\mathrm{t}} = \boldsymbol{y}_2\boldsymbol{y}_2^{\mathrm{t}} + \widehat{\boldsymbol{\beta}}_1\boldsymbol{B}_1\boldsymbol{B}_1^{\mathrm{t}}\widehat{\boldsymbol{\beta}}_1^{\mathrm{t}},
$$

with the rank $m = k + q_1$. When $m \geqslant p$, Theorem 2.1 can be used to get improved procedures. Otherwise, we can use the results given in the case $m < p$ in Section 2.

[2] *Estimation of* $\mathbf{\Sigma}_A$: We next consider the estimation of the multivariate 'between' component of variance $\mathbf{\Sigma}_A$ in the context of the simultaneous estimation of the 'within' and 'between' components $(\mathbf{\Sigma}, \mathbf{\Sigma}_A)$. Especially, we here treat the balanced case that $\tau_1 = \cdots = \tau_\ell = \tau$, since it is intractable to deal with the general cases of τ_i's. Then from (3.4), $\boldsymbol{S}_2$ is distributed as

$$\boldsymbol{S}_2 \sim \mathcal{W}_p(\mathbf{\Sigma}_2, n_2) \quad \text{for } \mathbf{\Sigma}_2 = \mathbf{\Sigma} + \tau\mathbf{\Sigma}_A,$$

and it is assumed that $n_2 \geqslant p$. An unbiased estimator of $\mathbf{\Sigma}_A$ is given by $\widehat{\mathbf{\Sigma}}_A^0 = \tau^{-1}\left\{n_2^{-1}\boldsymbol{S}_2 - n^{-1}\boldsymbol{S}\right\}$, which has a drawback of taking negative values with a positive probability. This is why we cannot employ the Stein loss $L_S(\widehat{\mathbf{\Sigma}}_A, \mathbf{\Sigma}_A)$ for the function $L_S(\cdot,\cdot)$ given by (1.2). Instead of the Stein loss, Srivastava and Kubokawa [18] proposed the use of the Kullback–Leibler loss function

$$\begin{aligned} L_{KL}(\widehat{\mathbf{\Sigma}}, \widehat{\mathbf{\Sigma}}_A; \mathbf{\Sigma}, \mathbf{\Sigma}_A) &= nL_S(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}) + n_2 L_S(\widehat{\mathbf{\Sigma}} + \tau\widehat{\mathbf{\Sigma}}_A, \mathbf{\Sigma} + \tau\mathbf{\Sigma}_A) \\ &= nL_S(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma}) + n_2 L_S(\widehat{\mathbf{\Sigma}}_2, \mathbf{\Sigma}_2) \end{aligned} \tag{3.8}$$

for $\widehat{\mathbf{\Sigma}}_2 = \widehat{\mathbf{\Sigma}} + \tau\widehat{\mathbf{\Sigma}}_A$, and considered the simultaneous estimation of $(\mathbf{\Sigma}, \mathbf{\Sigma}_A)$. Since the estimation of $\mathbf{\Sigma}$ in terms of the risk $R_1(\omega; \widehat{\mathbf{\Sigma}}_1) = E[L_S(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma})]$ has been treated previously, we address the problem of estimating $\mathbf{\Sigma}_2$ in terms of the risk $R_2(\omega; \widehat{\mathbf{\Sigma}}_2) = E[L_S(\widehat{\mathbf{\Sigma}}_2, \mathbf{\Sigma}_2)]$.

The general class of scale-equivariant estimators of $\mathbf{\Sigma}_2$ are given by $\widehat{\mathbf{\Sigma}}_2(\mathbf{\Phi}) = \boldsymbol{Q}\mathbf{\Phi}(\mathbf{\Lambda})\boldsymbol{Q}^t$ where $\boldsymbol{Q}$ and $\mathbf{\Lambda}$ are defined by (3.5), and $\mathbf{\Phi}(\mathbf{\Lambda}) = \mathrm{diag}\,(\phi_1(\mathbf{\Lambda}), \ldots, \phi_p(\mathbf{\Lambda}))$. The use of the information that $\mathbf{\Sigma} \leqslant \mathbf{\Sigma} + \tau\mathbf{\Sigma}_A = \mathbf{\Sigma}_2$ leads to the improvement on $\widehat{\mathbf{\Sigma}}_2(\mathbf{\Phi})$ by the truncated estimator

$$\widehat{\mathbf{\Sigma}}_2([\mathbf{\Phi}]_{TR}) = \boldsymbol{Q}[\mathbf{\Phi}(\mathbf{\Lambda})]_{TR}\boldsymbol{Q}^t, \tag{3.9}$$

for $[\mathbf{\Phi}(\mathbf{\Lambda})]_{TR} = \max\left\{\mathbf{\Phi}(\mathbf{\Lambda}),\ (n+n_2)^{-1}(\boldsymbol{I}_p + \mathbf{\Lambda})\right\}$, where the notation $\max\{\boldsymbol{A}, \boldsymbol{B}\}$ denotes that $\max\{\boldsymbol{A}, \boldsymbol{B}\} = \mathrm{diag}\,\left(\max\{a_1, b_1\}, \ldots, \max\{a_p, b_p\}\right)$ for two diagonal matrices $\boldsymbol{A} = \mathrm{diag}\,(a_1, \ldots, a_p)$ and $\boldsymbol{B} = \mathrm{diag}\,(b_1, \ldots, b_p)$. The following proposition can be proved by the same arguments as in the proof of Theorem 2.1 based on the Stein–Haff identity.

Proposition 3.1. *Assume that $n_2 \geqslant p$ and $\tau_1 = \cdots = \tau_\ell$. Then the scale-equivariant estimator $\widehat{\mathbf{\Sigma}}_2(\mathbf{\Phi})$ is dominated by the truncated one $\widehat{\mathbf{\Sigma}}_2([\mathbf{\Phi}]_{TR})$ relative to the Stein loss function $L_S(\widehat{\mathbf{\Sigma}}_2, \mathbf{\Sigma}_2)$ if $P_\omega\,[[\mathbf{\Phi}(\mathbf{\Lambda})]_{TR} \neq \mathbf{\Phi}(\mathbf{\Lambda})] > 0$ at some ω.*

From (3.6) and (3.9), we obtain the truncated estimator $\widehat{\mathbf{\Sigma}}_A^{TR}$ of $\mathbf{\Sigma}_A$, given by

$$\widehat{\mathbf{\Sigma}}_A([\mathbf{\Phi}]_{TR}, [\mathbf{\Psi}]^{TR}) = \frac{1}{\tau}\boldsymbol{Q}\left[[\mathbf{\Phi}]_{TR} - [\mathbf{\Psi}]^{TR}\right]\boldsymbol{Q}^t,$$

which is always nonnegative definite. Combining Corollary 3.1 and Proposition 3.1 gives the following dominance result.

Corollary 3.3. *Assume that $n_2 \geqslant p$. In the framework of the simultaneous estimation of $(\mathbf{\Sigma}, \mathbf{\Sigma}_A)$, a scale-equivariant estimator $(\widehat{\mathbf{\Sigma}}(\mathbf{\Psi}),\ \widehat{\mathbf{\Sigma}}_A(\mathbf{\Phi}, \mathbf{\Psi}))$ for $\widehat{\mathbf{\Sigma}}_A(\mathbf{\Phi}, \mathbf{\Psi}) = \tau^{-1}\boldsymbol{Q}\{\mathbf{\Phi} - \mathbf{\Psi}\}\boldsymbol{Q}^t$ is improved on by the truncated one $(\widehat{\mathbf{\Sigma}}([\mathbf{\Psi}]^{TR}),\ \widehat{\mathbf{\Sigma}}_A([\mathbf{\Phi}]_{TR}, [\mathbf{\Psi}]^{TR}))$ relative to the Kullback–Leibler loss $L_{KL}(\widehat{\mathbf{\Sigma}}, \widehat{\mathbf{\Sigma}}_A; \mathbf{\Sigma}, \mathbf{\Sigma}_A)$ if the two estimators are different with a positive probability.*

Applying Proposition 3.1 to the unbiased estimator $\widehat{\mathbf{\Sigma}}_2^0 = n_2^{-1}\boldsymbol{S}_2 = \boldsymbol{Q}(n_2^{-1}\mathbf{\Lambda})\boldsymbol{Q}^t$ yields the improved truncated one $\widehat{\mathbf{\Sigma}}_2([n_2^{-1}\mathbf{\Lambda}]_{TR}) = \boldsymbol{Q}\max\{n_2^{-1}\mathbf{\Lambda},\ (n+n_2)^{-1}(\boldsymbol{I}_p + \mathbf{\Lambda})\}\boldsymbol{Q}^t$, which leads to the REML estimator of the 'between' component of variance $\boldsymbol{\Sigma}_A$, given by

$$\widehat{\boldsymbol{\Sigma}}_A^{\mathrm{REML}} = \frac{1}{\tau}\left\{\widehat{\boldsymbol{\Sigma}}_2^{\mathrm{TR}} - \widehat{\boldsymbol{\Sigma}}^{\mathrm{REML}}\right\} = \frac{1}{\tau}\max\left\{n_2^{-1}\boldsymbol{S}_2 - n^{-1}\boldsymbol{S}, \boldsymbol{0}\right\}.$$

We thus see that the REML estimator $(\widehat{\boldsymbol{\Sigma}}^{\mathrm{REML}}, \widehat{\boldsymbol{\Sigma}}_A^{\mathrm{REML}})$ dominates the unbiased one $(\widehat{\boldsymbol{\Sigma}}^0, \widehat{\boldsymbol{\Sigma}}_A^0)$. It can be shown that a Stein type minimax estimator of $\boldsymbol{\Sigma}_2$ is given by $\widehat{\boldsymbol{\Sigma}}_2^{\mathrm{S}} = \boldsymbol{QD\Lambda Q}^{\mathrm{t}}$, where $\boldsymbol{D} = \mathrm{diag}\,(d_1, \ldots, d_p)$ for $d_i = 1/(n_2 + p + 1 - 2i)$. Applying Proposition 3.1 produces the improved truncated one $\boldsymbol{\Sigma}_2^{\mathrm{STR}} = \boldsymbol{Q}\max\{\boldsymbol{D\Lambda}, (n+n_2)^{-1}(\boldsymbol{I}_p + \boldsymbol{\Lambda})\}\boldsymbol{Q}^{\mathrm{t}}$, leading to the estimator of $\boldsymbol{\Sigma}_A$:

$$\widehat{\boldsymbol{\Sigma}}_A^{\mathrm{STR}} = \frac{1}{\tau}\left\{\widehat{\boldsymbol{\Sigma}}_2^{\mathrm{STR}} - \widehat{\boldsymbol{\Sigma}}^{\mathrm{STR}}\right\} = \frac{1}{\tau}\boldsymbol{Q}\max\left\{\boldsymbol{D\Lambda} - \boldsymbol{D}^*, \boldsymbol{0}\right\}\boldsymbol{Q}^{\mathrm{t}}. \tag{3.10}$$

It can be observed that $(\widehat{\boldsymbol{\Sigma}}^{\mathrm{STR}}, \widehat{\boldsymbol{\Sigma}}_A^{\mathrm{STR}})$ dominates the set of the James–Stein estimators as well as $(\widehat{\boldsymbol{\Sigma}}^0, \widehat{\boldsymbol{\Sigma}}_A^0)$.

Although we would get similar dominance results for $\boldsymbol{\Sigma}_A$ in the general cases of $\tau_1, \ldots, \tau_\ell$, it is intractable to treat the cases due to technical difficulty.

4. Concluding remarks

In this paper, we have developed the truncation rules which provide truncated estimators dominating the general scale equivariant estimators. Although applying the truncation rules to ordinary types of estimators yields the improved estimators having good risk behaviors, such truncated estimators have theoretical drawbacks of inadmissibility. Alternative methods are the Bayesian rules, and the most interesting issue is to find prior distributions such that the resulting Bayes estimators dominate the James–Stein minimax estimator. Yang and Berger [22] and Berger et al. [2] derived the reference priors and proposed the use of the Bayes estimators against the reference priors. Although the Bayes estimator of Yang and Berger [22] is not in the class of scale equivariant estimators, some types of reference priors in Berger et al. [2] will enjoy the scale equivariant property. It would be of great interest to investigate dominance properties of their Bayes estimators, which will be studied in a future.

Appendix A.

We here provide the proofs of the theorems and the propositions based on the Stein–Haff identity due to Stein [21] and Haff [4], which is described below. For the Kronecker delta δ_{ij} and $\boldsymbol{S} = (s_{ij})$, let us define the differential operator $\mathrm{d}_{ij}^{\mathrm{S}}$ by $\mathrm{d}_{ij}^{\mathrm{S}} = 2^{-1}\left(1 + \delta_{ij}\right)\partial/\partial s_{ij}$, and denote $\mathcal{D}_{\mathrm{S}} = (\mathrm{d}_{ij}^{\mathrm{S}})$. Let $\boldsymbol{G}(\boldsymbol{S}) = (g_{ij}(\boldsymbol{S}))$ be a $p \times p$ matrix of absolutely continuous functions of $\boldsymbol{S}$, and define $\mathcal{D}_{\mathrm{S}}\boldsymbol{G}(\boldsymbol{S}) = ([\mathcal{D}_{\mathrm{S}}\boldsymbol{G}(\boldsymbol{S})]_{ij})$ by $[\mathcal{D}_{\mathrm{S}}\boldsymbol{G}(\boldsymbol{S})]_{ij} = \sum_{c=1}^{p} \mathrm{d}_{ic}^{\mathrm{S}} g_{cj}(\boldsymbol{S})$. Then the Stein–Haff identity is given by

$$E\left[\mathrm{tr}\,\boldsymbol{G}(\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right] = E\left[(n-p-1)\mathrm{tr}\,\boldsymbol{G}(\boldsymbol{S})\boldsymbol{S}^{-1} + 2\mathrm{tr}\,[\mathcal{D}_{\mathrm{S}}\boldsymbol{G}(\boldsymbol{S})]\right]. \tag{A.1}$$

The notations d_{ij}^W and $\mathcal{D}_W$ are similarly defined for the statistic $\boldsymbol{W}$. For $\boldsymbol{Q}$ and $\boldsymbol{\Lambda}$ defined by (2.5), the following calculus is very helpful.

$$\operatorname{tr}\mathcal{D}_S\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{Q}^{\mathrm{t}}=\sum_{i=1}^{p}\left\{p\psi_i-\lambda_i\frac{\partial\psi_i}{\partial\lambda_i}-\sum_{j>i}\frac{\lambda_i\psi_i-\lambda_j\psi_j}{\lambda_i-\lambda_j}\right\}, \tag{A.2}$$

$$\operatorname{tr}\mathcal{D}_W\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{Q}^{\mathrm{t}}=\sum_{i=1}^{p}\left\{\frac{\partial\psi_i}{\partial\lambda_i}+\sum_{j>i}\frac{\psi_i-\psi_j}{\lambda_i-\lambda_j}\right\}, \tag{A.3}$$

$$\operatorname{tr}\mathcal{D}_W[\boldsymbol{U}\boldsymbol{T}]=\operatorname{tr}[\mathcal{D}_W\boldsymbol{U}]\boldsymbol{T}+\operatorname{tr}\boldsymbol{U}^{\mathrm{t}}[\mathcal{D}_W\boldsymbol{T}^{\mathrm{t}}], \tag{A.4}$$

where $\boldsymbol{U}$ and $\boldsymbol{T}$ are $p\times p$ matrices of functions of $\boldsymbol{S}$. Eqs. (A.2) and (A.3) are from Loh [13] and Konno [9], and the Eq. (A.4) is from Haff [4].

Proof of Theorem 2.1. Since the estimators $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Phi})$ and $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Phi}]^{\mathrm{TR}})$ are scale-equivariant, we can assume that $\boldsymbol{\Sigma}=\boldsymbol{I}_p$ without any loss of generality. The risk difference $\Delta\equiv R(\boldsymbol{\Sigma},\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}))-R(\boldsymbol{\Sigma},\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{\mathrm{TR}}))$ is written as

$$\begin{aligned}\Delta&=E\left[\operatorname{tr}\boldsymbol{Q}\{\boldsymbol{\Psi}-[\boldsymbol{\Psi}]^{\mathrm{TR}}\}\boldsymbol{Q}^{\mathrm{t}}\right]+E\left[\operatorname{tr}\{\boldsymbol{I}_p-\boldsymbol{\Psi}\{[\boldsymbol{\Psi}]^{\mathrm{TR}}\}^{-1}\right]\\&\quad+E\left[\operatorname{tr}\boldsymbol{\Psi}\{[\boldsymbol{\Psi}]^{\mathrm{TR}}\}^{-1}-\log|\boldsymbol{\Psi}\{[\boldsymbol{\Psi}]^{\mathrm{TR}}\}^{-1}|-p\right]\\&=I_1+I_2+I_3.\end{aligned} \tag{A.5}$$

From the nonnegativeness of the loss function, it follows that $I_3\geqslant 0$. We shall show that $I_1+I_2\geqslant 0$.

To evaluate the first term I_1, we need to write the expectation $E[\operatorname{tr}\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{Q}^{\mathrm{t}}]$ in the integral form as

$$\begin{aligned}&E[\operatorname{tr}\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{Q}^{\mathrm{t}}]\\&\quad=c_0(\boldsymbol{\theta})\int\!\!\int[\operatorname{tr}\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{Q}^{\mathrm{t}}]|\boldsymbol{S}|^{(n-p-1)/2}\exp\left\{-\operatorname{tr}(\boldsymbol{S}+\boldsymbol{X}\boldsymbol{X}^{\mathrm{t}}-2\boldsymbol{X}\boldsymbol{\theta}^{\mathrm{t}})/2\right\}\,\mathrm{d}\boldsymbol{X}\,\mathrm{d}\boldsymbol{S}\\&\quad=c_0(\boldsymbol{\theta})\int\!\!\int[\operatorname{tr}\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{Q}^{\mathrm{t}}]|\boldsymbol{S}|^{(n-p-1)/2}\exp\left\{-\operatorname{tr}(\boldsymbol{S}+\boldsymbol{X}\boldsymbol{X}^{\mathrm{t}})/2\right\}\\&\qquad\times\int_{O(m)}\exp\left\{\operatorname{tr}\boldsymbol{X}\boldsymbol{O}\boldsymbol{\theta}^{\mathrm{t}}/2\right\}\,\mathrm{d}\mu_{(m)}(\boldsymbol{O})\,\mathrm{d}\boldsymbol{X}\,\mathrm{d}\boldsymbol{S},\end{aligned} \tag{A.6}$$

where $\mathrm{d}\mu_{(m)}(\boldsymbol{O})$ denotes an invariant probability measure on the group $O(m)$ of $m\times m$ orthogonal matrices, and $c_0(\boldsymbol{\theta})$ is the normalizing function. The second equality in (A.6) can be seen from the fact that $\boldsymbol{X}\boldsymbol{X}^{\mathrm{t}}$ is invariant under the transformation $\boldsymbol{X}\to\boldsymbol{X}\boldsymbol{O}$ for any $m\times m$ orthogonal matrix $\boldsymbol{O}$. A basic property of zonal polynomials gives that

$$\int_{O(m)}\exp\left\{\operatorname{tr}\boldsymbol{X}\boldsymbol{O}\boldsymbol{\theta}^{\mathrm{t}}/2\right\}\,\mathrm{d}\mu_{(m)}(\boldsymbol{O})=\sum_{\kappa}\alpha_{\kappa}^{(m)}C_{\kappa}\left(\boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{t}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{t}}\right),$$

where $\alpha_{\kappa}^{(m)}$ is given in James [7] and $C_{\kappa}(\boldsymbol{Z})$ denotes the normalized zonal polynomials of a positive definite matrix $\boldsymbol{Z}$ of order p corresponding to partitions $\kappa=\{\kappa_1,\ldots,\kappa_p\}$ so that $(\operatorname{tr}\boldsymbol{Z})^k=\sum_{\{\kappa:\kappa_1+\cdots+\kappa_p=k\}}C_{\kappa}(\boldsymbol{Z})$ for all $k=0,1,2,\ldots$. Making the transformations $\boldsymbol{S}\to\boldsymbol{O}\boldsymbol{S}\boldsymbol{O}^{\mathrm{t}}$ and $\boldsymbol{W}\to\boldsymbol{O}\boldsymbol{W}\boldsymbol{O}^{\mathrm{t}}$ for a $p\times p$ orthogonal matrix $\boldsymbol{O}$, we see that the last expression in (A.6) is rewritten by

$$c_0(\boldsymbol{\theta}) \int \int [\mathrm{tr}\, \boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{Q}^{\mathrm{t}}] |\boldsymbol{S}|^{(n-p-1)/2} \exp\left\{-\mathrm{tr}\,(\boldsymbol{S}+\boldsymbol{X}\boldsymbol{X}^{\mathrm{t}})/2\right\} \\ \times \sum_{\kappa} \alpha_{\kappa}^{(m)} \int_{O(p)} C_{\kappa}(\boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{t}}\boldsymbol{O}\boldsymbol{X}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{O}^{\mathrm{t}})\, \mathrm{d}\mu_{(p)}(\boldsymbol{O})\, \mathrm{d}\boldsymbol{S}\, \mathrm{d}\boldsymbol{X}. \tag{A.7}$$

By using the property of zonal polynomials given by

$$\int_{O(p)} C_{\kappa}(\boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{t}}\boldsymbol{O}\boldsymbol{X}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{O}^{\mathrm{t}})\, \mathrm{d}\mu_{(p)}(\boldsymbol{O}) = C_{\kappa}(\boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{t}}) C_{\kappa}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{t}}) / C_{\kappa}(\boldsymbol{I}_p),$$

the quantity in (A.7) is expressed by

$$c_1(\boldsymbol{\theta}) \int \int [\mathrm{tr}\, \boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{Q}^{\mathrm{t}}] \sum_{\kappa} \alpha_{\kappa}^{(m)} \frac{C_{\kappa}(\boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{t}}) C_{\kappa}(\boldsymbol{W})}{C_{\kappa}(\boldsymbol{I}_p)} \\ \times |\boldsymbol{S}|^{(n-p-1)/2} |\boldsymbol{W}|^{(m-p-1)/2} \exp\{-\mathrm{tr}\,(\boldsymbol{S}+\boldsymbol{W})/2\}\, \mathrm{d}\boldsymbol{S}\, \mathrm{d}\boldsymbol{W}, \tag{A.8}$$

where $c_1(\boldsymbol{\theta})$ is the normalizing function. When $\boldsymbol{S}$ and $\boldsymbol{W}$ are mutually independently distributed as $\boldsymbol{S} \sim \mathcal{W}_p(n, \boldsymbol{I})$ and $\boldsymbol{W} \sim \mathcal{W}_p(m, \boldsymbol{I})$, the joint density of $(\boldsymbol{S}, \boldsymbol{W})$ is given by

$$f_*(\boldsymbol{S}, \boldsymbol{W}) = c_* |\boldsymbol{S}|^{(n-p-1)/2} |\boldsymbol{W}|^{(m-p-1)/2} \exp\{-\mathrm{tr}\,(\boldsymbol{S}+\boldsymbol{W})/2\}$$

for the normalizing constant c_*. Let the notation $E_*[\cdot]$ denote an expectation with respect to the density $f_*(\boldsymbol{S}, \boldsymbol{W})$, and we have the expression that

$$E[\mathrm{tr}\, \boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{Q}^{\mathrm{t}}] = E_*[\mathrm{tr}\, \boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{Q}^{\mathrm{t}} \cdot c(\boldsymbol{W})],$$

where

$$c(\boldsymbol{W}) = \frac{c_1(\boldsymbol{\theta})}{c_*} \sum_{\kappa} \alpha_{\kappa}^{(m)} \frac{C_{\kappa}(\boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{t}}) C_{\kappa}(\boldsymbol{W})}{C_{\kappa}(\boldsymbol{I}_p)}.$$

Letting $\boldsymbol{G} = \mathrm{diag}\,(g_1, \ldots, g_p) = \{\boldsymbol{\Psi} - [\boldsymbol{\Psi}]^{\mathrm{TR}}\}(\boldsymbol{I}+\boldsymbol{\Lambda})^{-1}$ and using the notation $E_*[\cdot]$, we can rewrite I_1 as

$$I_1 = E_*[\mathrm{tr}\, \boldsymbol{Q}\boldsymbol{G}\boldsymbol{Q}^{\mathrm{t}} c(\boldsymbol{W})] + E_*[\mathrm{tr}\, \boldsymbol{Q}\boldsymbol{G}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{t}} c(\boldsymbol{W})] = I_{11} + I_{12}.$$

The Stein–Haff identity (A.1) with respect to $\boldsymbol{S}$ and Eq. (A.2) are useful for evaluating I_{11} as

$$I_{11} = E_*\left[(n-p-1)\mathrm{tr}\, \boldsymbol{Q}\boldsymbol{G}\boldsymbol{Q}^{\mathrm{t}} c(\boldsymbol{W})\boldsymbol{S}^{-1} + 2\,\mathrm{tr}\, \mathcal{D}_{\mathrm{S}}[\boldsymbol{Q}\boldsymbol{G}\boldsymbol{Q}^{\mathrm{t}} c(\boldsymbol{W})]\right] \\ = \sum_{i=1}^{p} E_*\left[(n-p-1) g_i c(\boldsymbol{W}) + 2\left\{p g_i - \lambda_i \frac{\partial g_i}{\partial \lambda_i} - \sum_{j>i} \frac{\lambda_i g_i - \lambda_j g_j}{\lambda_i - \lambda_j}\right\} c(\boldsymbol{W})\right]. \tag{A.9}$$

On the other hand, applying the Stein–Haff identity with respect to $\boldsymbol{W}$ with the density $f_*(\boldsymbol{S}, \boldsymbol{W})$ gives that

$$I_{12} = E^*[(m-p-1)\,\mathrm{tr}\, \boldsymbol{Q}\boldsymbol{G}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{t}} c(\boldsymbol{W})\boldsymbol{W} + 2\,\mathrm{tr}\, \mathcal{D}_W[\boldsymbol{Q}\boldsymbol{G}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{t}} c(\boldsymbol{W})]].$$

From (A.4), it is seen that

$$\mathrm{tr}\, \mathcal{D}_W[\boldsymbol{Q}\boldsymbol{G}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{t}} c(W)] = \{\mathrm{tr}\, \mathcal{D}_W[\boldsymbol{Q}\boldsymbol{G}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{t}}]\} c(\boldsymbol{W}) + \mathrm{tr}\, \boldsymbol{Q}\boldsymbol{G}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{t}} \mathcal{D}_W[c(\boldsymbol{W})\boldsymbol{I}_p],$$

so that from (A.3), I_{12} is evaluated as

$$I_{12} = \sum_{i=1}^{p} E^* \left[(m-p-1) g_i c(\boldsymbol{W}) + 2 \left\{ \frac{\partial(\lambda_i g_i)}{\partial \lambda_i} + \sum_{j>i} \frac{\lambda_i g_i - \lambda_j g_j}{\lambda_i - \lambda_j} \right\} c(\boldsymbol{W}) \right] \\ +2E^* \left[\operatorname{tr} \boldsymbol{QG\Lambda Q}^{\mathrm{t}} \mathcal{D}_W [c(\boldsymbol{W}) \boldsymbol{I}_p] \right]. \tag{A.10}$$

Combining (A.9) and (A.10), we observe that if we can prove the inequality

$$\operatorname{tr} \boldsymbol{QG\Lambda Q}^{\mathrm{t}} \mathcal{D}_W [c(\boldsymbol{W}) \boldsymbol{I}_p] \geqslant 0, \tag{A.11}$$

then I_1 can be evaluated as

$$I_1 \geqslant E^*[(n+m) \operatorname{tr} \boldsymbol{G} c(\boldsymbol{W})] = E[(n+m) \operatorname{tr} \{\boldsymbol{\Psi} - [\boldsymbol{\Psi}]^{\mathrm{TR}}\} (\boldsymbol{I} + \boldsymbol{\Lambda})^{-1}]. \tag{A.12}$$

Combing (A.5) and (A.12), we get the following inequality for the quantity $I_1 + I_2$:

$$I_1 + I_2 \geqslant \sum_{i=1}^{p} E\left[\left\{ \psi_i - \min\left(\psi_i, \frac{1+\lambda_i}{n+m} \right) \right\} \left\{ \frac{n+m}{1+\lambda_i} + \frac{1}{\min(\psi_i, (1+\lambda_i)/(n+m))} \right\} \right],$$

which can be seen to be zero. Since $I_3 \geqslant 0$, we can conclude that $\Delta = I_1 + I_2 + I_3 \geqslant 0$ for any ω if the inequality (A.11) is verified.

To complete the proof, we shall prove the inequality (A.11). Since the zonal polynomials $C_\kappa(\boldsymbol{W})$ are polynomials of the eigenvalues $\ell_1, \ldots, \ell_p$ ($\ell_1 \geqslant \cdots \geqslant \ell_p$) of $\boldsymbol{W}$ with nonnegative coefficients, we can put $c^*(\boldsymbol{L}) = c(\boldsymbol{W})$ for $\boldsymbol{L} = \operatorname{diag}(\ell_1, \ldots, \ell_p)$, and note that

$$\frac{\partial}{\partial \ell_i} c^*(\boldsymbol{L}) \equiv c_i^*(\boldsymbol{L}) (= c_i^*) \geqslant 0.$$

Let $\boldsymbol{H} = (h_{ij})$ be a $p \times p$ orthogonal matrix such that $\boldsymbol{W} = \boldsymbol{HLH}^{\mathrm{t}}$. From the calculus given by Konno [9], it can be seen that

$$\mathrm{d}_{ij}^{W} \ell_s = h_{is} h_{js}. \tag{A.13}$$

Using Eq. (A.13), we get that

$$\mathrm{d}_{ij}^{W} c^*(\boldsymbol{L}) = \sum_{s=1}^{p} (\mathrm{d}_{ij}^{W} \ell_s) \frac{\partial}{\partial \ell_s} c^*(\boldsymbol{L}) = \sum_{s=1}^{p} c_s^* h_{is} h_{js},$$

which implies that $\mathcal{D}_W[c(\boldsymbol{W})\boldsymbol{I}_p] = \boldsymbol{H} \operatorname{diag}(c_1^*, \ldots, c_p^*) \boldsymbol{H}^{\mathrm{t}}$. Hence, the l.h.s. of (A.11) is expressed by

$$\operatorname{tr} \boldsymbol{QG\Lambda Q}^{\mathrm{t}} \mathcal{D}_W [c(\boldsymbol{W}) \boldsymbol{I}_p] = \operatorname{tr} \boldsymbol{QG\Lambda Q}^{\mathrm{t}} \boldsymbol{H} \operatorname{diag}(c_1^*, \ldots, c_p^*) \boldsymbol{H}^{\mathrm{t}} \\ = \sum_{i=1}^{p} \sum_{j=1}^{p} \gamma_{ij}^2 g_i \lambda_i c_i^*,$$

which is nonnegative, where $(\gamma_{ij}) = \boldsymbol{H}^{\mathrm{t}} \boldsymbol{Q}$. Thus, the proof of Theorem 2.1 is complete. □

Proposition 2.1 can be easily proved by using the Stein–Haff identity (A.1) and Eq. (A.2) although the details are omitted here.

We next prove Theorem 2.2. Let ∇_X be the differential operator $\nabla_X = (\partial/\partial X_{ij})$. We shall use the Stein identity due to Stein [20], given by

$$E[\mathrm{tr}\,(\boldsymbol{X}-\boldsymbol{\theta})\boldsymbol{G}(\boldsymbol{X})] = E[\mathrm{tr}\,\boldsymbol{\Sigma}\nabla_X \boldsymbol{G}(\boldsymbol{X})] \tag{A.14}$$

for an $m \times p$ matrix $\boldsymbol{G}(\boldsymbol{X})$ of functions of $\boldsymbol{X}$. For $\boldsymbol{F} = (f_{ij}) = \boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^{\mathrm{t}}$, let $\mathcal{D}_F = (\mathrm{d}_{ij}^F)$ where $\mathrm{d}_{ij}^F = 2^{-1}(1+\delta_{ij})\partial/\partial f_{ij}$. The following calculus due to Konno [9] is also very helpful:

$$\nabla_X \boldsymbol{U} = 2\boldsymbol{S}^{-1}\boldsymbol{X}\mathcal{D}_F\boldsymbol{U}, \tag{A.15}$$

$$\mathrm{tr}\,\nabla_X(\boldsymbol{U}\boldsymbol{X}^{\mathrm{t}}) = p\,\mathrm{tr}\,\boldsymbol{U} + \mathrm{tr}\,\boldsymbol{X}^{\mathrm{t}}\nabla_X\boldsymbol{U}, \tag{A.16}$$

$$\boldsymbol{X}^{\mathrm{t}}\mathcal{D}_{\mathrm{S}}(\boldsymbol{X}\boldsymbol{T}) = -\boldsymbol{F}\mathcal{D}_F(\boldsymbol{F}\boldsymbol{T}) + 2^{-1}(m+1)\boldsymbol{F}\boldsymbol{T}, \tag{A.17}$$

for $m \times m$ matrices $\boldsymbol{T}$ and $\boldsymbol{U}$ which are functions of $\boldsymbol{F}$.

Proof of Theorem 2.2. Since the estimators $\widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{\Phi})$ and $\widehat{\boldsymbol{\Sigma}}(a, [\boldsymbol{\Phi}]^{\mathrm{TR}})$ are scale-equivariant, we can assume that $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ without any loss of generality. The risk difference $\Delta \equiv R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{\Psi})) - R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}(a, [\boldsymbol{\Phi}]^{\mathrm{TR}}))$ is written as

$$\begin{aligned}\Delta &= E\left[\mathrm{tr}\,\boldsymbol{X}\boldsymbol{R}\{\boldsymbol{\Psi} - [\boldsymbol{\Psi}]^{\mathrm{TR}}\}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}\right] + E\left[\mathrm{tr}\,\{\boldsymbol{I}_m - (a\boldsymbol{I}_m + \boldsymbol{\Lambda}\boldsymbol{\Psi})(a\boldsymbol{I}_m + \boldsymbol{\Lambda}[\boldsymbol{\Phi}]^{\mathrm{TR}})^{-1}\}\right] \\ &\quad + E\left[\mathrm{tr}\,\boldsymbol{B}(\boldsymbol{\Lambda}) - \log|\boldsymbol{B}(\boldsymbol{\Lambda})| - m\right] \\ &= I_1 + I_2 + I_3,\end{aligned} \tag{A.18}$$

for $\boldsymbol{B}(\boldsymbol{\Lambda}) = (a\boldsymbol{I}_m + \boldsymbol{\Lambda}\boldsymbol{\Psi})(a\boldsymbol{I}_m + \boldsymbol{\Lambda}[\boldsymbol{\Phi}]^{\mathrm{TR}})^{-1}$. From the nonnegativeness of the loss function, it follows that $I_3 \geqslant 0$. We shall show that $I_1 + I_2 \geqslant 0$.

To evaluate the first term I_1, we can use the same arguments as given between (A.6) and (A.8) in the proof of Theorem 2.1 and write I_1 as

$$\begin{aligned}I_1 &= c_0(\boldsymbol{\theta}) \int\int \left[\mathrm{tr}\,\boldsymbol{X}\boldsymbol{R}\{\boldsymbol{\Psi} - [\boldsymbol{\Psi}]^{\mathrm{TR}}\}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}\right] \sum_\kappa \alpha_\kappa^{(p)} \frac{C_\kappa(\boldsymbol{\theta}^{\mathrm{t}}\boldsymbol{\theta})C_\kappa(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})}{C_\kappa(\boldsymbol{I}_m)} \\ &\quad \times |\boldsymbol{S}|^{(n-p-1)/2} \exp\left\{-\mathrm{tr}\,(\boldsymbol{S} + \boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})/2\right\}\,\mathrm{d}\boldsymbol{S}\,\mathrm{d}\boldsymbol{X}.\end{aligned}$$

When $\boldsymbol{S}$ and $\boldsymbol{X}$ are mutually independently distributed as $\boldsymbol{S} \sim \mathcal{W}_p(n, \boldsymbol{I})$ and $\boldsymbol{X} \sim \mathcal{N}_{p,m}(\boldsymbol{0}, \boldsymbol{I}_p, \boldsymbol{I}_m)$, the joint density of $(\boldsymbol{S}, \boldsymbol{X})$ is given by

$$f_*(\boldsymbol{S}, \boldsymbol{X}) = c_*|\boldsymbol{S}|^{(n-p-1)/2} \exp\left\{-\mathrm{tr}\,(\boldsymbol{S} + \boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})/2\right\}$$

for the normalizing constant c_*. Let the notation $E_*[\cdot]$ denote an expectation with respect to the density $f_*(\boldsymbol{S}, \boldsymbol{X})$, and we have the expression that

$$I_1 = E[\mathrm{tr}\,\boldsymbol{X}\boldsymbol{R}\{\boldsymbol{\Psi} - [\boldsymbol{\Psi}]^{\mathrm{TR}}\}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}] = E_*[\mathrm{tr}\,\boldsymbol{X}\boldsymbol{R}\{\boldsymbol{\Psi} - [\boldsymbol{\Psi}]^{\mathrm{TR}}\}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}} \cdot c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})],$$

where

$$c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}) = \frac{c_1(\boldsymbol{\theta})}{c_*} \sum_\kappa \alpha_\kappa^{(m)} \frac{C_\kappa(\boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{t}})C_\kappa(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})}{C_\kappa(\boldsymbol{I}_p)}.$$

Letting $\boldsymbol{G} = \mathrm{diag}\,(g_1, \ldots, g_p) = \{\boldsymbol{\Psi} - [\boldsymbol{\Psi}]^{\mathrm{TR}}\}(\boldsymbol{I} + \boldsymbol{\Lambda})^{-1}$ and using the notation $E_*[\cdot]$, we can rewrite I_1 as

$$I_1 = E_*[\mathrm{tr}\,\boldsymbol{X}\boldsymbol{R}\boldsymbol{G}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})] + E_*[\mathrm{tr}\,\boldsymbol{X}\boldsymbol{R}\boldsymbol{G}\boldsymbol{\Lambda}\boldsymbol{R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})] = I_{11} + I_{12}.$$

Using the Stein–Haff identity (A.1) and Eq. (A.17), we see that I_{11} can be rewritten by

$$\begin{aligned} I_{11} &= E_*\left[(n-p-1)\mathrm{tr}\,\boldsymbol{XRGR}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})\boldsymbol{S}^{-1} + 2\,\mathrm{tr}\,\mathcal{D}_{\mathrm{S}}[\boldsymbol{XRGR}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}]c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})\right] \\ &= E_*\left[(n+m-p)\mathrm{tr}\,\boldsymbol{G\Lambda}c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}) - 2\,\mathrm{tr}\,\boldsymbol{F}\mathcal{D}_F[\boldsymbol{FRGR}^{\mathrm{t}}]c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})\right]. \end{aligned} \tag{A.19}$$

For I_{12}, on the other hand, we use the Stein identity (A.14) for $\boldsymbol{X} \sim \mathcal{N}_{p,m}(\boldsymbol{0}, \boldsymbol{I}_p, \boldsymbol{I}_m)$ in the expectation $E_*[\cdot]$ and get that

$$I_{12} = E_*[p\,\mathrm{tr}\,\boldsymbol{RG\Lambda R}^{\mathrm{t}}c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})] + E_*[\mathrm{tr}\,\boldsymbol{X}^{\mathrm{t}}\nabla_X[\boldsymbol{RG\Lambda R}^{\mathrm{t}}c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})]], \tag{A.20}$$

where Eq. (A.16) is used at the last equality in (A.20). Using Eq. (A.15), we see that the term $\mathrm{tr}\,\boldsymbol{X}^{\mathrm{t}}\nabla_X[\boldsymbol{RG\Lambda R}^{\mathrm{t}}c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})]$ is rewritten as

$$2\,\mathrm{tr}\,\boldsymbol{F}(\mathcal{D}_F[\boldsymbol{RG\Lambda R}^{\mathrm{t}}])c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}) + \mathrm{tr}\,\boldsymbol{X}^{\mathrm{t}}(\nabla_X[c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})\boldsymbol{I}_m])\boldsymbol{RG\Lambda R}^{\mathrm{t}}. \tag{A.21}$$

Combining (A.19), (A.20) and (A.21) yields that

$$\begin{aligned} I_1 &= E_*\left[(n+m)\mathrm{tr}\,\boldsymbol{G\Lambda}c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})\right] + E_*\left[\mathrm{tr}\,\boldsymbol{X}^{\mathrm{t}}(\nabla_X[c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})\boldsymbol{I}_m])\boldsymbol{RG\Lambda R}^{\mathrm{t}}\right] \\ &\geqslant E[(n+m)\mathrm{tr}\,\boldsymbol{G\Lambda}], \end{aligned} \tag{A.22}$$

if we can show the inequality $\mathrm{tr}\,\boldsymbol{X}^{\mathrm{t}}(\nabla_X[c(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})\boldsymbol{I}_m])\boldsymbol{RG\Lambda R}^{\mathrm{t}} \geqslant 0$, which can be shown by using the same arguments as in the proof of (A.11). Combining (A.18) and (A.22), we observe that

$$I_1 + I_2 \geqslant \sum_{i=1}^{m} E\left[(\psi_i - \psi_i^{\mathrm{TR}})\left\{\frac{n+m}{1+\lambda_i} - \frac{1}{a+\lambda_i\psi_i^{\mathrm{TR}}}\right\}\lambda_i\right],$$

which is equal to zero as seen from the definition of ψ_i^{TR}. Therefore the proof of Theorem 2.2 is complete. □

Proof of Proposition 2.2. The risk function of the estimator $\widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{\Phi})$ is written by

$$\begin{aligned} R(\boldsymbol{\Sigma}, \widehat{\boldsymbol{\Sigma}}(a, \boldsymbol{\Psi})) = E&\left[\mathrm{tr}\,(a\boldsymbol{S} + \boldsymbol{XR\Psi R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}})\boldsymbol{\Sigma}^{-1} - p\log a\right] - p \\ &- E\left[\log|\boldsymbol{I}_p + a^{-1}\boldsymbol{S}^{-1/2}\boldsymbol{XR\Psi R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{S}^{-1/2}| + \log|\boldsymbol{S\Sigma}^{-1}|\right]. \end{aligned} \tag{A.23}$$

It is easy to see that $E[\mathrm{tr}\,a\boldsymbol{S\Sigma}^{-1}] = pna$ and that $|\boldsymbol{I}_p + a^{-1}\boldsymbol{S}^{-1/2}\boldsymbol{XR\Psi R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{S}^{-1/2}| = |\boldsymbol{I}_m + a^{-1}\boldsymbol{\Lambda\Psi}|$. Applying the Stein–Haff identity (A.1) and using Eq. (A.17), we observe that $E[\mathrm{tr}\,\boldsymbol{XR\Psi R}^{\mathrm{t}}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{\Sigma}^{-1}] = E\left[(n+m-p)\,\mathrm{tr}\,\boldsymbol{\Lambda\Psi} - 2\,\mathrm{tr}\,\boldsymbol{F}\mathcal{D}_F[\boldsymbol{R\Lambda\Psi R}^{\mathrm{t}}]\right]$. From Stein [4], it follows that $\mathcal{D}_F[\boldsymbol{R\Psi R}^{\mathrm{t}}] = \boldsymbol{R\Psi}^{(1)}\boldsymbol{R}^{\mathrm{t}}$ where $\boldsymbol{\Psi}^{(1)} = \mathrm{diag}\,(\psi_1^{(1)}, \ldots, \psi_m^{(1)})$ for $\psi_i^{(1)} = 2^{-1}\sum_{j\neq i}(\psi_i - \psi_j)/(\lambda_i - \lambda_j) + \partial\psi_i/\partial\lambda_i$. Hence we get the expression in Proposition 2.2. □

Acknowledgements

The authors are grateful to the two reviewers for their helpful and valuable comments. This research was supported in part by the grant from the National Science Council of Republic of China under Contract no. NSC93-2118-M-001-027, in part by grants from the Ministry of Education and Science, Japan, nos. 15200021, 15200022 and 16500172 and in part by a grant from the 21st Century COE Program at the Faculty of Economics in the University of Tokyo.

References

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

[2] J.O. Berger, W. Strawderman, D. Tang, Posterior propriety and admissibility of hyperpriors in normal hierarchical models, Ann. Statist. 33 (2005) 606–646.

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

[4] L.R. Haff, An identity for the Wishart distribution with applications, J. Multivariate Anal. 9 (1979) 531–542.

[5] L.R. Haff, Empirical Bayes estimation of the multivariate normal covariance matrix, Ann. Statist. 8 (1980) 586–597.

[6] H. Hara, Estimation of covariance matrix and mean squared error for shrinkage estimators in multivariate normal distribution, Doctoral Dissertation, Faculty of Engineering, University of Tokyo, 1999.

[7] A.T. James, Distribution of matrix variates and latent roots derived from normal samples, Ann. Math. Statist. 35 (1964) 475–501.

[8] W. James, C. Stein, Estimation with quadratic loss, in: Proceedings of Fourth Berkeley Symposium of Mathematical Statistics Probabability, vol. 1, University of California Press, Berkeley, 1961, pp. 361–379.

[9] Y. Konno, Improved estimation of matrix of normal mean and eigenvalues in the multivariate *F*-distribution, Doctoral Dissertation, Institute of Mathematics, University of Tsukuba, 1992.

[10] T. Kubokawa, Shrinkage and modification techniques in estimation of variance and the related problems: a review, Commun. Statist. Theory Methods 28 (1999) 613–650.

[11] T. Kubokawa, C. Robert, A.K.Md.E. Saleh, Empirical Bayes estimation of the variance parameter of a normal distribution with unknown mean under an entropy loss, Sankhya Ser. A 54 (1999) 402–410.

[12] T. Kubokawa, M.S. Srivastava, Estimating the covariance matrix: a new approach, J. Multivariate Anal. 86 (2003) 28–47.

[13] W.-L. Loh, Estimating covariance matrices, Ph.D. Thesis. Stanford University, 1988.

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

[15] F. Perron, Equivariant estimators of the covariance matrix, Canad. J. Statist. 18 (1990) 179–182.

[16] B.K. Sinha, M. Ghosh, Inadmissibility of the best equivariant estimators of the variance–covariance matrix, and the generalized variance under entropy loss, Statist. Decisions 5 (1987) 201–227.

[17] M.S. Srivastava, C.G. Khatri, An Introduction to Multivariate Statistics, North-Holland, New York, 1979.

[18] M.S. Srivastava, T. Kubokawa, Improved nonnegative estimation of multivariate components of variance, Ann. Statist. 27 (1999) 2008–2032.

[19] C. Stein, Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean, Ann. Inst. Statist. Math. 16 (1964) 155–160.

[20] C. Stein, Estimation of the mean of a multivariate normal distribution, in: Proceedings of the Prague Symposium on Asymptotic Statistics, 1973, pp. 345–381.

[21] C. Stein, Lectures on the theory of estimation of many parameters, in: I.A. Ibragimov, M.S. Nikulin (Eds.), Studies in the Statistical Theory of Estimation, Part I Proceedings of Scientific Seminars of the Steklov Institute, Leningrad Division, vol. 74, 1977, pp. 4–65. (in Russian).

[22] R. Yang, J.O. Berger, Estimation of a covariance matrix using the reference prior, Ann. Statist. 22 (1994) 1195–1211.