CIRJE-F-52

Estimating the Covariance Matrix: A New Approach

Tatsuya Kubokawa
University of Tokyo

M. S. Srivastava
University of Toronto

June 1999

Estimating the Covariance Matrix: A New Approach

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

In this paper, we consider the problem of estimating the covariance matrix and the generalized variance when the observations follow a nonsingular multivariate normal distribution with unknown mean. A new method is presented to obtain a truncated estimator that utilizes the information available in the sample mean matrix and dominates the James-Stein minimax estimator. Several scale equivariant minimax estimators are also given. This method is then applied to obtain new truncated and improved estimators of the generalized variance; it also provides a new proof to the results of Shorrock and Zidek (1976) and Sinha (1976).

Key words and phrases: Covariance matrix, generalized variance, minimax estimation, improvement, decision theory, Stein result, Bartlett's decomposition.

AMS subject classifications: Primary 62F11, 62J12, Secondary 62C15, 62C20.

Tatsuya Kubokawa
Faculty of Economics, University of Tokyo
Hongo, Bunkyo-ku, Tokyo 113, JAPAN
E-Mail: tatsuya@e.u-tokyo.ac.jp
Fax: +81-3-5841-5521
Telephone: +81-3-5841-5656(DI)

M. S. Srivastava
Department of Statistics, University of Toronto
100 St George Street, Toronto, Ontario, CANADA M5S 3G3
E-Mail: srivasta@utstat.utstat.toronto.edu
Fax: +1-416-978-5133
Telephone: +1-416-978-4450

1 Introduction

Consider the canonical form of the multivariate normal linear model in which the $p \times m$ random matrix $\boldsymbol{X}$ and the $p \times p$ random symmetric matrix $\boldsymbol{S}$ are independently distributed as $\mathcal{N}(\boldsymbol{\Xi}, \boldsymbol{\Sigma}, \boldsymbol{I}_m)$ and $\mathcal{W}_p(\boldsymbol{\Sigma}, n)$, respectively, where we follow the notation of Srivastava and Khatri (1979, p.54, 76). We shall assume that the covariance matrix $\boldsymbol{\Sigma}$ is positive definite (p.d.) and that the sample size $n \geq p$, and thus $\boldsymbol{S}$ is positive definite with probability one, see Stein (1969). In this paper, we consider the problem of estimating the covariance matrix $\boldsymbol{\Sigma}$ and the generalized variance $|\boldsymbol{\Sigma}|$, the determinant of the matrix $\boldsymbol{\Sigma}$ under the Stein loss function

$$L(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \operatorname{tr} \widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1} - |\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}| - p, \tag{1.1}$$

where $\widehat{\boldsymbol{\Sigma}}$ is the estimator of $\boldsymbol{\Sigma}$ and every estimator is evaluated in terms of the risk functions $R(\omega, \widehat{\boldsymbol{\Sigma}}) = E_\omega[L(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})]$, $\omega = (\boldsymbol{\Sigma}, \boldsymbol{\Xi})$.

Beginning with the work of James and Stein (1961), where they showed that the estimator

$$\widehat{\boldsymbol{\Sigma}}^{JS} = \boldsymbol{T}\boldsymbol{D}\boldsymbol{T}^t, \tag{1.2}$$

where $\boldsymbol{S} = \boldsymbol{T}\boldsymbol{T}^t$, $\boldsymbol{T}$ is a lower triangular matrix with diagonal elements (and hence unique), and

$$\boldsymbol{D} = \operatorname{diag}(d_1, \ldots, d_p), \quad d_i = (n + p + 1 - 2i)^{-1}, \quad i = 1, \ldots, p. \tag{1.3}$$

dominates the uniformly minimum unbiased estimator $\widehat{\boldsymbol{\Sigma}}^{UB} = n^{-1}\boldsymbol{S}$, many estimators have been proposed in the literature dominating $\widehat{\boldsymbol{\Sigma}}^{UB}$, see Stein (1977) and Haff (1979), among them, who developed what is now called Stein-Haff identity that led to a substantial development in this area, see Kubokawa (1998) for an extensive review.

The estimators mentioned above did not use the information available in the observation matrix $\boldsymbol{X}$ while Stein (1964) has shown in the univariate case, $p = 1$, that a truncated estimator that utilizes the information in the sample mean dominates the uniformly minimum variance unbiased estimators of the variance σ^2. Attempts in this direction utilizing the information contained in the sample mean were first made by Shorrock and Zidek (1976) and Sinha (1976) who provided minimax estimators for the generalized variance using the information available in the observation matrix $\boldsymbol{X}$.

The mathematical tools used in the above two papers to obtain these minimax estimators were, respectively, the use of zonal polynomials and Fubini-type theorem of Karlin (1960). Sarkar (1989, 1991) and Iliopoulos and Kourouklis (1999) used the above two mentioned approaches to obtain the confidence interval for the generalized variance $|\boldsymbol{\Sigma}|$. Sinha and Ghosh (1987) also provided a truncated estimator of the covariance matrix $\boldsymbol{\Sigma}$ utilizing the information contained in the observation matrix $\boldsymbol{X}$. This truncated estimator is given by

$$\widehat{\boldsymbol{\Sigma}}^{SG} = \begin{cases} (n+m)^{-1}(\boldsymbol{S} + \boldsymbol{X}\boldsymbol{X}^t) & \text{if } (n+m)^{-1}(\boldsymbol{S} + \boldsymbol{X}\boldsymbol{X}^t) \leq n^{-1}\boldsymbol{S} \\ n^{-1}\boldsymbol{S} & \text{otherwise,} \end{cases} \tag{1.4}$$

improving on the UMVU one $\widehat{\boldsymbol{\Sigma}}^{UB} = n^{-1}\boldsymbol{S}$ under the Stein loss, where $\boldsymbol{A} \geq \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is nonnegative definite. Based on the same technique, Hara (1999) recently showed that $\widehat{\boldsymbol{\Sigma}}^{SG}$ is dominated by

$$\widehat{\boldsymbol{\Sigma}}^{HR} = \boldsymbol{S}^{1/2}\boldsymbol{Q}\mathrm{diag}\,(\phi_1, \ldots, \phi_p)\boldsymbol{Q}^t\boldsymbol{S}^{1/2}$$

for

$$\phi_i = \begin{cases} \min\{n^{-1}, (n+m)^{-1}(1+\gamma_i)\} & \text{if } \gamma_i > 0 \\ n^{-1} & \text{if } \gamma_i = 0, \end{cases}$$

where $\boldsymbol{Q}$ is an orthogonal matrix such that $\boldsymbol{Q}^t\boldsymbol{S}^{-1/2}\boldsymbol{X}\boldsymbol{X}^t\boldsymbol{S}^{-1/2}\boldsymbol{Q} = \mathrm{diag}\,(\gamma_1, \ldots, \gamma_p)$. When the rank of $\boldsymbol{X}$ is one, namely $m = 1$, the risk functions can be easily handled and several further observations have been given by Kubokawa *et al.* (1992, 93) and Perron (1990). Especially, Kubokawa *et al.* (1992) derived an empirical Bayes estimator improving on the Sinha-Ghosh estimator $\widehat{\boldsymbol{\Sigma}}^{SG}$. However, none of these estimators were shown to dominate the initial James-Stein minimax estimator $\widehat{\boldsymbol{\Sigma}}^{JS}$. Thus, our aim is to obtain an estimator that dominates $\widehat{\boldsymbol{\Sigma}}^{JS}$ as well as in which, as suggested by the above estimators $\widehat{\boldsymbol{\Sigma}}^{SG}$ and $\widehat{\boldsymbol{\Sigma}}^{HR}$, the coefficients $(n+p+1-2i)^{-1}$ should be changed to $(n+m+p+1-2i)^{-1}$ when we utilize both $\boldsymbol{S}$ and $\boldsymbol{X}$ in estimation of $\boldsymbol{\Sigma}$. In Section 2, we develop a new type of estimator with such a natural analogy. For this purpose, we introduce a new method for the improvement. This method can be also applied in Section 3 not only to construct a new form of an improved estimator of $|\boldsymbol{\Sigma}|$ but also to give another proof of the result of Shorrock and Zidek (1976) and Sinha (1976). When $\boldsymbol{X}$ has full rank, namely, $m \geq p$, another type of minimax improved estimator motivated by Srivastava and Kubokawa (1999) are provided in Section 2, and the improvements on any scale equivariant estimator are shown. Monte Carlo simulations are carried out in Section 4 to compare risk behaviors of the proposed estimators.

2 Estimation of the Covariance Matrix

2.1 Improvements on the James-Stein minimax estimator

Consider the problem of estimating the covariance matrix $\boldsymbol{\Sigma}$ based on $(\boldsymbol{S}, \boldsymbol{X})$ relative to the Stein loss function Every estimator is evaluated in terms of the risk function $R(\omega, \widehat{\boldsymbol{\Sigma}}) = E_\omega[L(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})]$, where $\omega = (\boldsymbol{\Sigma}, \boldsymbol{\Xi})$.

Let G_T^+ be the triangular group consisting of $p \times p$ lower triangular matrices with positive diagonal elements. Let $\boldsymbol{T} = (t_{ij}) \in G_T^+$ such that $\boldsymbol{S} = \boldsymbol{T}\boldsymbol{T}^t$. For constructing an estimator improving on the James-Stein minimax estimator (1.2), define an $m \times p$ matrix $\boldsymbol{Y}$ by

$$\boldsymbol{Y} = (y_{ij}) \quad = \quad \left(\boldsymbol{T}^{-1}\boldsymbol{X}\right)^t$$

$$\begin{aligned} &= (\boldsymbol{y}_1, \dots, \boldsymbol{y}_p) \\ &= (\boldsymbol{y}_1, \dots, \boldsymbol{y}_{j-1}, \boldsymbol{Y}_j), \end{aligned}$$

for $\boldsymbol{Y}_j = (\boldsymbol{y}_j, \dots, \boldsymbol{y}_p)$ and $j = 2, \dots, p$. Also for $j = 1, \dots, p$, define $m \times m$ matrix $\boldsymbol{C}_j$ inductively by

$$\begin{aligned} \boldsymbol{C}_j &= \boldsymbol{C}_j(\boldsymbol{y}_1, \dots, \boldsymbol{y}_{j-1}) \\ &= \boldsymbol{C}_{j-1} - (1 + \boldsymbol{y}_{j-1}^t \boldsymbol{C}_{j-1} \boldsymbol{y}_{j-1})^{-1} \boldsymbol{C}_{j-1} \boldsymbol{y}_{j-1} \boldsymbol{y}_{j-1}^t \boldsymbol{C}_{j-1} \end{aligned} \tag{2.1}$$

where $\boldsymbol{C}_1 = \boldsymbol{I}_m$. Then we can see that

$$\begin{aligned} |\boldsymbol{I}_p + \boldsymbol{Y}^t \boldsymbol{Y}| &= \begin{vmatrix} 1 + \boldsymbol{y}_1^t \boldsymbol{y}_1 & \boldsymbol{y}_1^t \boldsymbol{Y}_2 \\ \boldsymbol{Y}_2^t \boldsymbol{y}_1 & \boldsymbol{I}_{p-1} + \boldsymbol{Y}_2^t \boldsymbol{Y}_2 \end{vmatrix} \\ &= (1 + \boldsymbol{y}_1^t \boldsymbol{y}_1) \left| \boldsymbol{I}_{p-1} + \boldsymbol{Y}_2^t \boldsymbol{C}_2 \boldsymbol{Y}_2 \right| \\ &= (1 + \boldsymbol{y}_1^t \boldsymbol{y}_1) \begin{vmatrix} 1 + \boldsymbol{y}_2^t \boldsymbol{C}_2 \boldsymbol{y}_2 & \boldsymbol{y}_2^t \boldsymbol{C}_2 \boldsymbol{Y}_3 \\ \boldsymbol{Y}_3^t \boldsymbol{C}_2 \boldsymbol{y}_2 & \boldsymbol{I}_{p-2} + \boldsymbol{Y}_3^t \boldsymbol{C}_2 \boldsymbol{Y}_3 \end{vmatrix} \\ &= (1 + \boldsymbol{y}_1^t \boldsymbol{y}_1)(1 + \boldsymbol{y}_2^t \boldsymbol{C}_2 \boldsymbol{y}_2) \left| \boldsymbol{I}_{p-2} + \boldsymbol{Y}_3^t \boldsymbol{C}_3 \boldsymbol{Y}_3 \right| \\ &= \prod_{i=1}^p (1 + \boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{y}_i). \end{aligned} \tag{2.2}$$

Using the statistics $\boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{y}_i$'s, we want to propose a new estimator given by

$$\widehat{\boldsymbol{\Sigma}}^{TR} = \boldsymbol{T} \boldsymbol{G} \boldsymbol{T}^t, \tag{2.3}$$

where $\boldsymbol{G} = \mathrm{diag}\,(g_1, \dots, g_p)$ for

$$g_i = \min \left\{ \frac{1}{n + p + 1 - 2i}, \ \frac{1 + \boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{y}_i}{n + m + p + 1 - 2i} \right\}.$$

Theorem 1. *The truncated estimator $\widehat{\boldsymbol{\Sigma}}^{TR}$ dominates the James-Stein minimax estimator $\widehat{\boldsymbol{\Sigma}}^{JS}$ relative to the Stein loss (1.1).*

Proof. For sake of convenience, let

$$\begin{aligned} \boldsymbol{t}_{j,j-1} &= (t_{j,j-1}, \dots, t_{p,j-1})^t, \\ \boldsymbol{T}_{jj} &= \begin{pmatrix} t_{jj} & & & \text{\Large 0} \\ t_{j+1,j} & t_{j+1,j+1} & & \\ \vdots & \vdots & \ddots & \\ t_{pj} & t_{p,j+1} & \cdots & t_{pp} \end{pmatrix} \end{aligned}$$

for $j = 2, \ldots, p$. $\boldsymbol{T}_{11}$ corresponds to $\boldsymbol{T}$. By making the transformation, it is supposed that $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ without loss of generality. The risk difference of the two estimators is expressed as

$$\begin{aligned} & R(\omega; \widehat{\boldsymbol{\Sigma}}^{JS}) - R(\omega; \widehat{\boldsymbol{\Sigma}}^{TR}) \\ = \;& E\left[\mathrm{tr}\,(\boldsymbol{D} - \boldsymbol{G})\boldsymbol{T}^t\boldsymbol{T} - \log|\boldsymbol{D}\boldsymbol{G}^{-1}|\right] \\ = \;& \sum_{i=1}^{p} \Delta_i, \end{aligned}$$

where

$$\Delta_i = E\left[\left\{(d_i - d_i^* a_{ii})(t_{ii}^2 + \boldsymbol{t}_{i+1,i}^t \boldsymbol{t}_{i+1,i}) - \log d_i/(d_i^* a_{ii})\right\} I(d_i \geq a_{ii})\right], \tag{2.4}$$

for

$$\begin{aligned} a_{ii} &= 1 + \boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{y}_i, \\ d_i^* &= (n + m + p + 1 - 2i)^{-1}. \end{aligned}$$

We shall show that $\Delta_i \geq 0$ for $i = 1, \ldots, p$. For this purpose, we write the joint density function of $(\boldsymbol{T}, \boldsymbol{Y})$ as

$$c_0(\boldsymbol{\Xi}) \prod_{i=1}^{p} t_{ii}^{n+m-i} \mathrm{etr}\left[-2^{-1}\left\{\boldsymbol{T}(\boldsymbol{I}_p + \boldsymbol{Y}^t\boldsymbol{Y})\boldsymbol{T}^t - 2\boldsymbol{T}\boldsymbol{Y}^t\boldsymbol{\Xi}^t\right\}\right], \tag{2.5}$$

which is given by making the transformations $\boldsymbol{S} \to \boldsymbol{T}\boldsymbol{T}^t$ and $\boldsymbol{X} \to \boldsymbol{Y}^t = \boldsymbol{T}^{-1}\boldsymbol{X}$ with the Jacobians $2^p \prod_{i=1}^p t_{ii}^{p-i+1}$ and $|\boldsymbol{T}|^m$, where $c_0(\boldsymbol{\Xi})$ is a normalizing function. Let us decompose $\boldsymbol{I}_p + \boldsymbol{Y}^t\boldsymbol{Y}$ and $\boldsymbol{Y}^t\boldsymbol{\Xi}^t$ as

$$\begin{aligned} \boldsymbol{I}_p + \boldsymbol{Y}^t\boldsymbol{Y} &= \boldsymbol{I}_p + \begin{pmatrix} \boldsymbol{y}_1^t \\ \boldsymbol{Y}_2^t \end{pmatrix} (\boldsymbol{y}_1, \boldsymbol{Y}_2) \\ &= \begin{pmatrix} a_{11} & \boldsymbol{a}_{21}^t \\ \boldsymbol{a}_{21} & \boldsymbol{A}_{22} \end{pmatrix}, \\ \boldsymbol{Y}^t\boldsymbol{\Xi}^t &= \begin{pmatrix} \boldsymbol{y}_1^t \\ \boldsymbol{Y}_2^t \end{pmatrix} (\boldsymbol{\xi}_1, \boldsymbol{\Xi}_2) \\ &= \begin{pmatrix} \theta_{11} & \boldsymbol{\theta}_{12} \\ \boldsymbol{\theta}_{21} & \boldsymbol{\Theta}_{22} \end{pmatrix}, \end{aligned}$$

where $a_{11} = 1 + \boldsymbol{y}_1^t\boldsymbol{y}_1$, $\boldsymbol{a}_{21} = \boldsymbol{Y}_2^t\boldsymbol{y}_1$, $\boldsymbol{A}_{22} = \boldsymbol{I}_p + \boldsymbol{Y}_2^t\boldsymbol{Y}_2$, $\theta_{11} = \boldsymbol{y}_1^t\boldsymbol{\xi}_1$, $\boldsymbol{\theta}_{12} = \boldsymbol{y}_1^t\boldsymbol{\Xi}_2$, $\boldsymbol{\theta}_{21} = \boldsymbol{Y}_2^t\boldsymbol{\xi}_1$ and $\boldsymbol{\Theta}_{22} = \boldsymbol{Y}_2^t\boldsymbol{\Xi}_2$ for $\boldsymbol{\xi}_1$ being the first column vector of $\boldsymbol{\Xi}^t$. Then we have

$$\begin{aligned} & \mathrm{tr}\left\{\boldsymbol{T}(\boldsymbol{I}_p + \boldsymbol{Y}^t\boldsymbol{Y})\boldsymbol{T}^t - 2\boldsymbol{T}\boldsymbol{Y}^t\boldsymbol{\Xi}^t\right\} \\ = \;& \mathrm{tr}\left\{\begin{pmatrix} t_{11} & \boldsymbol{0} \\ \boldsymbol{t}_{21} & \boldsymbol{T}_{22} \end{pmatrix} \begin{pmatrix} a_{11} & \boldsymbol{a}_{21}^t \\ \boldsymbol{a}_{21} & \boldsymbol{A}_{22} \end{pmatrix} \begin{pmatrix} t_{11} & \boldsymbol{t}_{21}^t \\ 0 & \boldsymbol{T}_{22}^t \end{pmatrix}\right. \end{aligned}$$

$$
\begin{aligned}
& \qquad -2\begin{pmatrix} t_{11} & \mathbf{0} \\ \boldsymbol{t}_{21} & \boldsymbol{T}_{22} \end{pmatrix}\begin{pmatrix} \theta_{11} & \boldsymbol{\theta}_{12} \\ \boldsymbol{\theta}_{21} & \boldsymbol{\Theta}_{22} \end{pmatrix}\Bigg\} \\
= & \ \Big(a_{11}t_{11}^2 - 2\theta_{11}t_{11}\Big) + \Big(a_{11}\boldsymbol{t}_{21}^t\boldsymbol{t}_{21}^t + 2\boldsymbol{t}_{21}^t(\boldsymbol{T}_{22}\boldsymbol{a}_{21} - \boldsymbol{\theta}_{12}^t)\Big) \\
& + \Big(\mathrm{tr}\,\boldsymbol{T}_{22}\boldsymbol{A}_{22}\boldsymbol{T}_{22}^t - 2\mathrm{tr}\,\boldsymbol{T}_{22}\boldsymbol{\Theta}22\Big) \\
= & \ \Big(a_{11}t_{11}^2 - 2\theta_{11}t_{11}\Big) + a_{11}||\boldsymbol{t}_{21} + a_{11}^{-1}(\boldsymbol{T}_{22}\boldsymbol{a}_{21} - \boldsymbol{\theta}_{12}^t)||^2 - a_{11}^{-1}\boldsymbol{\theta}_{12}\boldsymbol{\theta}_{12}^t \\
& + \mathrm{tr}\,\boldsymbol{T}_{22}(\boldsymbol{A}_{22} - a_{11}^{-1}\boldsymbol{a}_{21}\boldsymbol{a}_{21}^t)\boldsymbol{T}_{22}^t - 2\mathrm{tr}\,\boldsymbol{T}_{22}(\boldsymbol{\Theta}_{22} - a_{11}^{-1}\boldsymbol{a}_{21}\boldsymbol{\theta}_{12}) \qquad (2.6) \\
= & \ \Big(a_{11}t_{11}^2 - 2\theta_{11}t_{11}\Big) + a_{11}||\boldsymbol{t}_{21} + \boldsymbol{z}_1||^2 + h_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_{22}),
\end{aligned}
$$

where $\boldsymbol{C}_2$ is defined in (2.1), $||\boldsymbol{u}||^2 = \boldsymbol{u}^t\boldsymbol{u}$ for suitable column vector $\boldsymbol{u}$,

$$
\begin{aligned}
\boldsymbol{z}_1 &= a_{11}^{-1}(\boldsymbol{T}_{22}\boldsymbol{Y}_2^t - \boldsymbol{\Xi}_2^t)\boldsymbol{y}_1, \\
h_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_{22}) &= \mathrm{tr}\,\boldsymbol{T}_{22}(\boldsymbol{I}_{p-1} + \boldsymbol{Y}_2^t\boldsymbol{C}_2\boldsymbol{Y}_2)\boldsymbol{T}_{22}^t - 2\mathrm{tr}\,\boldsymbol{T}_{22}\boldsymbol{Y}_2^t\boldsymbol{C}_2\boldsymbol{\Xi}_2 \\
& \quad - a_{11}^{-1}\boldsymbol{y}_1^t\boldsymbol{\Xi}_2\boldsymbol{\Xi}_2^t\boldsymbol{y}_1.
\end{aligned}
$$

We are now ready to prove that $\Delta_1 \geq 0$. Combining (2.4), (2.5) and (2.6) gives that

$$
\begin{aligned}
\Delta_1 = & \int\cdots\int \Big\{(d_1 - d_1^* a_{11})(t_{11}^2 + \boldsymbol{t}_{21}^t\boldsymbol{t}_{21}) - \log d_1/(d_1^* a_{11})\Big\} I(d_1 \geq d_1^* a_{11}) \\
& \times c_0(\boldsymbol{\Xi}) \prod_{i=1}^p t_{ii}^{n+m-i} \\
& \times \exp\left[-\frac{1}{2}\Big\{a_{11}t_{11}^2 - 2\theta_{11}t_{11} + a_{11}||\boldsymbol{t}_{21} + \boldsymbol{z}_1||^2 + h_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_{22})\Big\}\right] \qquad (2.7) \\
& \times dt_{11}d\boldsymbol{t}_{21}d\boldsymbol{T}_{22}d\boldsymbol{y}_1 d\boldsymbol{Y}_2.
\end{aligned}
$$

Noting that

$$
\begin{aligned}
& \int ||\boldsymbol{t}_{21}||^2 e^{-a_{11}||\boldsymbol{t}_{21} + \boldsymbol{z}_1||^2/2} d\boldsymbol{t}_{21} \\
& \quad = \int (||\boldsymbol{z}_1||^2 + ||\boldsymbol{x}||^2) e^{-a_{11}||\boldsymbol{x}||^2/2} d\boldsymbol{x} \qquad (2.8) \\
& \quad = \left(||\boldsymbol{z}_1||^2 + \frac{p-1}{a_{11}}\right)(2\pi a_{11})^{(p-1)/2},
\end{aligned}
$$

we can demonstrate that

$$
\begin{aligned}
\Delta_1 \geq & \int\cdots\int \left\{(d_1 - d_1^* a_{11})\left(t_{11}^2 + \frac{p-1}{a_{11}}\right) - \log d_1/(d_1^* a_{11})\right\} I(d_1 \geq d_1^* a_{11}) \qquad (2.9) \\
& \times c_1(\boldsymbol{\Xi}, a_{11}) \prod_{i=1}^p t_{ii}^{n+m-i} \exp\left[-\frac{1}{2}\Big\{a_{11}t_{11}^2 - 2\theta_{11}t_{11} + h_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_{22})\Big\}\right] \\
& \times dt_{11}d\boldsymbol{T}_{22}d\boldsymbol{y}_1 d\boldsymbol{Y}_2,
\end{aligned}
$$

where $c_1(\boldsymbol{\Xi}, a_{11}) = c_0(\boldsymbol{\Xi})(2\pi a_{11})^{(p-1)/2}$. Note that $h_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_{22}) = h_1(-\boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_{22})$, $a_{11} = 1 + \boldsymbol{y}_1^t \boldsymbol{y}_1$ and $\theta_{11} = \boldsymbol{y}_1^t \boldsymbol{\xi}_1$. When we denote the integrand of the r.h.s. of (2.9) by $G_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, t_{11}\boldsymbol{T}_{22})$, the r.h.s. of (2.9) is rewritten by

$$
\begin{aligned}
&\int \cdots \int G_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, t_{11}\boldsymbol{T}_{22}) d\boldsymbol{y}_1 d\boldsymbol{Y}_2 dt_{11} d\boldsymbol{T}_{22} \\
&= \frac{1}{2} \int \cdots \int \{G_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, t_{11}\boldsymbol{T}_{22}) + G_1(-\boldsymbol{y}_1, \boldsymbol{Y}_2, t_{11}\boldsymbol{T}_{22})\} d\boldsymbol{y}_1 d\boldsymbol{Y}_2 dt_{11} d\boldsymbol{T}_{22} \\
&= \int \cdots \int \left\{ (d_1 - d_1^* a_{11}) \left(B_1 + \frac{p-1}{a_{11}} \right) - \log \frac{d_1}{d_1^* a_{11}} \right\} I(d_1 \geq d_1^* a_{11}) \qquad (2.10) \\
&\quad \times c_1(\boldsymbol{\Xi}, a_{11}) \prod_{i=1}^{p} t_{ii}^{n+m-i} \frac{1}{2} \left(e^{\theta_{11} t_{11}} + e^{-\theta_{11} t_{11}} \right) \\
&\quad \times \exp\left\{ -\frac{1}{2} \left(a_{11} t_{11}^2 + h_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_{22}) \right) \right\} dt_{11} d\boldsymbol{T}_{22} d\boldsymbol{y}_1 d\boldsymbol{Y}_2,
\end{aligned}
$$

where

$$
B_1 = \frac{\int_0^\infty t_{11}^{n+m+1} \left(e^{\theta_{11} t_{11}} + e^{-\theta_{11} t_{11}} \right) e^{-a_{11} t_{11}^2/2} dt_{11}}{\int_0^\infty t_{11}^{n+m-1} \left(e^{\theta_{11} t_{11}} + e^{-\theta_{11} t_{11}} \right) e^{-a_{11} t_{11}^2/2} dt_{11}}.
$$

Making the Taylor expansions for $e^{\theta_{11} t_{11}}$ and $e^{-\theta_{11} t_{11}}$, we see that

$$
\begin{aligned}
B_1 &= \frac{\int_0^\infty t_{11}^{n+m+1} \sum_{\ell=0}^\infty \{\theta_{11}^{2\ell} t_{11}^{2\ell}/(2\ell)!\} e^{-a_{11} t_{11}^2/2} dt_{11}}{\int_0^\infty t_{11}^{n+m-1} \sum_{\ell=0}^\infty \{\theta_{11}^{2\ell} t_{11}^{2\ell}/(2\ell)!\} e^{-a_{11} t_{11}^2/2} dt_{11}} \\
&\geq \inf_\ell \frac{\int_0^\infty t_{11}^{n+m+1+2\ell} e^{-a_{11} t_{11}^2/2} dt_{11}}{\int_0^\infty t_{11}^{n+m-1+2\ell} e^{-a_{11} t_{11}^2/2} dt_{11}} \qquad (2.11) \\
&= \inf_\ell \left\{ \frac{1}{a_{11}} \frac{\int_0^\infty x^{(n+m+2+2\ell)/2-1} e^{-x/2} dx}{\int_0^\infty x^{(n+m+2\ell)/2-1} e^{-x/2} dx} \right\} \\
&= \inf_\ell \left\{ \frac{n+m+2\ell}{a_{11}} \right\} \\
&= \frac{n+m}{a_{11}}.
\end{aligned}
$$

Hence the non-negativeness of Δ_1 can be established since

$$
\begin{aligned}
&(d_1 - d_1^* a_{11}) \left(\frac{n+m}{a_{11}} + \frac{p-1}{a_{11}} \right) - \log \frac{d_1}{d_1^* a_{11}} \\
&= \frac{d_1}{d_1^* a_{11}} - \log \frac{d_1}{d_1^* a_{11}} - 1 \\
&\geq 0.
\end{aligned}
$$

Next we shall prove that $\Delta_i \geq 0$ for $i = 2, \ldots, p$. To employ the same arguments as in the above proof, we need to verify that for $i = 2, \ldots, p-1$,

$$
\mathrm{tr} \left\{ \boldsymbol{T}(\boldsymbol{I}_p + \boldsymbol{Y}^t \boldsymbol{Y}) \boldsymbol{T}^t - 2\boldsymbol{T}\boldsymbol{Y}^t \boldsymbol{\Xi}^t \right\} \qquad (2.12)
$$

$$
\begin{aligned}
&= \sum_{j=1}^{i} \left\{ a_{jj} t_{jj}^2 - 2\boldsymbol{y}_j^t \boldsymbol{C}_j \boldsymbol{\xi}_j t_{jj} + a_{jj} ||\boldsymbol{t}_{j+1,j} + \boldsymbol{z}_j||^2 - a_{jj}^{-1} \boldsymbol{y}_j^t \boldsymbol{C}_j \boldsymbol{\Xi}_{j+1} \boldsymbol{\Xi}_{j+1}^t \boldsymbol{C}_j \boldsymbol{y}_j \right\} \\
&\quad + \mathrm{tr}\, \boldsymbol{T}_{i+1,i+1} \left(\boldsymbol{I}_{p-i} + \boldsymbol{Y}_{i+1}^t \boldsymbol{C}_{i+1} \boldsymbol{Y}_{i+1} \right) \boldsymbol{T}_{i+1,i+1}^t - 2 \mathrm{tr}\, \boldsymbol{T}_{i+1,i+1} \boldsymbol{Y}_{i+1}^t \boldsymbol{C}_{i+1} \boldsymbol{\Xi}_{i+1},
\end{aligned}
$$

where $a_{ii} = 1 + \boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{y}_i$,

$$
\boldsymbol{z}_i = a_{ii}^{-1} \left(\boldsymbol{T}_{i+1,i+1} \boldsymbol{Y}_{i+1}^t - \boldsymbol{\Xi}_{i+1}^t \right) \boldsymbol{C}_i \boldsymbol{y}_i
$$

and $\boldsymbol{\Xi}^t = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_i, \boldsymbol{\Xi}_{i+1})$ for column vectors $\boldsymbol{\xi}_i$'s. The same arguments as in (2.6) are used to check the expression (2.12). In fact, we can observe that

$$
\begin{aligned}
&\mathrm{tr}\, \boldsymbol{T}_{ii} \left(\boldsymbol{I}_{p-i+1} + \boldsymbol{Y}_i^t \boldsymbol{C}_i \boldsymbol{Y}_i \right) \boldsymbol{T}_{ii}^t - 2 \mathrm{tr}\, \boldsymbol{T}_{ii} \boldsymbol{Y}_i^t \boldsymbol{C}_i \boldsymbol{\Xi} \\
&= \mathrm{tr} \left\{ \begin{pmatrix} t_{ii} & \boldsymbol{0} \\ \boldsymbol{t}_{i+1,i} & \boldsymbol{T}_{i+1,i+1} \end{pmatrix} \begin{pmatrix} a_{ii} & \boldsymbol{a}_{i+1,i}^t \\ \boldsymbol{a}_{i+1,i} & \boldsymbol{A}_{i+1,i+1} \end{pmatrix} \begin{pmatrix} t_{ii} & \boldsymbol{t}_{i+1,i}^t \\ 0 & \boldsymbol{T}_{i+1,i+1}^t \end{pmatrix} \right. \\
&\qquad \left. -2 \begin{pmatrix} t_{ii} & \boldsymbol{0} \\ \boldsymbol{t}_{i+1,i} & \boldsymbol{T}_{i+1,i+1} \end{pmatrix} \begin{pmatrix} \theta_{ii} & \boldsymbol{\theta}_{i,i+1} \\ \boldsymbol{\theta}_{i+1,i} & \boldsymbol{\Theta}_{i+1,i+1} \end{pmatrix} \right\} \\
&= \left(a_{ii} t_{ii}^2 - 2\theta_{ii} t_{ii} \right) + a_{ii} ||\boldsymbol{t}_{i+1,i} + a_{ii}^{-1} (\boldsymbol{T}_{i+1,i+1} \boldsymbol{a}_{i+1,i} - \boldsymbol{\theta}_{i,i+1}^t)||^2 - a_{ii}^{-1} \boldsymbol{\theta}_{i,i+1} \boldsymbol{\theta}_{i,i+1}^t \\
&\quad + \mathrm{tr}\, \boldsymbol{T}_{i+1,i+1} (\boldsymbol{A}_{i+1,i+1} - a_{ii}^{-1} \boldsymbol{a}_{i+1,i} \boldsymbol{a}_{i+1,i}^t) \boldsymbol{T}_{i+1,i+1}^t \\
&\quad - 2 \mathrm{tr}\, \boldsymbol{T}_{i+1,i+1} (\boldsymbol{\Theta}_{i+1,i+1} - a_{ii}^{-1} \boldsymbol{a}_{i+1,i} \boldsymbol{\theta}_{i,i+1}) \\
&= \left(a_{ii} t_{ii}^2 - 2\boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{\xi}_i t_{ii} \right) + a_{ii} ||\boldsymbol{t}_{i+1,i} + \boldsymbol{z}_i||^2 - a_{ii}^{-1} \boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{\Xi}_{i+1} \boldsymbol{\Xi}_{i+1}^t \boldsymbol{C}_i \boldsymbol{y}_i \\
&\quad + \mathrm{tr}\, \boldsymbol{T}_{i+1,i+1} \left(\boldsymbol{I}_{p-i} + \boldsymbol{Y}_{i+1}^t \left(\boldsymbol{C}_i - a_{ii}^{-1} \boldsymbol{C}_i \boldsymbol{y}_i \boldsymbol{y}_i^t \boldsymbol{C}_i \right) \boldsymbol{Y}_{i+1} \right) \boldsymbol{T}_{i+1,i+1}^t \\
&\quad - 2 \mathrm{tr}\, \boldsymbol{T}_{i+1,i+1} \boldsymbol{Y}_{i+1}^t \left(\boldsymbol{C}_i - a_{ii}^{-1} \boldsymbol{C}_i \boldsymbol{y}_i \boldsymbol{y}_i^t \boldsymbol{C}_i \right) \boldsymbol{\Xi}_{i+1},
\end{aligned}
$$

which proves the expression (2.12), where $\boldsymbol{a}_{i+1,i} = \boldsymbol{Y}_{i+1}^t \boldsymbol{C}_i \boldsymbol{y}_i$, $\boldsymbol{A}_{i+1,i+1} = \boldsymbol{I}_{p-i} + \boldsymbol{Y}_{i+1}^t \boldsymbol{C}_i \boldsymbol{Y}_{i+1}$, $\theta_{ii} = \boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{\xi}_i$, $\theta_{i,i+1} = \boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{\Xi}_{i+1}$ and $\boldsymbol{\Theta}_{i+1,i+1} = \boldsymbol{Y}_{i+1}^t \boldsymbol{C}_i \boldsymbol{\Xi}_{i+1}$. Integrating out with respect to $\boldsymbol{t}_{21}, \boldsymbol{t}_{32}, \ldots, \boldsymbol{t}_{i+1,i}$ and using the expression (2.12) and the same arguments as in (2.7), we see that

$$
\begin{aligned}
\Delta_i \geq & \int \cdots \int \left[(d_i - d_i^* a_{ii}) \left(t_{ii}^2 + \frac{p-i}{a_{ii}} \right) - \log \frac{d_i}{d_i^* a_{ii}} \right] I(d_i \geq d_i^* a_{ii}) \\
& \times c_j(\boldsymbol{\Xi}, a_{11}, \ldots, a_{ii}) \prod_{i=1}^{p} t_{ii}^{n+m-i} \\
& \times \exp \left[-\frac{1}{2} \left\{ \sum_{j=1}^{i} \left(a_{jj} t_{jj}^2 - 2\theta_{jj} t_{jj} \right) + h_i(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i, \boldsymbol{Y}_{i+1}, \boldsymbol{T}_{i+1,i+1}) \right\} \right] \\
& \times \prod_{j=1}^{i} (dt_{jj} d\boldsymbol{y}_j) d\boldsymbol{Y}_{i+1} \boldsymbol{T}_{i+1},
\end{aligned}
$$

where

$$
\begin{aligned}
& h_i(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i, \boldsymbol{Y}_{i+1}, \boldsymbol{T}_{i+1,i+1}) \\
= \quad & -\sum_{j=1}^{i} \left\{ a_{jj}^{-1} \boldsymbol{y}_j^t \boldsymbol{C}_j \boldsymbol{\Xi}_{j+1} \boldsymbol{\Xi}_{j+1}^t \boldsymbol{C}_j \boldsymbol{y}_j \right\} \qquad (2.13) \\
& + \mathrm{tr}\, \boldsymbol{T}_{i+1,i+1} \left(\boldsymbol{I}_{p-i} + \boldsymbol{Y}_{i+1}^t \boldsymbol{C}_{i+1} \boldsymbol{Y}_{i+1} \right) \boldsymbol{T}_{i+1,i+1}^t - 2\mathrm{tr}\, \boldsymbol{T}_{i+1,i+1} \boldsymbol{Y}_{i+1}^t \boldsymbol{C}_{i+1} \boldsymbol{\Xi}_{i+1}.
\end{aligned}
$$

Similarly to (2.10), the non-negativeness of Δ_i can be verified if we can show that

$$
\left\{ (d_i - d_i^* a_{ii}) \left(B_i + \frac{p-i}{a_{ii}} \right) - \log \frac{d_i}{d_i^* a_{ii}} \right\} I(d_i \geq d_i^* a_{ii}) \geq 0, \qquad (2.14)
$$

where

$$
B_i = \frac{\int_0^\infty t_{ii}^{n+m-i+2} \left(e^{\theta_{ii} t_{ii}} + e^{-\theta_{ii} t_{ii}} \right) e^{-a_{ii} t_{ii}^2/2} dt_{ii}}{\int_0^\infty t_{ii}^{n+m-i} \left(e^{\theta_{ii} t_{ii}} + e^{-\theta_{ii} t_{ii}} \right) e^{-a_{ii} t_{ii}^2/2} dt_{ii}}.
$$

From (2.11), we have that $B_i \geq (n+m-i+1)/a_{ii}$, so that the inequality (2.14) is guaranteed. Therefore the proof of Theorem 1 is complete. □□

2.2 Improvements on scale equivariant minimax estimators

It is known that the James-Stein minimax estimator treated in the previous subsection has a drawback that it depends on the coordinate system. We here try to construct truncated procedures improving on minimax estimators not depending on the coordinate system when $m \geq p$ or $\boldsymbol{X}\boldsymbol{X}^t$ is of full rank.

Assume that $m \geq p$ in this subsection. We consider the following equivariant estimators under a scale transformation:

$$
\widehat{\boldsymbol{\Sigma}}(\boldsymbol{H}^t \boldsymbol{A} \boldsymbol{S} \boldsymbol{A} \boldsymbol{H}, \boldsymbol{H}^t \boldsymbol{A} \boldsymbol{X} \boldsymbol{O}) = \boldsymbol{H}^t \boldsymbol{A} \widehat{\boldsymbol{\Sigma}}(\boldsymbol{S}, \boldsymbol{X}) \boldsymbol{A} \boldsymbol{H}, \qquad (2.15)
$$

for any $\boldsymbol{H} \in O(p)$, any $\boldsymbol{O} \in O(m)$ and any $p \times p$ symmetric matrix $\boldsymbol{A}$, where $O(p)$ is the group of $p \times p$ orthogonal matrices. Then it can be seen that (2.15) is equivalent to

$$
\widehat{\boldsymbol{\Sigma}}(\boldsymbol{S}, \boldsymbol{X}) = (\boldsymbol{X}\boldsymbol{X}^t)^{1/2} \boldsymbol{H} \boldsymbol{\Psi}(\boldsymbol{H}^t \boldsymbol{F} \boldsymbol{H}) \boldsymbol{H}^t (\boldsymbol{X}\boldsymbol{X}^t)^{1/2}, \qquad (2.16)
$$

for any $\boldsymbol{H} \in O(p)$, where $\boldsymbol{F} = (\boldsymbol{X}\boldsymbol{X}^t)^{-1/2} \boldsymbol{S} (\boldsymbol{X}\boldsymbol{X}^t)^{-1/2}$, and $(\boldsymbol{X}\boldsymbol{X}^t)^{1/2}$ is a symmetric matrix such that $(\boldsymbol{X}\boldsymbol{X}^t) = ((\boldsymbol{X}\boldsymbol{X}^t)^{1/2})^2$. Let $\boldsymbol{P}$ be an orthogonal $p \times p$ matrix such that

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

with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$. Then the estimator (2.16) can expressed by

$$
\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}) = (\boldsymbol{X}\boldsymbol{X}^t)^{1/2} \boldsymbol{P} \boldsymbol{\Psi}(\boldsymbol{\Lambda}) \boldsymbol{P}^t (\boldsymbol{X}\boldsymbol{X}^t)^{1/2} \qquad (2.17)
$$

for

$$\boldsymbol{\Psi}(\boldsymbol{\Lambda}) = \mathrm{diag}\,(\psi_1(\boldsymbol{\Lambda}), \ldots, \psi_p(\boldsymbol{\Lambda})),$$

where $\psi_i(\boldsymbol{\Lambda})$'s are non-negative functions of $\boldsymbol{\Lambda}$. The diagonalization of $\boldsymbol{\Psi}(\boldsymbol{\Lambda})$ follows from the requirement that $\boldsymbol{\Psi}(\boldsymbol{\Lambda}) = \boldsymbol{\epsilon}\boldsymbol{\Psi}(\boldsymbol{\epsilon}\boldsymbol{\Lambda}\boldsymbol{\epsilon})\boldsymbol{\epsilon}$ for any $\boldsymbol{\epsilon} = \mathrm{diag}\,(\pm 1, \ldots, \pm 1)$. This type of estimators is motivated by Srivastava and Kubokawa (1999). We call them scale equivariant in this paper.

For given estimator $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$, we define a truncation rule $[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}$ by

$$\begin{aligned} [\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR} &= \mathrm{diag}\,(\psi_1^{TR}(\boldsymbol{\Lambda}), \ldots, \psi_p^{TR}(\boldsymbol{\Lambda})), \\ \psi_i^{TR}(\boldsymbol{\Lambda}) &= \min\left\{\psi_i(\boldsymbol{\Lambda}), \frac{\lambda_i + 1}{n+m}\right\}, \quad i = 1, \ldots, p, \end{aligned} \tag{2.18}$$

which gives the corresponding truncated estimator of the form

$$\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR}) = (\boldsymbol{X}\boldsymbol{X}^t)^{1/2}\boldsymbol{P}\mathrm{diag}\,(\psi_1^{TR}(\boldsymbol{\Lambda}), \ldots, \psi_p^{TR}(\boldsymbol{\Lambda}))\boldsymbol{P}^t(\boldsymbol{X}\boldsymbol{X}^t)^{1/2}. \tag{2.19}$$

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

Theorem 2. *The truncated estimator $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR})$ dominates the scale equivariant estimator $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ relative to the Stein loss (1.1) if $P\left[[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR} \neq \boldsymbol{\Psi}(\boldsymbol{\Lambda})\right] > 0$ at some ω.*

It is interesting to show that $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ is minimax under the same conditions on $\boldsymbol{\Psi}$ for the minimaxity of an orthogonally equivariant estimators based on $\boldsymbol{S}$ only, given by

$$\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}) = \boldsymbol{R}\boldsymbol{\Psi}(\boldsymbol{L}^*)\boldsymbol{R}^t, \tag{2.20}$$

where $\boldsymbol{R}$ is an orthogonal matrix such that $\boldsymbol{S} = \boldsymbol{R}\boldsymbol{L}^*\boldsymbol{R}^t$ and $\boldsymbol{L}^* = \mathrm{diag}\,(\ell_1^*, \ldots, \ell_p^*)$ for eigen values $\ell_1^* \geq \ldots \geq \ell_p^*$.

Proposition 1.

(1) *If the orthogonally equivariant estimator $\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ is minimax, then for the same function $\boldsymbol{\Psi}$, $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ is minimax and scale equivariant one improving on $\widehat{\boldsymbol{\Sigma}}^{JS}$ relative to the Stein loss (1.1).*

(2) *If $P[\psi_i(\boldsymbol{\Lambda}) < \psi_j(\boldsymbol{\Lambda})] > 0$ for some $i < j$, then $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^O)$ dominates $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$, where $\boldsymbol{\Psi}^O(\boldsymbol{\Lambda}) = \mathrm{diag}\,(\psi_1^O(\boldsymbol{\Lambda}), \ldots, \psi_p^O(\boldsymbol{\Lambda}))$ majorizes $(\psi_1(\boldsymbol{\Lambda}), \ldots, \psi_p(\boldsymbol{\Lambda}))$, that is, $\sum_{i=1}^j \psi_i^O \geq \sum_{i=1}^j \psi_i$ for $1 \leq j \leq p-1$ and $\sum_{i=1}^p \psi_i^O = \sum_{i=1}^p \psi_i$.*

Proof. Recall that $\boldsymbol{F} = (\boldsymbol{X}\boldsymbol{X}^t)^{-1/2}\boldsymbol{S}(\boldsymbol{X}\boldsymbol{X}^t)^{-1/2} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^t$ and that $\boldsymbol{S} \sim \mathcal{W}_p(n, \boldsymbol{I}_p)$. Then it is seen that the conditional distribution of $\boldsymbol{F}$ given $\boldsymbol{X}$ has $\mathcal{W}_p(n, \boldsymbol{\Sigma}_*)$ for $\boldsymbol{\Sigma}_* = (\boldsymbol{X}\boldsymbol{X}^t)^{-1}$. Then the risk function of $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ is represented by

$$R(\omega, \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})) = E^X\left[E^{F|X}\left[\mathrm{tr}\,\boldsymbol{P}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{P}^t\boldsymbol{\Sigma}_*^{-1} - \log|\boldsymbol{P}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{P}^t\boldsymbol{\Sigma}_*^{-1}| - p\Big|\boldsymbol{X}\right]\right], \tag{2.21}$$

so that given $\boldsymbol{X}$, conditionally $\boldsymbol{P\Psi P}^t$ corresponds to the orthogonally invariant estimator $\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ of $\boldsymbol{\Sigma}_*$ with $\boldsymbol{S} \sim \mathcal{W}(n, \boldsymbol{\Sigma}^*)$. Hence the minimaxity of $\widetilde{\boldsymbol{\Sigma}}(\Psi)$ implies the minimaxity of $\widehat{\boldsymbol{\Sigma}}(\Psi)$, which proves the part (1). The part (2) follows from (2.21) and the results of Sheena and Takemura (1992). □□

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

[1] Stein type estimator. Let $\widehat{\boldsymbol{\Sigma}}^S = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^S)$ for

$$\boldsymbol{\Psi}^S(\boldsymbol{\Lambda}) = \mathrm{diag}\,(d_1\lambda_1, \ldots, d_p\lambda_p). \tag{2.22}$$

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

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

$$\widehat{\boldsymbol{\Sigma}}^T = (\boldsymbol{X}\boldsymbol{X}^t)^{1/2}\left\{\int_{O(p)} \boldsymbol{\Gamma}\boldsymbol{U}_\Gamma \boldsymbol{D}_m \boldsymbol{U}_\Gamma^t \boldsymbol{\Gamma}^t d\mu(\boldsymbol{\Gamma})\right\}(\boldsymbol{X}\boldsymbol{X}^t)^{1/2}, \tag{2.23}$$

where $\boldsymbol{U}_\Gamma \in G_T^+$ with $\boldsymbol{U}_\Gamma\boldsymbol{U}_\Gamma^t = \boldsymbol{\Gamma}^t\boldsymbol{F}\boldsymbol{\Gamma}$ for $\boldsymbol{F} = (\boldsymbol{X}\boldsymbol{X}^t)^{-1/2}\boldsymbol{S}(\boldsymbol{X}\boldsymbol{X}^t)^{-1/2} = \boldsymbol{P\Lambda P}^t$. Takemura (1984) provided another expression as $\widehat{\boldsymbol{\Sigma}}^T = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^T)$ for $\boldsymbol{\Psi}^T(\boldsymbol{\Lambda}) = \mathrm{diag}\,(\psi_1^T, \ldots, \psi_p^T)$, where

$$(\psi_1^T, \ldots, \psi_p^T)^t = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_p)\boldsymbol{W}(\boldsymbol{\Lambda})(d_1, \ldots, d_p)^t, \tag{2.24}$$

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

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

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

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

$$\tilde{w}_{ij}(\boldsymbol{\Lambda}) = \frac{\mathrm{tr}_{\,j-1}(\boldsymbol{\Lambda}_i)}{\mathrm{tr}_{\,j-1}(\boldsymbol{\Lambda})} - \frac{\mathrm{tr}_{\,j}(\boldsymbol{\Lambda}_i)}{\mathrm{tr}_{\,j}(\boldsymbol{\Lambda})},$$

for

$$\mathrm{tr}_{\,j}(\boldsymbol{\Lambda}) = \begin{cases} 1 & \text{if } j = 0, \\ \sum_{1\leq i_1<\cdots<i_j\leq p}\prod_{k=1}^{j}\lambda_{i_k} & \text{if } j = 1, \ldots, p, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\boldsymbol{\Lambda}_i = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_{i-1}, 0, \lambda_{i+1}, \ldots, \lambda_p).$$

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

$$\left(\psi_1^P, \ldots, \psi_p^P\right)' = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_p)\widetilde{\boldsymbol{W}}(\boldsymbol{\Lambda})(d_1, \ldots, d_p)'. \tag{2.25}$$

For $p = 2$, they are given by

$$\begin{aligned}\psi_1^P &= \lambda_1\left(\frac{\lambda_1}{\lambda_1+\lambda_2}d_1 + \frac{\lambda_2}{\lambda_1+\lambda_2}d_2\right),\\ \psi_2^P &= \lambda_2\left(\frac{\lambda_2}{\lambda_1+\lambda_2}d_1 + \frac{\lambda_1}{\lambda_1+\lambda_2}d_2\right).\end{aligned}$$

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

[4] Haff type estimator. Let

$$\widehat{\boldsymbol{\Sigma}}^H = \frac{1}{n}\left(\boldsymbol{S} + \frac{a_0}{\mathrm{tr}\,\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{X}^t}\boldsymbol{X}\boldsymbol{X}^t\right). \tag{2.26}$$

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

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

Now, applying the truncation rule (2.18) to the above estimators yields the improved estimators.

Corollary 1. *For $\boldsymbol{\Psi} = \boldsymbol{\Psi}^S$, $\boldsymbol{\Psi}^T$ and $\boldsymbol{\Psi}^P$, the estimator $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR})$ is scale-equivariant, minimax and improving on the corresponding estimator $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ relative to the Stein loss (1.1). Also $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^H]^{TR})$ dominates $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^H)$.*

It should be noted that Corollary 1 does not imply the dominance of $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR})$ over $\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$, but states the dominance of $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR})$ over $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$. Although $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ is not identical to $\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$, if $\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ is a superior minimax estimator, $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ inherits the same good risk properties with minimaxity and improvement. Corollary 1 states that these minimax estimators can be further improved on by $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR})$ by employing the information in $\boldsymbol{X}$.

Proof of Theorem 2. Without any loss of generality, let $\boldsymbol{\Sigma} = \boldsymbol{I}_p$. We first consider the expectation of the general function $h(\boldsymbol{F}, \boldsymbol{X}\boldsymbol{X}^t)$ of $\boldsymbol{F}$ and $\boldsymbol{X}\boldsymbol{X}^t$. The expectation is evaluated as

$$
\begin{aligned}
& E\left[h(\boldsymbol{F}, \boldsymbol{X}\boldsymbol{X}^t)\right] \\
= \ & c_0(\boldsymbol{\Xi}) \int\int h(\boldsymbol{F}, \boldsymbol{X}\boldsymbol{X}^t) |\boldsymbol{S}|^{(n-p-1)/2} \\
& \exp\left\{-\mathrm{tr}\,(\boldsymbol{S} + \boldsymbol{X}\boldsymbol{X}^t - 2\boldsymbol{X}\boldsymbol{\Xi}^t)/2\right\} d\boldsymbol{X} d\boldsymbol{S} \qquad (2.27) \\
= \ & c_0(\boldsymbol{\Xi}) \int\int h(\boldsymbol{F}, \boldsymbol{X}\boldsymbol{X}^t) |\boldsymbol{S}|^{(n-p-1)/2} \\
& \exp\left\{-\mathrm{tr}\,(\boldsymbol{S} + \boldsymbol{X}\boldsymbol{X}^t)/2\right\} \int \exp\left\{\mathrm{tr}\, \boldsymbol{X}\boldsymbol{H}\boldsymbol{\Xi}^t/2\right\} \mu(d\boldsymbol{H}) d\boldsymbol{X} d\boldsymbol{S},
\end{aligned}
$$

where $\mu(d\boldsymbol{H})$ denotes an invariant probability measure on the group of orthogonal matrices. Here the second equality in (2.27) follows from the fact that $\boldsymbol{F}$ and $\boldsymbol{X}\boldsymbol{X}^t$ are invariant under the transformation $\boldsymbol{X} \to \boldsymbol{X}\boldsymbol{H}$ for $m \times m$ orthogonal matrix $\boldsymbol{H}$. One of the essential properties of zonal polynomials gives

$$
\int \exp\left\{\mathrm{tr}\, \boldsymbol{X}\boldsymbol{H}\boldsymbol{\Xi}^t/2\right\} \mu(d\boldsymbol{H}) = \sum_{\kappa} \alpha_{\kappa}^{(m)} C_{\kappa}\left(\boldsymbol{\Xi}\boldsymbol{\Xi}^t \boldsymbol{X}\boldsymbol{X}^t\right),
$$

where $\alpha_{\kappa}^{(m)}$ is given in James (1964) and $C_{\kappa}(\boldsymbol{Z})$ denotes the normalized zonal polynomials of the positive definite matrix $\boldsymbol{Z}$ of order p corresponding to partitions $\kappa = \{\kappa_1, \ldots, \kappa_p\}$ so that for all $k = 0, 1, 2, \ldots$,

$$
(\mathrm{tr}\, \boldsymbol{Z})^k = \sum_{\{\kappa : \kappa_1 + \cdots + \kappa_p = k\}} C_{\kappa}(\boldsymbol{Z}).
$$

Let $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{X}^t$, and the r.h.s. of (2.27) is written by

$$
\begin{aligned}
& c_1(\boldsymbol{\Xi}) \int\int h(\boldsymbol{F}, \boldsymbol{W}) |\boldsymbol{S}|^{(n-p-1)/2} |\boldsymbol{W}|^{(m-p-1)/2} \\
& \qquad \exp\left\{-\mathrm{tr}\,(\boldsymbol{S} + \boldsymbol{W})/2\right\} \sum_{\kappa} \alpha_{\kappa}^{(m)} C_{\kappa}(\boldsymbol{\Xi}\boldsymbol{\Xi}^t \boldsymbol{W}) d\boldsymbol{S} d\boldsymbol{W},
\end{aligned}
$$

for the normalizing function $c_1(\boldsymbol{\Xi})$. Making the transformation $\boldsymbol{F} = \boldsymbol{W}^{-1/2}\boldsymbol{S}\boldsymbol{W}^{-1/2}$ with $J(\boldsymbol{S} \to \boldsymbol{F}) = |\boldsymbol{W}|^{(p+1)/2}$ gives that

$$
\begin{aligned}
E\left[h(\boldsymbol{F}, \boldsymbol{X}\boldsymbol{X}^t)\right] = \ & c_1(\boldsymbol{\Xi}) \int\int h(\boldsymbol{F}, \boldsymbol{W}) |\boldsymbol{F}|^{(n-p-1)/2} |\boldsymbol{W}|^{(n+m-p-1)/2} \qquad (2.28) \\
& \times \exp\left\{-\mathrm{tr}\,(\boldsymbol{F} + \boldsymbol{I})\boldsymbol{W}/2\right\} \sum_{\kappa} \alpha_{\kappa}^{(m)} C_{\kappa}(\boldsymbol{\Xi}\boldsymbol{\Xi}^t \boldsymbol{W}) d\boldsymbol{F} d\boldsymbol{W}.
\end{aligned}
$$

Again making the transformations $\boldsymbol{F} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^t$ and $\boldsymbol{W} = \boldsymbol{P}\boldsymbol{V}\boldsymbol{P}^t$ in order, we see that (2.28) is represented as

$$
E\left[h(\boldsymbol{F}, \boldsymbol{X}\boldsymbol{X}^t)\right]
$$

$$
\begin{aligned}
&= c_2(\boldsymbol{\Xi}) \int\int\int h(\boldsymbol{P\Lambda P}^t, \boldsymbol{W}) h(\boldsymbol{\Lambda}) |\boldsymbol{W}|^{(n+m-p-1)/2} \\
&\quad \times \exp\left\{-\mathrm{tr}\,(\boldsymbol{\Lambda}+\boldsymbol{I})\boldsymbol{P}^t\boldsymbol{W}\boldsymbol{P}/2\right\} \sum_{\kappa} \alpha_{\kappa}^{(m)} C_{\kappa}(\boldsymbol{\Xi}\boldsymbol{\Xi}^t\boldsymbol{W})\mu(d\boldsymbol{P})d\boldsymbol{\Lambda}d\boldsymbol{W} \\
&= c_2(\boldsymbol{\Xi}) \int\int\int h(\boldsymbol{P\Lambda P}^t, \boldsymbol{PVP}^t) h(\boldsymbol{\Lambda}) |\boldsymbol{V}|^{(n+m-p-1)/2} \qquad (2.29) \\
&\quad \times \exp\{-\mathrm{tr}\,(\boldsymbol{\Lambda}+\boldsymbol{I})\boldsymbol{V}/2\} \sum_{\kappa} \alpha_{\kappa}^{(m)} C_{\kappa}(\boldsymbol{\Xi}\boldsymbol{\Xi}^t\boldsymbol{PVP}^t)\mu(d\boldsymbol{P})d\boldsymbol{\Lambda}d\boldsymbol{V},
\end{aligned}
$$

where $h(\boldsymbol{\Lambda})$ is a function of $\boldsymbol{\Lambda}$ (see Srivastava and Khatri (1979)).

Based on the expression (2.29), we can evaluate the risk difference of the two estimators, which is given by

$$
\begin{aligned}
&R(\omega, \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})) - R(\omega, \widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR})) \\
&= E\left[\mathrm{tr}\left\{\boldsymbol{P\Psi}(\boldsymbol{\Lambda})\boldsymbol{P}^t - \boldsymbol{P}[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\boldsymbol{P}^t\right\}\boldsymbol{W} - \log|\boldsymbol{\Psi}(\boldsymbol{\Lambda})\{[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\}^{-1}|\right] \qquad (2.30) \\
&= E^{\Lambda}\left[\mathrm{tr}\left\{\boldsymbol{\Psi}(\boldsymbol{\Lambda}) - [\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\right\}E[\boldsymbol{V}|\boldsymbol{\Lambda}] - \log|\boldsymbol{\Psi}(\boldsymbol{\Lambda})\{[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\}^{-1}|\right].
\end{aligned}
$$

By the basic property of zonal polynomials,

$$
\int C_{\kappa}(\boldsymbol{\Xi}\boldsymbol{\Xi}^t\boldsymbol{PVP}^t)\mu(d\boldsymbol{P}) = C_{\kappa}(\boldsymbol{\Xi}\boldsymbol{\Xi}^t)C_{\kappa}(\boldsymbol{V})/C_{\kappa}(\boldsymbol{I}_p). \qquad (2.31)
$$

For simplicity, let us put

$$
\begin{aligned}
\boldsymbol{A} &= \mathrm{diag}\,(a_1, \ldots, a_p) \\
&= \left\{\boldsymbol{\Psi}(\boldsymbol{\Lambda}) - [\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\right\}(\boldsymbol{\Lambda}+\boldsymbol{I})^{-1}, \\
\boldsymbol{B} &= (\boldsymbol{\Lambda}+\boldsymbol{I})^{-1}.
\end{aligned}
$$

Then from (2.31), it can be seen that

$$
\begin{aligned}
&\mathrm{tr}\left\{\boldsymbol{\Psi}(\boldsymbol{\Lambda}) - [\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\right\}E[\boldsymbol{V}C_{\kappa}(\boldsymbol{V})|\boldsymbol{\Lambda}]/E[C_{\kappa}(\boldsymbol{V})|\boldsymbol{\Lambda}] \qquad (2.32) \\
&\geq \inf_{\kappa}\left\{\frac{c_3(\boldsymbol{B})}{E[C_{\kappa}(\boldsymbol{V})|\boldsymbol{\Lambda}]}\int\left(\mathrm{tr}\,\boldsymbol{AVB}^{-1}\right)C_{\kappa}(\boldsymbol{V})|\boldsymbol{V}|^{(n+m-p-1)/2}e^{-\mathrm{tr}\,\boldsymbol{VB}^{-1}/2}d\boldsymbol{V}\right\},
\end{aligned}
$$

where $c_3(\boldsymbol{B})$ ia a normalizing function in $\mathcal{W}_p(n+m, \boldsymbol{B})$. If we can show that for any κ,

$$
\begin{aligned}
&E\left[\mathrm{tr}\,\boldsymbol{AVB}^{-1}C_{\kappa}(\boldsymbol{V})|\boldsymbol{\Lambda}\right] \\
&= c_3(\boldsymbol{B})\int\left(\mathrm{tr}\,\boldsymbol{AVB}^{-1}\right)C_{\kappa}(\boldsymbol{V})|\boldsymbol{V}|^{(n+m-p-1)/2}e^{-\mathrm{tr}\,\boldsymbol{VB}^{-1}/2}d\boldsymbol{V} \\
&\geq (n+m)(\mathrm{tr}\,\boldsymbol{A})E\left[C_{\kappa}(\boldsymbol{V})|\boldsymbol{\Lambda}\right], \qquad (2.33)
\end{aligned}
$$

where conditionally, $\boldsymbol{V}|\boldsymbol{\Lambda} \sim \mathcal{W}_p(n+m, \boldsymbol{B})$, then the r.h.s. of the extreme equation in (2.30) is evaluated as

(the r.h.s. of (2.30))

$$
\begin{aligned}
&\geq \; E\left[\mathrm{tr}\,\left\{\boldsymbol{\Psi}(\boldsymbol{\Lambda})-[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\right\}(n+m)(\boldsymbol{\Lambda}+\boldsymbol{I})^{-1}-\log|\boldsymbol{\Psi}(\boldsymbol{\Lambda})\{[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\}^{-1}|\right]\\
&= \; \sum_{i=1}^{p} E\left[\left\{\frac{n+m}{\lambda_i+1}\psi_i(\boldsymbol{\Lambda})-1-\log\frac{n+m}{\lambda_i+1}\psi_i(\boldsymbol{\Lambda})\right\} I\left(\frac{n+m}{\lambda_i+1}\psi_i(\boldsymbol{\Lambda})\geq 1\right)\right]\\
&\geq \; 0.
\end{aligned}
$$

Hence we complete the proof of Theorem 2 with verifying the inequality (2.33).

We shall use the Stein-Haff identity due to Stein (1977) and Haff (1979) to prove the inequality (2.33). For the Kronecker's delta δ_{ij} and $\boldsymbol{V}=(v_{ij})$, let

$$
d_{ij}=\frac{1}{2}\,(1+\delta_{ij})\,\frac{\partial}{\partial v_{ij}},
$$

and denote $\mathcal{D}=(d_{ij})$. For $p\times p$ matrix $\boldsymbol{G}(\boldsymbol{V})=(g_{ij}(\boldsymbol{V}))$, define $\mathcal{D}\boldsymbol{G}(\boldsymbol{V})$ by

$$
[\mathcal{D}\boldsymbol{G}(\boldsymbol{V})]_{ij}=\sum_{s=1}^{p} d_{is}g_{sj}(\boldsymbol{V}).
$$

Then the Stein-Haff identity is given by

$$
E\left[\mathrm{tr}\,\boldsymbol{G}(\boldsymbol{V})\boldsymbol{B}^{-1}|\boldsymbol{\Lambda}\right]=E\left[2\mathrm{tr}\,[\mathcal{D}\boldsymbol{G}(\boldsymbol{V})]+(n+m-p-1)\mathrm{tr}\,\boldsymbol{G}(\boldsymbol{V})\boldsymbol{V}^{-1}|\boldsymbol{\Lambda}\right]. \tag{2.34}
$$

This identity is applied to the conditional expectation (2.33), which is rewritten as

$$
\begin{aligned}
&E\left[\mathrm{tr}\,\boldsymbol{A}\boldsymbol{V}\boldsymbol{B}^{-1}C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}\right]\\
&\quad = \; E\left[2\mathrm{tr}\,[\mathcal{D}\{\boldsymbol{A}\boldsymbol{V}C_\kappa(\boldsymbol{V})\}]+(n+m-p-1)\mathrm{tr}\,\boldsymbol{A}C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}\right].
\end{aligned} \tag{2.35}
$$

For evaluating the first term in the r.h.s. of (2.35), we observe that

$$
[\mathcal{D}\{\boldsymbol{A}\boldsymbol{V}C_\kappa(\boldsymbol{V})\}]_{ij}=\sum_{s=1}^{p} a_s\left(\frac{1+\delta_{is}}{2}\right)\delta_{ij}C_\kappa(\boldsymbol{V})+\sum_{s=1}^{p} a_s v_{sj}\{d_{is}C_\kappa(\boldsymbol{V})\},
$$

which yields that

$$
\mathrm{tr}\,[\mathcal{D}\{\boldsymbol{A}\boldsymbol{V}C_\kappa(\boldsymbol{V})\}]=\frac{p+1}{2}(\mathrm{tr}\,\boldsymbol{A})C_\kappa(\boldsymbol{V})+\sum_{i=1}^{p}\sum_{s=1}^{p} a_s v_{si}\{d_{is}C_\kappa(\boldsymbol{V})\}. \tag{2.36}
$$

Here we need to evaluate the second term in the r.h.s. of (2.36). Since zonal polynomials $C_\kappa(\boldsymbol{V})$ are polynomials of the eigen values $\ell_1,\ldots,\ell_p$ ($\ell_1\geq\cdots\ell_p$) of $\boldsymbol{V}$ with nonnegative coefficients, we can put

$$
f(\boldsymbol{L})=C_\kappa(\boldsymbol{V})
$$

for $\boldsymbol{L}=\mathrm{diag}\,(\ell_1,\ldots,\ell_p)$, and note that

$$
\frac{\partial}{\partial\ell_i}f(\boldsymbol{L})\equiv f_i(\boldsymbol{L})\geq 0.
$$

Since $\boldsymbol{V} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t$ for $p \times p$ orthogonal matrix $\boldsymbol{H} = (h_{ij})$, we obtain by taking differentials, as in Srivastava and Khatri (1979, p.31),

$$d\boldsymbol{V} = (d\boldsymbol{H})\boldsymbol{L}\boldsymbol{H}^t + \boldsymbol{H}(d\boldsymbol{L})\boldsymbol{H}^t + \boldsymbol{H}\boldsymbol{L}(d\boldsymbol{H}^t),$$

so that

$$\boldsymbol{H}^t(d\boldsymbol{V})\boldsymbol{H} = \boldsymbol{H}^t(d\boldsymbol{H})\boldsymbol{L} + d\boldsymbol{L} + \boldsymbol{L}(d\boldsymbol{H}^t)\boldsymbol{H}. \tag{2.37}$$

Since $\boldsymbol{H}^t\boldsymbol{H} = \boldsymbol{I}$, we have the relation that $(d\boldsymbol{H}^t)\boldsymbol{H} + \boldsymbol{H}^t(d\boldsymbol{H}) = \boldsymbol{0}$, or

$$(d\boldsymbol{H}^t)\boldsymbol{H} = -\boldsymbol{H}^t(d\boldsymbol{H}) = -[(d\boldsymbol{H}^t)\boldsymbol{H}]^t,$$

which means that $(d\boldsymbol{H}^t)\boldsymbol{H}$ is a skew symmetric matrix. This fact implies that the diagonal elements of the first and the second terms in the r.h.s. of (2.37) are zero. Letting $d = \partial/\partial v_{ij}$ especially and considering the diagonal elements in (2.37), we can see that

$$\begin{aligned} d_{ij}\ell_s &= \frac{1}{2}(1+\delta_{ij})\frac{\partial \ell_s}{\partial v_{ij}} \\ &= \frac{1}{2}\Big[\boldsymbol{H}^t(\boldsymbol{E}_{ij} + \boldsymbol{E}_{ji})\boldsymbol{H}\Big]_{ss} \\ &= h_{is}h_{js}, \end{aligned} \tag{2.38}$$

where $\boldsymbol{E}_{ij}$ is a $p \times p$ matrix such that the (i,j)-th element is one and others are zero.

Using the equation (2.38), we get that

$$\begin{aligned} d_{ij}f(\boldsymbol{L}) &= \sum_{r=1}^{p}(d_{ij}\ell_r)\frac{\partial}{\partial \ell_r}f(\boldsymbol{L}) \\ &= \sum_{r=1}^{p} f_r h_{ir}h_{jr}, \end{aligned}$$

which is substituted in the second term in the r.h.s. of (2.36) to get that

$$\begin{aligned} \sum_{i=1}^{p}\sum_{s=1}^{p} a_s v_{si}\{d_{is}C_\kappa(\boldsymbol{V})\} &= \sum_{i=1}^{p}\sum_{s=1}^{p} a_s v_{si}\sum_{r=1}^{p} f_r h_{ir}h_{sr} \\ &= \sum_{s=1}^{p}\sum_{r=1}^{p} a_s\{\boldsymbol{V}\boldsymbol{H}\}_{sr}\{\mathrm{diag}\,(f_1,\ldots,f_p)\boldsymbol{H}^t\}_{rs} \qquad (2.39) \\ &= \mathrm{tr}\,\boldsymbol{A}\boldsymbol{V}\boldsymbol{H}\mathrm{diag}\,(f_1,\ldots,f_p)\boldsymbol{H}^t \\ &= \mathrm{tr}\,\Big(\boldsymbol{H}^t\boldsymbol{A}\boldsymbol{H}\Big)\,\mathrm{diag}\,(\ell_1 f_1,\ldots,\ell_p f_p) \\ &\geq 0. \end{aligned}$$

Combining (2.35), (2.36) and (2.39) gives that

$$\begin{aligned} &E\Big[\mathrm{tr}\,\boldsymbol{A}\boldsymbol{V}\boldsymbol{B}^{-1}C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}\Big] \\ &\quad\geq\; E\,[(p+1)(\mathrm{tr}\,\boldsymbol{A})C_\kappa(\boldsymbol{V}) + (n+m-p-1)(\mathrm{tr}\,\boldsymbol{A})C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}] \\ &\quad=\; (n+m)(\mathrm{tr}\,\boldsymbol{A})E[C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}], \end{aligned}$$

which proves (2.33) and the proof of Theorem 2 is complete. □□

3 Estimation of the Generalized Variance

In this section, we treat the problem of estimating the generalized variance $|\boldsymbol{\Sigma}|$ which has been studied as one of the multivariate extensions of the Stein result. The method used in Section 2.1 will be applied in Section 3 not only to construct a new improved estimator of $|\boldsymbol{\Sigma}|$ but also to give another proof of the conventional result given by Shorrock and Zidek (1976) and Sinha (1976). It is supposed that every estimator $\delta = \delta(\boldsymbol{S}, \boldsymbol{X})$ is evaluated in terms of the risk function $R(\omega, \delta) = E_\omega[L(\delta, |\boldsymbol{\Sigma}|)]$ for $\omega = (\boldsymbol{\Sigma}, \boldsymbol{\Xi})$ relative to the Stein (or entropy) loss function

$$L(\delta, |\boldsymbol{\Sigma}|) = \delta/|\boldsymbol{\Sigma}| - \log \delta/|\boldsymbol{\Sigma}| - 1. \tag{3.1}$$

Shorrock and Zidek (1976) and Sinha and Ghosh (1987) showed that the best affine equivariant estimator of $|\boldsymbol{\Sigma}|$ is given by

$$\delta_0 = \frac{(n-p)!}{n!}|\boldsymbol{S}| \tag{3.2}$$

and that it is improved upon by the truncated estimator

$$\delta^{SZ} = \min\left\{\frac{(n-p)!}{n!}|\boldsymbol{S}|,\ \frac{(n+m-p)!}{(n+m)!}|\boldsymbol{S} + \boldsymbol{X}\boldsymbol{X}^t|\right\}. \tag{3.3}$$

Shorrock and Zidek (1976) established this result on the basis of expressing the risk function with the zonal polynomials. Since their approach was somewhat complicated, Sinha (1976) gave another method base on the Fubini-type theorem of Karlin (1960) which derives the distribution of a square root matrix of $\boldsymbol{S}$ with respect to the Lebesgue measure. Using (2.2) and $\boldsymbol{T} = (t_{ij}) \in G_T^+$ such that $\boldsymbol{S} = \boldsymbol{T}\boldsymbol{T}^t$, we see that the estimator δ^{SZ} is rewritten by

$$\delta^{SZ} = \prod_{i=1}^p (n-i+1)^{-1} t_{ii}^2 \times \min\left\{1,\ \prod_{i=1}^p G_i\right\}, \tag{3.4}$$

where

$$G_i = (n-i+1)\frac{1 + \boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{y}_i}{n+m-i+1}. \tag{3.5}$$

Also we can consider another type of estimators which are sequentially defined by

$$\delta_k^{TR} = \prod_{i=1}^p (n-i+1)^{-1} t_{ii}^2 \times \min\left\{1,\ G_1,\ G_1G_2, \ldots, \prod_{j=1}^k G_j\right\}, \tag{3.6}$$

for $k = 1, \ldots, p$. Then the method used in Section 2.1 can be applied to establish that δ^{SZ} dominates δ_0 and that δ_k^{TR} beats δ_{k-1}^{TR} for $k = 1, \ldots, p$. The two improved estimators δ^{SZ} and δ_p^{TR} are possible choice though the preference between them cannot be compared analytically.

Theorem 3.

(1) *The estimators δ^{SZ} dominates the δ_0 relative to the loss (3.1).*

(2) *For $k = 1, \ldots, p$, the truncated estimator δ_k^{TR} dominates δ_{k-1}^{TR} relative to the loss (3.1), where δ_0^{TR} denotes δ_0.*

Proof. We first prove the part (1). Consider an estimator of the general form

$$\delta_\phi = (\prod_{i=1}^{p} e_i t_{ii}^2)\phi(a_{11}, \ldots, a_{pp}),$$

where $e_i = (n-i+1)^{-1}$ and $a_{ii} = 1 + \boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{y}_i$ for $i = 1, \ldots, p$. Since it is supposed that $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ without loss of generality, the risk function of δ_ϕ is written as

$$R(\omega, \delta_\phi) = E\left[\prod_{i=1}^{p} e_i t_{ii}^2 \phi - \log \prod_{i=1}^{p} e_i t_{ii}^2 \phi - 1\right].$$

From (2.12), it is noted that

$$\begin{aligned} & \mathrm{tr}\left\{\boldsymbol{T}(\boldsymbol{I}_p + \boldsymbol{Y}^t\boldsymbol{Y})\boldsymbol{T}^t - 2\boldsymbol{T}\boldsymbol{Y}^t\boldsymbol{\Xi}^t\right\} \\ = & \sum_{i=1}^{p}\left\{a_{ii}t_{ii}^2 - 2\theta_{ii}t_{ii} - k_i(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i)\right\} + \sum_{i=1}^{p-1} a_{ii}||\boldsymbol{t}_{i+1,i} + \boldsymbol{z}_i||^2, \end{aligned} \tag{3.7}$$

where

$$\begin{aligned} k_i(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_j) &= a_{ii}^{-1}\boldsymbol{y}_i^t\boldsymbol{C}_i\boldsymbol{\Xi}_{i+1}\boldsymbol{\Xi}_{i+1}^t\boldsymbol{C}_i\boldsymbol{y}_i, \\ \theta_{ii} &= \boldsymbol{y}_i^t\boldsymbol{C}_i\boldsymbol{\xi}_i. \end{aligned}$$

Hence, integrating out the density with respect to $\boldsymbol{t}_{21}, \ldots, \boldsymbol{t}_{p,p-1}$, we rewrite the risk as

$$\begin{aligned} R(\omega, \delta_\phi) = & \int \cdots \int \left(\prod_{i=1}^{p} e_i t_{ii}^2 \phi - \log \prod_{i=1}^{p} e_i t_{ii}^2 \phi - 1\right) \\ & \times \prod_{i=1}^{p} t_{ii}^{n+m-i} \exp\left\{-\sum_{i=1}^{p}\left\{a_{ii}t_{ii}^2 - 2\theta_{ii}t_{ii} - k_i(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i)\right\}/2\right\} \\ & \times c_2(\boldsymbol{\Xi}, a_{11}, \ldots, a_{pp}) \prod_{i=1}^{p} dt_{ii} d\boldsymbol{Y}. \end{aligned} \tag{3.8}$$

Note that for $i = 1, \ldots, p$ and $j = 1, \ldots, i$,

$$\begin{aligned} \theta_{ii}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_j, \ldots, \boldsymbol{y}_i) &= (-1)^{\delta_{ij}}\theta_{ii}(\boldsymbol{y}_1, \ldots, -\boldsymbol{y}_j, \ldots, \boldsymbol{y}_i), \\ k_i(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_j, \ldots, \boldsymbol{y}_i) &= k_i(\boldsymbol{y}_1, \ldots, -\boldsymbol{y}_j, \ldots, \boldsymbol{y}_i), \end{aligned}$$

where δ_{ij} is the Kronecker's delta. Then the risk can be further rewritten as

$$\begin{aligned} R(\omega,\delta_\phi) &= \int\cdots\int\left(\prod_{i=1}^p e_i t_{ii}^2\phi - \log\prod_{i=1}^p e_i t_{ii}^2\phi - 1\right) \\ &\quad\times\prod_{i=1}^p\left\{\frac{1}{2}\left(e^{\theta_{ii}t_{ii}} + e^{-\theta_{ii}t_{ii}}\right)t_{ii}^{n+m-i}e^{-a_{ii}t_{ii}^2/2}dt_{ii}\right\} \\ &\quad\times c_2(\boldsymbol{\Xi}, a_{11},\ldots,a_{pp})\exp\left\{\sum_{i=1}^p k_i(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_i)/2\right\}d\boldsymbol{Y}, \end{aligned} \tag{3.9}$$

which is minimized at $\phi = \phi_\Xi^*$ where

$$\phi_\Xi^* = \prod_{i=1}^p \frac{\int t_{ii}^{n+m-i}\left(e^{\theta_{ii}t_{ii}} + e^{-\theta_{ii}t_{ii}}\right)e^{-a_{ii}t_{ii}^2/2}dt_{ii}}{e_i\int t_{ii}^{n+m+2-i}\left(e^{\theta_{ii}t_{ii}} + e^{-\theta_{ii}t_{ii}}\right)e^{-a_{ii}t_{ii}^2/2}dt_{ii}}.$$

From (2.11), we get the inequality

$$\phi_\Xi^* \le \prod_{i=1}^p \frac{a_{ii}}{e_i(n+m-i+1)} = \phi_0^* \text{ (say)},$$

so that

$$\phi_\Xi^* \le \min(1, \phi_0^*) \le 1$$

in the case that $\phi_0^* < 1$. Therefore the convexity of the loss (3.1) completes the proof of the first part of Proposition 1.

Next we demonstrate the part (2). Let us define F_k by

$$F_k = \frac{\min\left(1, G_1, \ldots, \prod_{i=1}^{k-1} G_i\right)}{\prod_{i=1}^k G_i}.$$

From (2.10) and (2.12), it can be seen that

$$\begin{aligned} R(\omega,\delta_{k-1}) - R(\omega,\delta_k) &= E\left[\left\{(F_k - 1)\,(\prod_{i=1}^k G_i)(\prod_{i=1}^p e_i t_{ii}^2) - \log F_k\right\}I(F_k \ge 1)\right] \\ &= E\left[\left\{(F_k - 1)\,(\prod_{i=1}^k G_i)B_k^*(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k) - \log F_k\right\}I(F_k \ge 1)\right], \end{aligned}$$

where

$$\begin{aligned} B_k^*(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k) &= B_k^* \\ &= \frac{\int\cdots\int\prod_{i=1}^k e_i t_{ii}^2 f_1(t_{11},\ldots,t_{kk},\boldsymbol{T}_{k+1,k+1},\boldsymbol{Y}_{k+1})\prod_{i=1}^k dt_{ii}d\boldsymbol{T}_{k+1,k+1}d\boldsymbol{Y}_{k+1}}{\int\cdots\int f_1(t_{11},\ldots,t_{kk},\boldsymbol{T}_{k+1,k+1},\boldsymbol{Y}_{k+1})\prod_{i=1}^k dt_{ii}d\boldsymbol{T}_{k+1,k+1}d\boldsymbol{Y}_{k+1}}, \end{aligned}$$

for

$$
\begin{aligned}
f_1(t_{11},\ldots,t_{kk},\boldsymbol{T}_{k+1,k+1},\boldsymbol{Y}_{k+1}) &= \prod_{i=1}^{p}\left\{t_{ii}^{n+m-i}(e^{\theta_{ii}t_{ii}}+e^{-\theta_{ii}t_{ii}})e^{-a_{ii}t_{ii}^2/2}\right\} \\
&\quad\times\exp\{-h_k^*(\boldsymbol{T}_{k+1,k+1},\boldsymbol{Y}_{k+1})/2\}, \\
h_k^*(\boldsymbol{T}_{k+1,k+1},\boldsymbol{Y}_{k+1})=h_k^* &= \mathrm{tr}\left\{\boldsymbol{T}_{k+1,k+1}(\boldsymbol{I}_{p-k}+\boldsymbol{Y}_{k+1}^t\boldsymbol{C}_{k+1}\boldsymbol{Y}_{k+1})\boldsymbol{T}_{k+1,k+1}^t\right. \\
&\quad\left.-2\boldsymbol{T}_{k+1,k+1}\boldsymbol{Y}_{k+1}^t\boldsymbol{C}_{k+1}\boldsymbol{\Xi}_{k+1}\right\}.
\end{aligned}
$$

By the same arguments as in (2.11), we observe that

$$
B_k^* \geq \frac{1}{\prod_{i=1}^k G_i}\frac{\int\cdots\int\prod_{i=k+1}^p e_i t_{ii}^{n+m+2-i}e^{-h_k^*/2}d\boldsymbol{T}_{k+1,k+1}d\boldsymbol{Y}_{k+1}}{\int\cdots\int\prod_{i=k+1}^p t_{ii}^{n+m-i}e^{-h_k^*/2}d\boldsymbol{T}_{k+1,k+1}d\boldsymbol{Y}_{k+1}}. \tag{3.10}
$$

By making the transformation $\boldsymbol{T}_{k+1,k+1}\boldsymbol{Y}_{k+1}^t=\boldsymbol{X}_{k+1}$, the r.h.s. of (3.10) is expressed by

$$
\frac{1}{\prod_{i=1}^k G_i}\frac{\int\cdots\int\prod_{i=k+1}^p e_i t_{ii}^{n+2-i}f_2(\boldsymbol{T}_{k+1,K+1},\boldsymbol{X}_{k+1})d\boldsymbol{T}_{k+1,k+1}d\boldsymbol{X}_{k+1}}{\int\cdots\int\prod_{i=k+1}^p t_{ii}^{n-i}f_2(\boldsymbol{T}_{k+1,K+1},\boldsymbol{X}_{k+1})d\boldsymbol{T}_{k+1,k+1}d\boldsymbol{X}_{k+1}},
$$

which can be easily seen to be $1/\prod_{i=1}^k G_i$, where

$$
\begin{aligned}
&f_2(\boldsymbol{T}_{k+1,K+1},\boldsymbol{X}_{k+1}) \\
&= \exp\left[-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{T}_{k+1,k+1}\boldsymbol{T}_{k+1,k+1}^t+\boldsymbol{X}_{k+1}\boldsymbol{C}_{k+1}\boldsymbol{X}_{k+1}^t-2\boldsymbol{X}_{k+1}\boldsymbol{C}_{k+1}\boldsymbol{\Xi}_{k+1}\right)\right].
\end{aligned}
$$

Therefore we get that $R(\omega,\delta_{k-1})\geq R(\omega,\delta_k)$ for any ω, and the proof of Theorem 3 is complete. □□

4 Simulation Studies

It is of interest to investigate the risk behaviors of several estimators given in the previous sections. We provide the results of Monte Carlo simulation for the risks of the estimators where the values of the risks are given by average values of the loss functions based on 50,000 replications. These are done in the cases where $p=2$, $n=4$, $m=1,10$, $\boldsymbol{\Sigma}=\mathrm{diag}\,(1,1)$, $\xi_{1j}=a/3$ and $\xi_{2j}=a$ for $\boldsymbol{\Xi}=(\xi_{ij})$ and $0\leq a\leq 8$.

The risk performances of estimators of $\boldsymbol{\Sigma}$ are first investigated. For the sake of simplicity, we denote $\widehat{\boldsymbol{\Sigma}}^{UB}$, $\widehat{\boldsymbol{\Sigma}}^{SG}$, $\widehat{\boldsymbol{\Sigma}}^{JS}$, $\widehat{\boldsymbol{\Sigma}}^{TR}$, $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^S)$, $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^S]^{TR})$, $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^T]^{TR})$ and $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^H]^{TR})$ with $a_0=(p-1)/n$ by UB, SG, JS, TR, S*, STR, TTR and HTR, respectively.

Table 1 reports the values of the risks of the estimators UB, SG, JS and TR for $m=1$, $p=2$ and $a=0,0.5,1,2,3,4,5,6,7,8$. In this case, SG, JS and TR are possible candidates since $\widehat{\boldsymbol{\Sigma}}^{HR}$ given in Section 1 is identical to SG.

For $m=10$ and $p=2$, the scale equivariant minimax estimators proposed in Section 2.2 are added to candidates, and the risk behaviors of the estimators UB, JS, TR, S*,

Table 1. Risks of the Estimators UB, SG, JS and TR in Estimation of $\boldsymbol{\Sigma}$ for $m = 1$ and $p = 2$

a	0	0.5	1	2	3	4	5	6	7	8
UB	.925	.925	.925	.925	.925	.925	.925	.925	.925	.925
SG	.922	.922	.923	.924	.925	.925	.925	.925	.925	.925
JS	.861	.861	.861	.861	.861	.861	.861	.861	.861	.861
TR	.839	.839	.840	.844	.850	.853	.855	.856	.857	.858

STR, TTR and HTR are given in Figure 1 for $0 \le a \le 8$ where the risk of $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^P]^{TR})$ is not given there since it behaves similarly to TTR.

Table 1 and Figure 1 reveal that

(1) in the case that $m = 1 < p = 2$, the estimator TR is much better than UB, SG and JS,

(2) in the case that $m = 10 > p = 2$, the estimator HTR is the best of the seven,

(3) STR beats TTR for $0 \le a < 2$ while the reverse hods for $a > 3$,

(4) the risk gain of TR is not so much as the scale equivariant minimax estimators for $m = 10$.

The truncated minimax estimator TR is thus recommended when $m < p$. When $m \ge p$, the estimators HTR, STR and TTR are recommendable for the practical use.

The risk performances in estimation of the generalized variance $|\boldsymbol{\Sigma}|$ are investigated in Figure 2, where δ^{UB}, δ^{SZ} and δ^{TR} are denoted by UB, SZ and TR, respectively. Figure 2 reveals that TR has a smaller risk on a large parameter space while the risk gain of SZ is significant at $\boldsymbol{\Xi} = \mathbf{0}$.

Acknowledgments. The research of the first author was supported in part by a grant from the Center for International Research on the Japanese Economy, the University of Tokyo, and by the Ministry of Education, Japan, Grant Nos. 09780214, 11680320. The research of the second author was supported in part by Natural Sciences and Engineering Research Council of Canada. We are grateful to Mr. M. Ushijima for his help in the simulation experiments.

REFERENCES

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.

Haff, L.R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, **9**, 531-542.

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

Hara, H. (1999). *Estimation of Covariance Matrix and Mean Squared Error for Shrinkage Estimators in Multivariate Normal Distribution.* Doctoral Dissertation, Faculty of Engineering, University of Tokyo.

Iliopoulos, G. and Kourouklis, S. (1999). Improving on the best affine equivariant estimator of the ratio of generalized variances. *J. Multivariate Anal.*, **68**, 176-192.

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

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.

Karlin, S. (1960). Lecture notes on multivariate analysis. Unpublished. Department of Statistics, Stanford University.

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), NOVA Science Publishers, Inc., New York, to appear.

Kubokawa, T. (1999). Shrinkage and modification techniques in estimation of variance and the related problems: A review. *Commun. Statist. - Theory Methods*, **28**, 613-650.

Kubokawa, T., Honda, T., Morita, K. and Saleh, A.K.Md.E. (1993). Estimating a covariance matrix of a normal distribution with unknown mean. *J. Japan Statist. Soc.*, **23**, 131-144.

Kubokawa, T., Robert, C. and Saleh, A.K.Md.E. (1992). Empirical Bayes estimation of the variance parameter of a normal distribution with unknown mean under an entropy loss. *Sankhya* (Ser. A), **54**, 402-410.

Perron, F. (1990). Equivariant estimators of the covariance matrix. *Canadian J. Statist.*, **18**, 179-182.

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

Sarkar, S.K. (1989). On improving the shortest length confidence interval for the generalized variance. *J. Multivariate Anal.*, **31**, 136-147.

Sarkar, S.K. (1991). Stein-type improvements of confidence intervals for the generalized variance. *Ann. Inst. Statist. Math.*, **43**, 369-375.

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.

Shorrock, R.B. and Zidek, J.V. (1976). An improved estimator of the generalized variance. *Ann. Statist.*, **4**, 629-638.

Sinha, B.K. (1976). On improved estimators of the generalized variance. *J. Multivariate Anal.*, **6**, 617-626.

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

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

Srivastava, M.S. and Kubokawa, T. (1999). Improved nonnegative estimation of multivariate components of variance. Unpublished Manuscript.

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. (1969). Multivariate analysis I. Technical Report No. 42, Stanford University.

Stein, C. (1977). Estimating the covariance matrix. Unpublished Manuscript.

Sugiura, N. and Ishibayashi, H. (1997). Reference prior Bayes estimator for bivariate normal covariance matrix with risk comparison. *Commun. Statist.-Theory Methods*, **26**, 2203-2221.

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

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

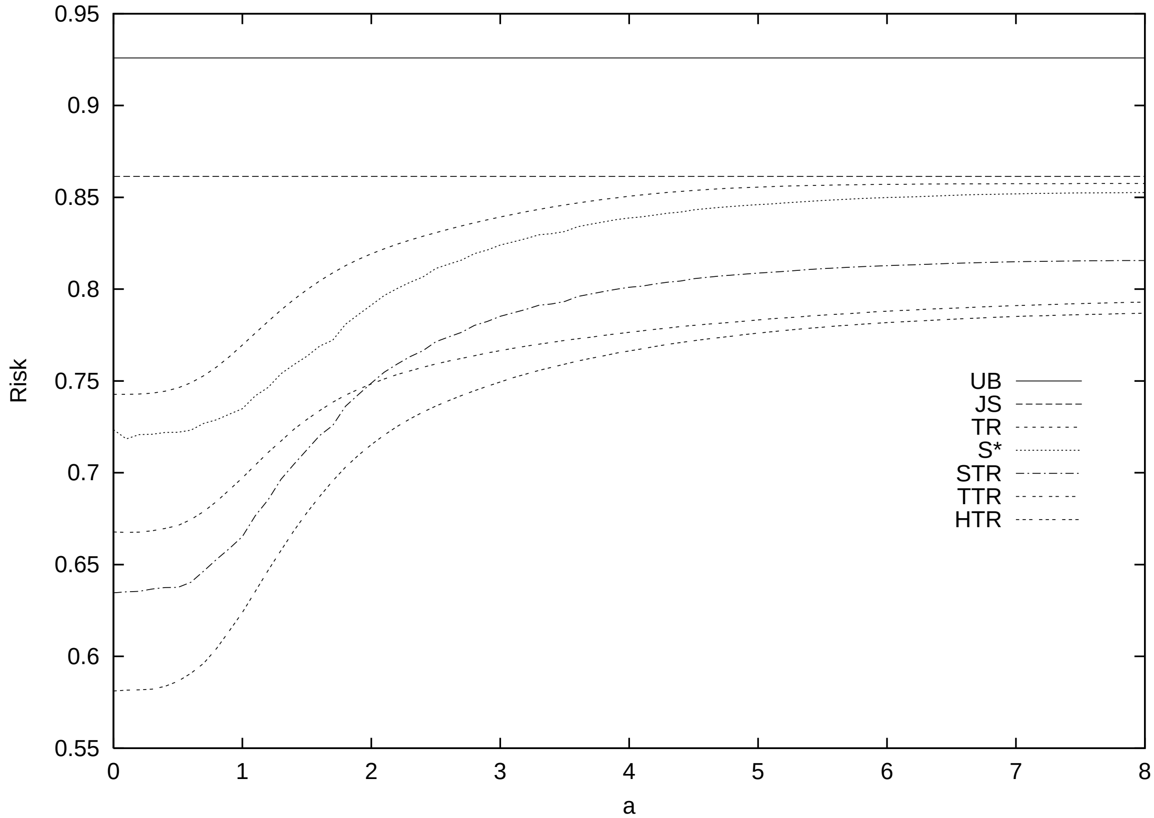

Figure 1. Risks of the Estimator UB, JS, TR, S*, STR, TTR and HTR in Estimation of $\boldsymbol{\Sigma}$ for $m = 10$ and $p = 2$

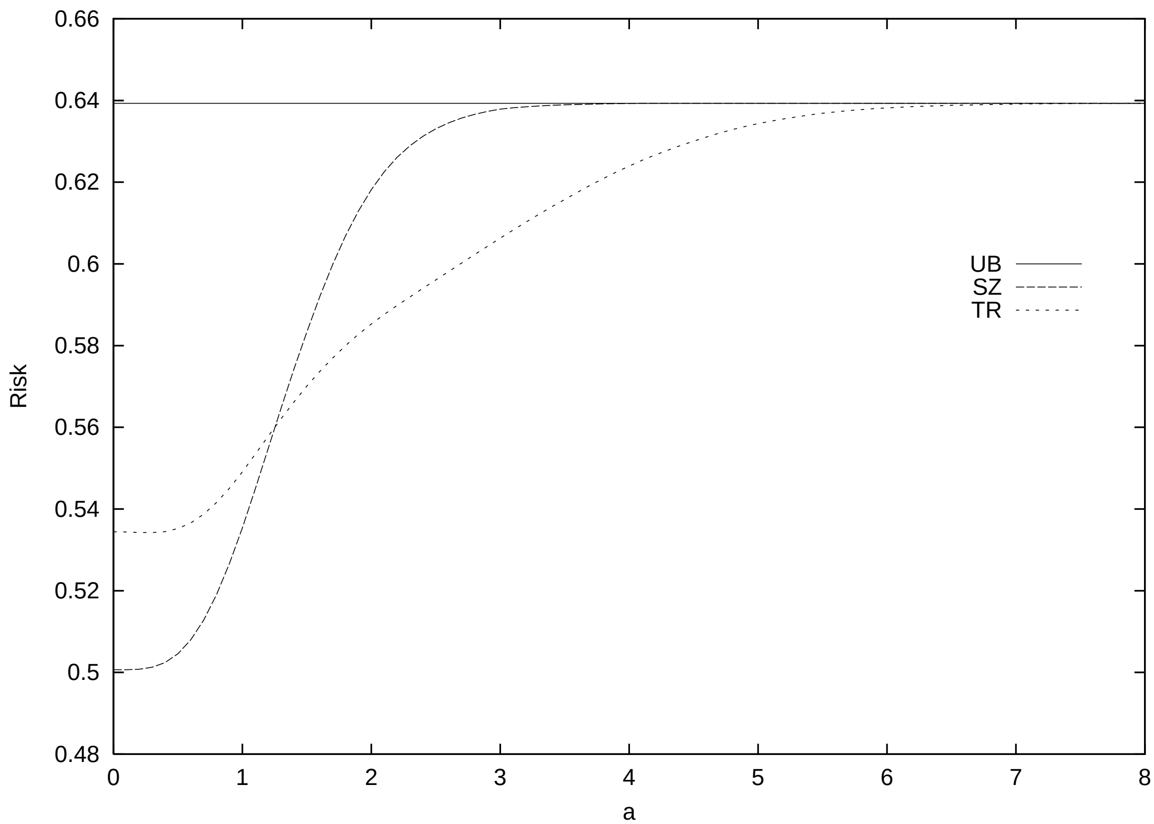

Figure 2. Risks of the Estimators UB, SZ and TR in Estimation of $|\boldsymbol{\Sigma}|$ for $m = 10$ and $p = 2$