CIRJE-F-162

Estimating the Covariance Matrix: A New Approach

Tatsuya Kubokawa
University of Tokyo

M. S. Srivastava
University of Toronto

July 2002

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 $\mathcal{N}_{p,m}(\boldsymbol{\Xi}, \boldsymbol{\Sigma}, \boldsymbol{I}_m)$ 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 positive 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 variance 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 (1980) among others.

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 estimator 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 $\boldsymbol{X}$.

Sinha and Ghosh (1987) provided a truncated estimator of the covariance matrix $\boldsymbol{\Sigma}$ utilizing the information contained in the observation matrix $\boldsymbol{X}$. Hara (1999) recently showed that Sinha and Ghosh's estimator is dominated by

$$\widehat{\boldsymbol{\Sigma}}^{HR} = \boldsymbol{S}^{1/2}\boldsymbol{Q}\operatorname{diag}(\phi_1, \ldots, \phi_p)\boldsymbol{Q}^t\boldsymbol{S}^{1/2} \tag{1.4}$$

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 $\boldsymbol{Q}^t\boldsymbol{S}^{-1/2}\boldsymbol{X}\boldsymbol{X}^t\boldsymbol{S}^{-1/2}\boldsymbol{Q} = \mathrm{diag}\,(\gamma_1, \ldots, \gamma_p)$. Dominance results for $m = 1$ were earlier given by Perron (1990) and Kubokawa *et al.* (1992). 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 $\widehat{\boldsymbol{\Sigma}}^{JS}$ when we utilize both $\boldsymbol{S}$ and $\boldsymbol{X}$ in estimation of $\boldsymbol{\Sigma}$. For this purpose, we introduce a new method. This method is 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 the rank of $\boldsymbol{X}$, $\rho(\boldsymbol{X}) = m \geq p$, another type of minimax improved estimators motivated by Srivastava and Kubokawa (1999) is provided in Subsection 2.2. 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 $\boldsymbol{Y}$ and an $m \times (p - j + 1)$ matrix $\boldsymbol{Y}_j$ by

$$\boldsymbol{Y} = \left(\boldsymbol{T}^{-1}\boldsymbol{X}\right)^t = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p) = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{j-1}, \boldsymbol{Y}_j), \quad \boldsymbol{Y}_j = (\boldsymbol{y}_j, \ldots, \boldsymbol{y}_p),$$

for $j = 2, \ldots, p$. Also for $j = 1, \ldots, p$, define inductively an $m \times m$ matrix $\boldsymbol{C}_j$ based on $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{j-1})$ by

$$\boldsymbol{C}_j = \boldsymbol{C}_{j-1} - (1 + \boldsymbol{y}_{j-1}^t\boldsymbol{C}_{j-1}\boldsymbol{y}_{j-1})^{-1}\boldsymbol{C}_{j-1}\boldsymbol{y}_{j-1}\boldsymbol{y}_{j-1}^t\boldsymbol{C}_{j-1} \tag{2.1}$$

where $\boldsymbol{C}_1 = \boldsymbol{I}_m$. Then it can be shown that

$$|\boldsymbol{I}_p + \boldsymbol{Y}^t\boldsymbol{Y}| = \prod_{i=1}^{p}(1 + \boldsymbol{y}_i^t\boldsymbol{C}_i\boldsymbol{y}_i). \tag{2.2}$$

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

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

where $\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p) = \mathrm{diag}\,(g_1, \ldots, g_p)$ for

$$g_i = g_i(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i) = \min\left\{\frac{1}{n + p + 1 - 2i}, \ \frac{1 + \boldsymbol{y}_i^t\boldsymbol{C}_i\boldsymbol{y}_i}{n + m + p + 1 - 2i}\right\}.$$

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 the sake of convenience, let

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

for $j=2,\dots,p$. $\boldsymbol{T}_1$ corresponds to $\boldsymbol{T}$. For calculating the risk for the Stein loss function given in (1.1), we may assume that $\boldsymbol{\Sigma}=\boldsymbol{I}_p$ without any loss of generality. The risk difference of the two estimators is expressed as

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

where

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

for $a_{ii}=1+\boldsymbol{y}_i^t\boldsymbol{C}_i\boldsymbol{y}_i$ and $d_i^*=(n+m+p+1-2i)^{-1}$.

For the proof of Theorem 1, it is sufficient to show that $\Delta_i\ge 0$ for $i=1,\dots,p$. We shall first show that $\Delta_1\ge 0$. 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 obtained 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 $|\boldsymbol{T}|^m$ respectively, 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 $\boldsymbol{\Theta}_{22}=\boldsymbol{Y}_2^t\boldsymbol{\Xi}_2$. Then we can write the exponent in (2.5) as

$$\begin{aligned}&\mathrm{tr}\left\{\boldsymbol{T}(\boldsymbol{I}_p+\boldsymbol{Y}^t\boldsymbol{Y})\boldsymbol{T}^t-2\boldsymbol{T}\boldsymbol{Y}^t\boldsymbol{\Xi}^t\right\}\\ &= \mathrm{tr}\left\{\begin{pmatrix}t_{11} & \mathbf{0}\\ \boldsymbol{t}_1 & \boldsymbol{T}_2\end{pmatrix}\begin{pmatrix}a_{11} & \boldsymbol{a}_{21}^t\\ \boldsymbol{a}_{21} & \boldsymbol{A}_{22}\end{pmatrix}\begin{pmatrix}t_{11} & \boldsymbol{t}_1^t\\ 0 & \boldsymbol{T}_2^t\end{pmatrix}\right.\end{aligned}$$

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

where $||\boldsymbol{u}||^2 = \boldsymbol{u}^t\boldsymbol{u}$ for a column vector $\boldsymbol{u}$, $\boldsymbol{z}_1 = a_{11}^{-1}(\boldsymbol{T}_2\boldsymbol{Y}_2^t - \boldsymbol{\Xi}_2^t)\boldsymbol{y}_1$,

$$
h_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_2) = \mathrm{tr}\,\boldsymbol{T}_2(\boldsymbol{I}_{p-1} + \boldsymbol{Y}_2^t\boldsymbol{C}_2\boldsymbol{Y}_2)\boldsymbol{T}_2^t - 2\mathrm{tr}\,\boldsymbol{T}_2\boldsymbol{Y}_2^t\boldsymbol{C}_2\boldsymbol{\Xi}_2 - a_{11}^{-1}\boldsymbol{y}_1^t\boldsymbol{\Xi}_2\boldsymbol{\Xi}_2^t\boldsymbol{y}_1,
$$

and $\boldsymbol{C}_2$ is defined in (2.1).

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

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

From the middle expression in the last line of the equation (2.6), and the joint density in (2.5), it follows from that given $\boldsymbol{y}_1$, $\boldsymbol{Y}_2$ and $\boldsymbol{T}_2$, $w_1 = a_{11}\boldsymbol{t}_1^t\boldsymbol{t}_1$ is distributed as noncentral chisquare with $(p-1)$ degrees of freedom and noncentrality parameter $a_{11}\boldsymbol{z}_1^t\boldsymbol{z}_1$. We shall denote this conditional density of w_1 by $f_{p-1}(w_1; a_{11}\boldsymbol{z}_1^t\boldsymbol{z}_1)$. Hence Δ_1 is rewriten as

$$
\begin{aligned}
\Delta_1 = & \int\cdots\int \Big\{(d_1 - d_1^* a_{11})(t_{11}^2 + w_1) - \log d_1/(d_1^* a_{11})\Big\} I(d_1 \geq d_1^* a_{11}) \\
& \times c_1(\boldsymbol{\Xi}, a_{11})(\prod_{i=1}^p t_{ii}^{n+m-i}) e^{\theta_{11}t_{11} - a_{11}t_{11}^2/2} e^{-h_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_2)/2} \\
& \times f_{p-1}(w_1; a_{11}\boldsymbol{z}_1^t\boldsymbol{z}_1) dt_{11} dw_1 d\boldsymbol{T}_2 d\boldsymbol{y}_1 d\boldsymbol{Y}_2 \qquad (2.8)
\end{aligned}
$$

for a positive function $c_1(\boldsymbol{\Xi}, a_{11})$. Note that a_{11}, $\boldsymbol{z}_1^t\boldsymbol{z}_1$ and $h_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_2)$ do not change under the transformation $\boldsymbol{y}_1 \to -\boldsymbol{y}_1$, while θ_{11} changes to $-\theta_{11}$ under the same transformation since $\theta_{11} = \boldsymbol{y}_1^t\boldsymbol{\xi}_1$. Using this argument, we can rewrite Δ_1 as

$$
\begin{aligned}
\Delta_1 = & \int\cdots\int \Big\{(d_1 - d_1^* a_{11})(t_{11}^2 + w_1) - \log d_1/(d_1^* a_{11})\Big\} I(d_1 \geq d_1^* a_{11}) \\
& \times c_1(\boldsymbol{\Xi}, a_{11})(\prod_{i=1}^p t_{ii}^{n+m-i})\frac{1}{2}\left(e^{\theta_{11}t_{11}} + e^{-\theta_{11}t_{11}}\right) e^{-a_{11}t_{11}^2/2} e^{-h_1(\boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_2)/2} \\
& \times f_{p-1}(w_1; a_{11}\boldsymbol{z}_1^t\boldsymbol{z}_1) dt_{11} dw_1 d\boldsymbol{T}_2 d\boldsymbol{y}_1 d\boldsymbol{Y}_2 \qquad (2.9)
\end{aligned}
$$

We shall evaluate (2.9) in two stages, first as a conditional expectation given $\boldsymbol{y}_1$, $\boldsymbol{Y}_2$ and $\boldsymbol{T}_2$. In what follows, we shall only write as conditional expectation without mentioning the above random vector and matrices. Let conditionally v_1 be distributed as χ^2_{n+m} and is independently distributed of w_1 defined above. Then Δ_1 can be expressed as

$$\Delta_1 = c_1^*(\boldsymbol{\Xi}) E\left[E\left[k_1(v_1, w_1) g_1(v_1) | \boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_2\right]\right] \tag{2.10}$$

where $c_1^*(\boldsymbol{\Xi})$ is a constant,

$$\begin{aligned} k_1(v_1, w_1) &= \left\{ (d_1/a_{11} - d_1^*)(v_1 + w_1) - \log \frac{d_1}{d_1^* a_{11}} \right\} I(d_1 \geq d_1^* a_{11}), \\ g_1(v_1) &= \exp\{\theta_{11} \sqrt{v_1/a_{11}}\} + \exp\{-\theta_{11} \sqrt{v_1/a_{11}}\}. \end{aligned}$$

Since $E[w_1 | \boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_2] = p - 1 + a_{11} \boldsymbol{z}_1^t \boldsymbol{z}_1 \geq p - 1$, the conditional expectation in (2.10) is greater than or equal to

$$E\left[k_1(v_1, p-1) g_1(v_1) | \boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_2\right]. \tag{2.11}$$

Noting that both functions $k_1(v_1, p-1)$ and $g_1(v_1)$ are increasing in v_1, we see from Theorem 1.10.5 of Srivastava and Khatri (1979) that

$$\begin{aligned} & E\left[k_1(v_1, p-1) g_1(v_1) | \boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_2\right] \\ \geq \;& E\left[k_1(v_1, p-1) | \boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_2\right] \times E\left[g_1(v_1) | \boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_2\right]. \end{aligned} \tag{2.12}$$

Since $v_1 \sim \chi^2_{n+m}$ conditionally, we have on the set $\{d_1 \geq d_1^* a_{11}\}$,

$$\begin{aligned} E\left[k_1(v_1, p-1) | \boldsymbol{y}_1, \boldsymbol{Y}_2, \boldsymbol{T}_2\right] &= (d_1/a_{11} - d_1^*)(n + m + p - 1) - \log \frac{d_1}{d_1^* a_{11}} \\ &= \frac{d_1}{d_1^* a_{11}} - \log \frac{d_1}{d_1^* a_{11}} - 1 \geq 0. \end{aligned} \tag{2.13}$$

Combining (2.10), (2.11), (2.12) and (2.13) shows that $\Delta_1 \geq 0$. For an alternative proof, see Kubokawa and Srivastava (1999).

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 $i = 2, \ldots, p-1$,

$$\begin{aligned} & \mathrm{tr}\left\{ \boldsymbol{T}(\boldsymbol{I}_p + \boldsymbol{Y}^t \boldsymbol{Y}) \boldsymbol{T}^t - 2 \boldsymbol{T} \boldsymbol{Y}^t \boldsymbol{\Xi}^t \right\} \\ &= \sum_{j=1}^{i} \left\{ a_{jj} t_{jj}^2 - 2 \boldsymbol{y}_j^t \boldsymbol{C}_j \boldsymbol{\xi}_j t_{jj} + a_{jj} ||\boldsymbol{t}_j + \boldsymbol{z}_j||^2 - a_{jj}^{-1} \boldsymbol{y}_j^t \boldsymbol{C}_j \boldsymbol{\Xi}_{j+1} \boldsymbol{\Xi}_{j+1}^t \boldsymbol{C}_j \boldsymbol{y}_j \right\} \\ &\quad + \mathrm{tr}\, \boldsymbol{T}_{i+1} \left(\boldsymbol{I}_{p-i} + \boldsymbol{Y}_{i+1}^t \boldsymbol{C}_{i+1} \boldsymbol{Y}_{i+1} \right) \boldsymbol{T}_{i+1}^t - 2 \mathrm{tr}\, \boldsymbol{T}_{i+1} \boldsymbol{Y}_{i+1}^t \boldsymbol{C}_{i+1} \boldsymbol{\Xi}_{i+1}, \end{aligned} \tag{2.14}$$

where $a_{ii} = 1 + \boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{y}_i$, $\boldsymbol{z}_i = a_{ii}^{-1} \left(\boldsymbol{T}_{i+1} \boldsymbol{Y}_{i+1}^t - \boldsymbol{\Xi}_{i+1}^t \right) \boldsymbol{C}_i \boldsymbol{y}_i$ 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 (2.14). In fact, we observe that

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

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 $\boldsymbol{\Theta}_{i+1,i+1} = \boldsymbol{Y}_{i+1}^t \boldsymbol{C}_i \boldsymbol{\Xi}_{i+1}$. Hence, the left side of the equation (2.14) is equal to the right side of that equation.

Using the expression (2.14), we can write Δ_i given by (2.4) as

$$
\begin{aligned}
\Delta_i &= \int \cdots \int k_i (a_{ii} t_{ii}^2, a_{ii} \boldsymbol{t}_i^t \boldsymbol{t}_i) \\
&\quad \times c_0(\boldsymbol{\Xi}) (\prod_{j=1}^p t_{jj}^{n+m-j}) \exp \left[- \sum_{j=1}^i \left\{ a_{jj} t_{jj}^2 - 2\theta_{jj} t_{jj} + a_{jj} \| \boldsymbol{t}_j + \boldsymbol{z}_j \|^2 \right\} / 2 \right] \\
&\quad \times e^{-h_i/2} (\prod_{j=1}^i dt_{jj} d\boldsymbol{t}_j d\boldsymbol{y}_j) d\boldsymbol{Y}_{i+1} d\boldsymbol{T}_{i+1},
\end{aligned} \tag{2.15}
$$

where

$$
k_i(x, y) = \left\{ (d_i / a_{ii} - d_i^*) (x + y) - \log \frac{d_i}{d_i^* a_{ii}} \right\} I(d_i \geq d_i^* a_{ii}),
$$

$$
\begin{aligned}
h_i &= h_i(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i, \boldsymbol{Y}_{i+1}, \boldsymbol{T}_{i+1}) \\
&= - \sum_{j=1}^i \left\{ a_{jj}^{-1} \boldsymbol{y}_j^t \boldsymbol{C}_j \boldsymbol{\Xi}_{j+1} \boldsymbol{\Xi}_{j+1}^t \boldsymbol{C}_j \boldsymbol{y}_j \right\} \\
&\quad + \operatorname{tr} \boldsymbol{T}_{i+1} \left(\boldsymbol{I}_{p-i} + \boldsymbol{Y}_{i+1}^t \boldsymbol{C}_{i+1} \boldsymbol{Y}_{i+1} \right) \boldsymbol{T}_{i+1}^t - 2 \operatorname{tr} \boldsymbol{T}_{i+1} \boldsymbol{Y}_{i+1}^t \boldsymbol{C}_{i+1} \boldsymbol{\Xi}_{i+1}.
\end{aligned} \tag{2.16}
$$

The same arguments as in the proof of $\Delta_1 \geq 0$ can be used to evaluate Δ_i. Note that given $\boldsymbol{Y}$ and $\boldsymbol{T}_{j+1}$, $\boldsymbol{t}_j$ has $\mathcal{N}_{p-j}(\boldsymbol{z}_j, a_{jj}^{-1})$. Integrating out the integrals in (2.15) with respect to $\boldsymbol{t}_j$ and t_{jj} inductively for $j = 1, \ldots, i-1$, we see that

$$\begin{aligned}\Delta_i &= \int \cdots \int k_i(a_{ii}t_{ii}^2, a_{ii}\boldsymbol{t}_i^t\boldsymbol{t}_i) \\ &\quad \times c_i(\boldsymbol{\Xi}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1})(\prod_{j=i}^{p} t_{jj}^{n+m-j})e^{\theta_{ii}t_{ii}-a_{ii}t_{ii}^2/2-a_{ii}\|\boldsymbol{t}_i+\boldsymbol{z}_i\|^2/2} \\ &\quad \times e^{-h_i/2}dt_{ii}d\boldsymbol{t}_i(\prod_{j=1}^{i} d\boldsymbol{y}_j)d\boldsymbol{Y}_{i+1}d\boldsymbol{T}_{i+1}, \end{aligned} \tag{2.17}$$

for a function $c_i(\boldsymbol{\Xi}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1})$. It is noted that given $\boldsymbol{Y}$ and $\boldsymbol{T}_{i+1}$, $w_i = a_{ii}\boldsymbol{t}_i^t\boldsymbol{t}_i$ is distributed as noncentral chisquare with $(p-i)$ degrees of freedom and noncentrality parameter $a_{ii}\boldsymbol{z}_i^t\boldsymbol{z}_i$. Also note that a_{ii}, $\boldsymbol{z}_i^t\boldsymbol{z}_i$ and $h_i(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_i, \boldsymbol{Y}_{i+1}, \boldsymbol{T}_{i+1})$ do not change under the transformation $\boldsymbol{y}_i \to -\boldsymbol{y}_i$, while θ_{ii} changes to $-\theta_{ii}$ under the same transformation. Hence Δ_i is rewriten as

$$\begin{aligned}\Delta_i &= \int \cdots \int k_i(a_{ii}t_{ii}^2, w_i) \\ &\quad \times c_i(\boldsymbol{\Xi}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1})(\prod_{j=i}^{p} t_{jj}^{n+m-j})\frac{1}{2}\left(e^{\theta_{ii}t_{ii}} + e^{-\theta_{ii}t_{ii}}\right)e^{-a_{ii}t_{ii}^2/2} \\ &\quad \times e^{-h_i/2}f_{p-i}(w_i; a_{ii}\boldsymbol{z}_i^t\boldsymbol{z}_i)dt_{ii}dw_i(\prod_{j=1}^{i} d\boldsymbol{y}_j)d\boldsymbol{Y}_{i+1}d\boldsymbol{T}_{i+1} \end{aligned} \tag{2.18}$$

where $f_{p-i}(w_i; a_{ii}\boldsymbol{z}_i^t\boldsymbol{z}_i)$ is a conditional density of w_i. Finally, Δ_i can be expressed as

$$\Delta_i = c_i^*(\boldsymbol{\Xi})E\left[E\left[k_i(v_i, w_i) \times \left(e^{\theta_{ii}\sqrt{v_i/a_{ii}}} + e^{-\theta_{ii}\sqrt{v_i/a_{ii}}}\right)\Big|\boldsymbol{Y}, \boldsymbol{T}_{i+1}\right]\right], \tag{2.19}$$

where $c_i^*(\boldsymbol{\Xi})$ is a constant and v_i is a random variable such that given $\boldsymbol{Y}$ and $\boldsymbol{T}_{i+1}$, v_i is conditionally independent of w_i and conditionally $v_i \sim \chi^2_{n+m-i+1}$. The same arguments as in (2.11), (2.12) and (2.13) are used to establish that $\Delta_i \geq 0$. 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. When the rank of the $p \times m$ matrix $\boldsymbol{X}$, $\rho(\boldsymbol{X}) = m \geq p$, then we show in this subsection that it is possible to construct truncated equivariant minimax estimators of $\boldsymbol{\Sigma}$. In this subsection, we shall assume that $m \geq p$.

We consider the following equivariant estimators under a scale transformation:

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{H}^t\boldsymbol{ASAH}, \boldsymbol{H}^t\boldsymbol{AXO}) = \boldsymbol{H}^t\boldsymbol{A}\widehat{\boldsymbol{\Sigma}}(\boldsymbol{S}, \boldsymbol{X})\boldsymbol{AH}, \tag{2.20}$$

for any $\boldsymbol{H} \in O(p)$, any $\boldsymbol{O} \in O(m)$ and any $p \times p$ nonsingular symmetric matrix $\boldsymbol{A}$, where $O(p)$ is the group of $p \times p$ orthogonal matrices. Then it can be seen that (2.20) is equivalent to

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{S}, \boldsymbol{X}) = (\boldsymbol{XX}^t)^{1/2}\boldsymbol{H\Psi}(\boldsymbol{H}^t\boldsymbol{FH})\boldsymbol{H}^t(\boldsymbol{XX}^t)^{1/2}, \tag{2.21}$$

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 (2.21) can be expressed by

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}) = (\boldsymbol{X}\boldsymbol{X}^t)^{1/2}\boldsymbol{P}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{P}^t(\boldsymbol{X}\boldsymbol{X}^t)^{1/2} \tag{2.22}$$

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 the value of $\boldsymbol{\Psi}(\boldsymbol{\Lambda}) = \boldsymbol{\epsilon}\boldsymbol{\Psi}(\boldsymbol{\epsilon}\boldsymbol{\Lambda}\boldsymbol{\epsilon})\boldsymbol{\epsilon}$ remains unchanged 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

$$\begin{aligned} [\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR} &= \mathrm{diag}\,(\psi_1^{TR}(\boldsymbol{\Lambda}), \ldots, \psi_p^{TR}(\boldsymbol{\Lambda})), \\ \psi_i^{TR}(\boldsymbol{\Lambda}) &= \min\left\{\psi_i(\boldsymbol{\Lambda}), \frac{\lambda_i + 1}{n+m}\right\}, \quad i = 1, \ldots, p, \end{aligned} \tag{2.23}$$

which gives the corresponding truncated estimator of the form

$$\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR}) = (\boldsymbol{X}\boldsymbol{X}^t)^{1/2}\boldsymbol{P}\mathrm{diag}\,(\psi_1^{TR}(\boldsymbol{\Lambda}), \ldots, \psi_p^{TR}(\boldsymbol{\Lambda}))\boldsymbol{P}^t(\boldsymbol{X}\boldsymbol{X}^t)^{1/2}. \tag{2.24}$$

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 $\boldsymbol{\Psi}$ as for the minimaxity of an orthogonally equivariant estimators based on $\boldsymbol{S}$ only, given by

$$\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}) = \boldsymbol{R}\boldsymbol{\Psi}(\boldsymbol{L}^*)\boldsymbol{R}^t, \tag{2.25}$$

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 eigenvalues $\ell_1^* \geq \ldots \geq \ell_p^*$.

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

$$R(\omega, \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})) = E^X\left[E^{F|X}[\operatorname{tr}\boldsymbol{P}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{P}^t\boldsymbol{\Sigma}_*^{-1} - \log|\boldsymbol{P}\boldsymbol{\Psi}(\boldsymbol{\Lambda})\boldsymbol{P}^t\boldsymbol{\Sigma}_*^{-1}| - p \,\Big|\, \boldsymbol{X}]\right], \tag{2.26}$$

so that given $\boldsymbol{X}$, conditionally $\boldsymbol{P}\boldsymbol{\Psi}\boldsymbol{P}^t$ 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 $\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ implies the minimaxity of $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$, which proves the part (1). The part (2) follows from (2.26) and the results of Sheena and Takemura (1992). □□

Combining Theorem 2 and Proposition 1 gives the following.

Proposition 2. *If an orthogonally equivariant estimator $\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ is minimax, then the truncated estimator $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR})$ is scale-equivariant, minimax and improving on $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ relative to the Stein loss (1.1).*

It should be noted that Proposition 2 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. Proposition 2 states that the minimax estimator can be further improved on by $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR})$ by employing the information in $\boldsymbol{X}$.

From Proposition 1, we can obtain some scale equivariant and minimax estimators by using the results derived previously for the estimation of $\boldsymbol{\Sigma}$. Of these, the Stein type scale equivariant minimax estimator is given by $\widehat{\boldsymbol{\Sigma}}^S = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^S)$ for $\boldsymbol{\Psi}^S(\boldsymbol{\Lambda}) = \operatorname{diag}(d_1\lambda_1, \ldots, d_p\lambda_p)$. The minimaxity of $\widehat{\boldsymbol{\Sigma}}^S$ follows from the result of Dey and Srinivasan (1985). Applying the truncation rule (2.23) to $\boldsymbol{\Psi}^S(\boldsymbol{\Lambda})$ yields the minimax estimator

$$\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^S]^{TR}) \quad \text{for} \quad [\boldsymbol{\Psi}^S]^{TR} = \operatorname{diag}\left(\min\left\{\frac{\lambda_i}{n+p+1-2i}, \frac{\lambda_i+1}{n+m}\right\}, i = 1, \ldots, p\right), \tag{2.27}$$

which improves on the Stein type scale equivariant minimax estimator $\widehat{\boldsymbol{\Sigma}}^S$. The scale equivariant minimax estimators based on estimators of Takemura (1984), Perron (1992) and Sheena and Takemura (1992) and their improved truncated estimators can also be derived, but the details are omitted from this paper; the reader is referred to Kubokawa and Srivastava (1999) for details.

The Haff type scale equivariant estimator is given by

$$\widehat{\boldsymbol{\Sigma}}^H = \frac{1}{n}\left(\boldsymbol{S} + \frac{a_0}{\operatorname{tr}\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{X}^t}\boldsymbol{X}\boldsymbol{X}^t\right). \tag{2.28}$$

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

$\boldsymbol{\Psi}^H = n^{-1}\boldsymbol{\Lambda} + a_0(\mathrm{tr}\,\boldsymbol{\Lambda}^{-1})^{-1}\boldsymbol{I}$. Applying the truncation rule to $\boldsymbol{\Psi}^H$ yields the estimator

$$\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^H]^{TR}) \quad \text{for} \quad [\boldsymbol{\Psi}^H]^{TR} = \mathrm{diag}\left(\min\left\{\frac{\lambda_i}{n} + \frac{a_0}{\mathrm{tr}\,\boldsymbol{\Lambda}^{-1}}, \frac{\lambda_i+1}{n+m}\right\}, i=1,\ldots,p\right), \quad (2.29)$$

which improves on the Haff type scale equivariant estimator $\widehat{\boldsymbol{\Sigma}}^H$.

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

$$\begin{aligned}
&E\left[h(\boldsymbol{F}, \boldsymbol{X}\boldsymbol{X}^t)\right] \\
&= c_0(\boldsymbol{\Xi})\int\int h(\boldsymbol{F}, \boldsymbol{X}\boldsymbol{X}^t)|\boldsymbol{S}|^{(n-p-1)/2}\exp\left\{-\mathrm{tr}\,(\boldsymbol{S}+\boldsymbol{X}\boldsymbol{X}^t - 2\boldsymbol{X}\boldsymbol{\Xi}^t)/2\right\}d\boldsymbol{X}d\boldsymbol{S} \\
&= c_0(\boldsymbol{\Xi})\int\int h(\boldsymbol{F}, \boldsymbol{X}\boldsymbol{X}^t)|\boldsymbol{S}|^{(n-p-1)/2} \qquad (2.30)\\
&\quad \times \exp\left\{-\mathrm{tr}\,(\boldsymbol{S}+\boldsymbol{X}\boldsymbol{X}^t)/2\right\}\int\exp\left\{\mathrm{tr}\,\boldsymbol{X}\boldsymbol{H}\boldsymbol{\Xi}^t/2\right\}\mu(d\boldsymbol{H})d\boldsymbol{X}d\boldsymbol{S},
\end{aligned}$$

where $\mu(d\boldsymbol{H})$ denotes an invariant probability measure on the group of orthogonal matrices. Here the second equality in (2.30) 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 (2.30) is written by

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

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

$$\begin{aligned}
E\left[h(\boldsymbol{F},\boldsymbol{X}\boldsymbol{X}^t)\right] &= c_1(\boldsymbol{\Xi})\int\int h(\boldsymbol{F},\boldsymbol{W})|\boldsymbol{F}|^{(n-p-1)/2}|\boldsymbol{W}|^{(n+m-p-1)/2} \qquad (2.31)\\
&\quad\times\exp\{-\mathrm{tr}\,(\boldsymbol{F}+\boldsymbol{I})\boldsymbol{W}/2\}\sum_\kappa\alpha_\kappa^{(m)}C_\kappa(\boldsymbol{\Xi}\boldsymbol{\Xi}^t\boldsymbol{W})d\boldsymbol{F}d\boldsymbol{W}.
\end{aligned}$$

Again making the transformations $\boldsymbol{F} = \boldsymbol{P\Lambda P}^t$ and $\boldsymbol{W} = \boldsymbol{PVP}^t$ in order, we see that (2.31) is represented as

$$
\begin{aligned}
& E\left[h(\boldsymbol{F}, \boldsymbol{XX}^t)\right] \\
&= c_2(\boldsymbol{\Xi}) \int\int\int h(\boldsymbol{P\Lambda P}^t, \boldsymbol{W}) g(\boldsymbol{\Lambda}) |\boldsymbol{W}|^{(n+m-p-1)/2} \\
&\quad \times \exp\left\{-\mathrm{tr}\,(\boldsymbol{\Lambda}+\boldsymbol{I})\boldsymbol{P}^t\boldsymbol{WP}/2\right\} \sum_\kappa \alpha_\kappa^{(m)} C_\kappa(\boldsymbol{\Xi\Xi}^t\boldsymbol{W}) \mu(d\boldsymbol{P}) d\boldsymbol{\Lambda} d\boldsymbol{W} \\
&= c_2(\boldsymbol{\Xi}) \int\int\int h(\boldsymbol{P\Lambda P}^t, \boldsymbol{PVP}^t) g(\boldsymbol{\Lambda}) |\boldsymbol{V}|^{(n+m-p-1)/2} \qquad (2.32) \\
&\quad \times \exp\{-\mathrm{tr}\,(\boldsymbol{\Lambda}+\boldsymbol{I})\boldsymbol{V}/2\} \sum_\kappa \alpha_\kappa^{(m)} C_\kappa(\boldsymbol{\Xi\Xi}^t\boldsymbol{PVP}^t) \mu(d\boldsymbol{P}) d\boldsymbol{\Lambda} d\boldsymbol{V},
\end{aligned}
$$

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

Based on the expression (2.32), we can evaluate the risk difference of the two estimators, which is given by

$$
\begin{aligned}
\Delta &= R(\omega, \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})) - R(\omega, \widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR})) \\
&= E\left[\mathrm{tr}\left\{\boldsymbol{P\Psi}(\boldsymbol{\Lambda})\boldsymbol{P}^t - \boldsymbol{P}[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\boldsymbol{P}^t\right\}\boldsymbol{W} - \log|\boldsymbol{\Psi}(\boldsymbol{\Lambda})\{[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\}^{-1}|\right] \qquad (2.33) \\
&= c_2(\boldsymbol{\Xi}) \int\int\int \left[\mathrm{tr}\left\{\boldsymbol{\Psi}(\boldsymbol{\Lambda}) - [\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\right\}\boldsymbol{V} - \log|\boldsymbol{\Psi}(\boldsymbol{\Lambda})\{[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\}^{-1}|\right] \\
&\quad \times g(\boldsymbol{\Lambda})|\boldsymbol{V}|^{(n+m-p-1)/2} \\
&\quad \times \exp\{-\mathrm{tr}\,(\boldsymbol{\Lambda}+\boldsymbol{I})\boldsymbol{V}/2\} \sum_\kappa \alpha_\kappa^{(m)} C_\kappa(\boldsymbol{\Xi\Xi}^t\boldsymbol{PVP}^t)\mu(d\boldsymbol{P}) d\boldsymbol{\Lambda} d\boldsymbol{V},
\end{aligned}
$$

where $\boldsymbol{V} = \boldsymbol{P}^t\boldsymbol{WP}$. By the basic property of zonal polynomials,

$$
\int C_\kappa(\boldsymbol{\Xi\Xi}^t\boldsymbol{PVP}^t)\mu(d\boldsymbol{P}) = C_\kappa(\boldsymbol{\Xi\Xi}^t)C_\kappa(\boldsymbol{V})/C_\kappa(\boldsymbol{I}_p). \qquad (2.34)
$$

For simplicity, let us put $\boldsymbol{A} = \{\boldsymbol{\Psi}(\boldsymbol{\Lambda}) - [\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\}(\boldsymbol{\Lambda}+\boldsymbol{I})^{-1}$ and $\boldsymbol{B} = (\boldsymbol{\Lambda}+\boldsymbol{I})^{-1}$. Then from (2.34), it can be seen that

$$
\begin{aligned}
\Delta &= c_2(\boldsymbol{\Xi}) \int\int\int \left[\mathrm{tr}\,\boldsymbol{AVB}^{-1} - \log|\boldsymbol{\Psi}(\boldsymbol{\Lambda})\{[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\}^{-1}|\right] g(\boldsymbol{\Lambda})|\boldsymbol{V}|^{(n+m-p-1)/2} \\
&\quad \times \exp\left\{-\mathrm{tr}\,\boldsymbol{VB}^{-1}/2\right\} \sum_\kappa \alpha_\kappa^{(m)} \frac{C_\kappa(\boldsymbol{\Xi\Xi}^t)C_\kappa(\boldsymbol{V})}{C_\kappa(\boldsymbol{I}_p)} d\boldsymbol{V} d\boldsymbol{\Lambda}. \qquad (2.35)
\end{aligned}
$$

Hence, we can see that $\Delta \geq 0$ if the following inequality is shown:

$$
\begin{aligned}
& \frac{\sum_\kappa \alpha_\kappa^{(n)} b_\kappa \int \mathrm{tr}\,(\boldsymbol{AVB}^{-1}) C_\kappa(\boldsymbol{V})|\boldsymbol{V}|^{(n+m-p-1)/2} \exp\left\{-\mathrm{tr}\,\boldsymbol{VB}^{-1}/2\right\} d\boldsymbol{V}}{\sum_\kappa \alpha_\kappa^{(n)} b_\kappa \int C_\kappa(\boldsymbol{V})|\boldsymbol{V}|^{(n+m-p-1)/2} \exp\left\{-\mathrm{tr}\,\boldsymbol{VB}^{-1}/2\right\} d\boldsymbol{V}} \\
&\qquad \geq \log|\boldsymbol{\Psi}(\boldsymbol{\Lambda})\{[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\}^{-1}|, \qquad (2.36)
\end{aligned}
$$

where $b_\kappa = C_\kappa(\boldsymbol{\Xi}\boldsymbol{\Xi}^t)/C_\kappa(\boldsymbol{I}_p)$. That is, we need to show that

$$\frac{\sum_\kappa \alpha_\kappa^{(n)} b_\kappa E[\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{V}\boldsymbol{B}^{-1})C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}]}{\sum_\kappa \alpha_\kappa^{(n)} b_\kappa E[C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}]} \geq \log|\boldsymbol{\Psi}(\boldsymbol{\Lambda})\{[\boldsymbol{\Psi}(\boldsymbol{\Lambda})]^{TR}\}^{-1}|, \tag{2.37}$$

where conditionally, $\boldsymbol{V}|\boldsymbol{\Lambda} \sim \mathcal{W}_p(n+m, \boldsymbol{B})$.

Here, we shall show that

$$E[\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{V}\boldsymbol{B}^{-1})C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}] \geq E[\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{V}\boldsymbol{B}^{-1})|\boldsymbol{\Lambda}] \cdot E[C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}]. \tag{2.38}$$

Let $\boldsymbol{H}$ be an orthogonal matrix such that $\boldsymbol{V} = \boldsymbol{H}\boldsymbol{D}\boldsymbol{H}^t$ for a diagonal matrix $\boldsymbol{D}$. Then the l.h.s. of (2.38) is written as

$$\begin{aligned} E[\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{V}\boldsymbol{B}^{-1})C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}] &= E[\mathrm{tr}\,(\boldsymbol{H}^t\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{H}\boldsymbol{D})C_\kappa(\boldsymbol{D})|\boldsymbol{\Lambda}] \\ &= E^H\left[E^{D|H}[\mathrm{tr}\,(\boldsymbol{H}^t\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{H}\boldsymbol{D})C_\kappa(\boldsymbol{D})|\boldsymbol{\Lambda}]\right], \end{aligned} \tag{2.39}$$

where $E^{D|H}[\cdot]$ denotes the conditional expectation with respect to $\boldsymbol{D}$ given $\boldsymbol{H}$. Since coefficients of eigenvalues in $C_\kappa(\boldsymbol{D})$ are nonnegative, $C_\kappa(\boldsymbol{D})$ is a monotone increasing function in $\boldsymbol{D}$. Also $\mathrm{tr}\,(\boldsymbol{H}^t\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{H}\boldsymbol{D})$ is a monotone increasing function in $\boldsymbol{D}$ since diagonal elements of $\boldsymbol{H}^t\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{H}$ are nonnegative. Hence Theorem 1.10.5 of Srivastava and Khatri (1979) is applied to get that

$$\begin{aligned} & E^H\left[E^{D|H}[\mathrm{tr}\,(\boldsymbol{H}^t\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{H}\boldsymbol{D})C_\kappa(\boldsymbol{D})|\boldsymbol{\Lambda}]\right] \\ \geq\ & E^H\left[E^{D|H}[\mathrm{tr}\,(\boldsymbol{H}^t\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{H}\boldsymbol{D})|\boldsymbol{\Lambda}] \cdot E^{D|H}[C_\kappa(\boldsymbol{D})|\boldsymbol{\Lambda}]\right] \\ =\ & E^H\left[E^{D|H}[\mathrm{tr}\,(\boldsymbol{H}^t\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{H}\boldsymbol{D})|\boldsymbol{\Lambda}]\right] \cdot E[C_\kappa(\boldsymbol{D})|\boldsymbol{\Lambda}] \\ =\ & E[\mathrm{tr}\,(\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{V})|\boldsymbol{\Lambda}] \cdot E[C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}], \end{aligned} \tag{2.40}$$

since $E^{D|H}[C_\kappa(\boldsymbol{D})|\boldsymbol{\Lambda}]$ does not depend on $\boldsymbol{H}$. We thus obtain the inequality in (2.38); for an alternative method of proving this inequality, see Kubokawa and Srivastava (1999).

Noting that $E[\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{V}\boldsymbol{B}^{-1})|\boldsymbol{\Lambda}] = (n+m)\mathrm{tr}\,\boldsymbol{A}$ and using the inequality (2.38), we see that the l.h.s. of (2.37) is evaluated as

$$\begin{aligned} \frac{E[\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{V}\boldsymbol{B}^{-1})C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}]}{E[C_\kappa(\boldsymbol{V})|\boldsymbol{\Lambda}]} &\geq (n+m)\mathrm{tr}\,\boldsymbol{A} \\ &= \sum_{i=1}^p \left\{\frac{n+m}{\lambda_i+1}\psi_i(\boldsymbol{\Lambda}) - 1\right\} I\left(\frac{n+m}{\lambda_i+1}\psi_i(\boldsymbol{\Lambda}) \geq 1\right). \end{aligned}$$

Since the r.h.s. of (2.37) is written by

$$\sum_{i=1}^p \log\frac{n+m}{\lambda_i+1}\psi_i(\boldsymbol{\Lambda}) I\left(\frac{n+m}{\lambda_i+1}\psi_i(\boldsymbol{\Lambda}) \geq 1\right),$$

the inequality (2.37) is satisfied. Therefore 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 $\delta_0 = \{(n-p)!/n!\}|\boldsymbol{S}|$ and that it is improved upon by the truncated estimator

$$\delta^{SZ} = \min\left\{ \frac{(n-p)!}{n!}|\boldsymbol{S}|,\ \frac{(n+m-p)!}{(n+m)!}|\boldsymbol{S}+\boldsymbol{X}\boldsymbol{X}^t| \right\}. \tag{3.2}$$

Shorrock and Zidek (1976) established this result by expressing the risk function in zonal polynomials. Since their approach was somewhat complicated, Sinha (1976) gave another method based on the distribution of a nonsymmetric 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

$$\delta^{SZ} = \prod_{i=1}^p (n-i+1)^{-1} t_{ii}^2 \times \min\left\{1,\ \prod_{i=1}^p G_i\right\}, \tag{3.3}$$

where

$$G_i = \frac{n-i+1}{n+m-i+1}\left(1 + \boldsymbol{y}_i^t \boldsymbol{C}_i \boldsymbol{y}_i\right). \tag{3.4}$$

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

$$\delta_k^{TR} = \prod_{i=1}^p (n-i+1)^{-1} t_{ii}^2 \times \min\left\{1,\ G_1,\ G_1G_2, \dots,\ \prod_{j=1}^k G_j\right\}, \tag{3.5}$$

for $k = 1, \dots, p$. Then the method used in Subsection 2.1 can be applied to establish that δ^{SZ} dominates δ_0 and that δ_k^{TR} beats δ_{k-1}^{TR} for $k = 1, \dots, p$. The two improved estimators δ^{SZ} and δ_p^{TR} are possible choices 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 $k = 1, \dots, p$, 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). The risk difference of the estimators δ_0 and δ^{SZ} is given by

$$\begin{aligned}\Delta &= R(\omega,\delta_0) - R(\omega,\delta^{SZ})\\ &= E\left[\left\{\prod_{i=1}^p e_i t_{ii}^2 (1-\prod_{i=1}^p G_i) + \log \prod_{i=1}^p G_i\right\} I(\prod_{i=1}^p G_i < 1)\right],\end{aligned}$$

where $e_i = (n-i+1)^{-1}$ for $i = 1,\dots,p$. Using the expression (2.14) gives that

$$\begin{aligned}&\mathrm{tr}\left\{\boldsymbol{T}(\boldsymbol{I}_p + \boldsymbol{Y}^t\boldsymbol{Y})\boldsymbol{T}^t - 2\boldsymbol{T}\boldsymbol{Y}^t\boldsymbol{\Xi}^t\right\} \qquad (3.6)\\ &= \sum_{i=1}^p \left\{a_{ii}t_{ii}^2 - 2\theta_{ii}t_{ii} - k_i(\boldsymbol{y}_1,\dots,\boldsymbol{y}_i)\right\} + \sum_{i=1}^{p-1} a_{ii}||\boldsymbol{t}_i + \boldsymbol{z}_i||^2,\end{aligned}$$

where $a_{ii} = 1 + \boldsymbol{y}_i^t\boldsymbol{C}_i\boldsymbol{y}_i$, $\theta_{ii} = \boldsymbol{y}_i^t\boldsymbol{C}_i\boldsymbol{\xi}_i$, $\boldsymbol{z}_i = a_{ii}^{-1}(\boldsymbol{T}_{i+1}\boldsymbol{Y}_{i+1}^t - \boldsymbol{\Xi}_{i+1}^t)\boldsymbol{C}_i\boldsymbol{y}_i$ and $k_i(\boldsymbol{y}_1,\dots,\boldsymbol{y}_j) = a_{ii}^{-1}\boldsymbol{y}_i^t\boldsymbol{C}_i\boldsymbol{\Xi}_{i+1}\boldsymbol{\Xi}_{i+1}^t\boldsymbol{C}_i\boldsymbol{y}_i$. Note that given $\boldsymbol{Y}$ and $\boldsymbol{T}_{i+1}$, $\boldsymbol{t}_i$ has conditionally $\mathcal{N}_{p-i}(-\boldsymbol{z}_i, a_{ii}^{-1})$. Integrating out the density with respect to $\boldsymbol{t}_1,\dots,\boldsymbol{t}_{p-1}$ in turn, we write the risk differnce Δ as

$$\begin{aligned}\Delta &= \int\cdots\int \left\{\prod_{i=1}^p e_i t_{ii}^2 (1-\prod_{i=1}^p G_i) + \log \prod_{i=1}^p G_i\right\} I(\prod_{i=1}^p G_i < 1) \qquad (3.7)\\ &\quad \times \prod_{i=1}^p t_{ii}^{n+m-i} \exp\left\{-\sum_{i=1}^p \left\{a_{ii}t_{ii}^2 - 2\theta_{ii}t_{ii} - k_i(\boldsymbol{y}_1,\dots,\boldsymbol{y}_i)\right\}/2\right\}\\ &\quad \times c(\boldsymbol{\Xi}, a_{11},\dots,a_{pp}) \prod_{i=1}^p dt_{ii} d\boldsymbol{Y},\end{aligned}$$

for a function $c(\boldsymbol{\Xi}, a_{11},\dots,a_{pp})$. Note that for $i = 1,\dots,p$ and $j = 1,\dots,i$,

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

where δ_{ij} is the Kronecker's delta. Then, similarly to (2.9), the risk difference Δ can be rewritten as

$$\begin{aligned}\Delta &= \int\cdots\int \left\{\prod_{i=1}^p e_i t_{ii}^2 (1-\prod_{i=1}^p G_i) + \log \prod_{i=1}^p G_i\right\} I(\prod_{i=1}^p G_i < 1) \qquad (3.8)\\ &\quad \times \prod_{i=1}^p \left\{\frac{1}{2}\left(e^{\theta_{ii}t_{ii}} + e^{-\theta_{ii}t_{ii}}\right) t_{ii}^{n+m-i} e^{-a_{ii}t_{ii}^2/2} dt_{ii}\right\}\\ &\quad \times c(\boldsymbol{\Xi}, a_{11},\dots,a_{pp}) \exp\left\{\sum_{i=1}^p k_i(\boldsymbol{y}_1,\dots,\boldsymbol{y}_i)/2\right\} d\boldsymbol{Y}.\end{aligned}$$

Letting v_i be a random variable such that given $\boldsymbol{Y}$, v_i is conditionally distributed as $\chi^2_{n+m-i+1}$, we can express the risk difference Δ as

$$\Delta = c^*(\boldsymbol{\Xi}) E\left[\left\{\prod_{i=1}^p e_i \frac{v_i}{a_{ii}} (1-\prod_{i=1}^p G_i) + \log \prod_{i=1}^p G_i\right\} I(\prod_{i=1}^p G_i < 1)\right.$$

$$\times \prod_{i=1}^{p} (e^{\theta_{ii}\sqrt{v_i/a_{ii}}} + e^{-\theta_{ii}\sqrt{v_i/a_{ii}}}) \Big] , \tag{3.9}$$

for a constant $c^*(\boldsymbol{\Xi})$. The same argument as in (2.12) shows that

$$\begin{aligned} & E\Big[(\prod_{i=1}^{p} v_i) \prod_{i=1}^{p} (e^{\theta_{ii}\sqrt{v_i/a_{ii}}} + e^{-\theta_{ii}\sqrt{v_i/a_{ii}}}) \,\Big|\, \boldsymbol{Y} \Big] \\ & = \prod_{i=1}^{p} \Big\{ E\Big[v_i (e^{\theta_{ii}\sqrt{v_i/a_{ii}}} + e^{-\theta_{ii}\sqrt{v_i/a_{ii}}}) \,\Big|\, \boldsymbol{Y} \Big] \Big\} \\ & \le \prod_{i=1}^{p} \Big\{ E[v_i|\boldsymbol{Y}] \times E[e^{\theta_{ii}\sqrt{v_i/a_{ii}}} + e^{-\theta_{ii}\sqrt{v_i/a_{ii}}} \,|\, \boldsymbol{Y}] \Big\} . \end{aligned} \tag{3.10}$$

Also it is seen that

$$E\Big[\prod_{i=1}^{p} e_i v_i / a_{ii} \Big| \boldsymbol{Y} \Big] = (\prod_{i=1}^{p} G_i)^{-1}. \tag{3.11}$$

Combining (3.9), (3.10) and (3.11), we can verify that $\Delta \ge 0$, which completes the proof of the first part of Theorem 3.

For the proof of the part (2), the risk difference can be written by

$$R(\omega, \delta_{k-1}) - R(\omega, \delta_k) = E\Big[\Big\{ (F_k - 1) (\prod_{i=1}^{k} G_i)(\prod_{i=1}^{p} e_i t_{ii}^2) - \log F_k \Big\} I(F_k \ge 1) \Big] ,$$

where

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

By using the same arguments as in the proof of (1), the risk difference can be expressed as

$$c^*(\boldsymbol{\Xi}) E\Big[\Big\{ (F_k - 1) (\prod_{i=1}^{k} G_i)(\prod_{i=1}^{p} e_i \frac{v_i}{a_{ii}}) - \log F_k \Big\} I(F_k \ge 1) \prod_{i=1}^{p} (e^{\theta_{ii}\sqrt{v_i/a_{ii}}} + e^{-\theta_{ii}\sqrt{v_i/a_{ii}}}) \Big] ,$$

which can be shown to be nonnegative from (3.10) and (3.11). Therefore, the part (2) is proved 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 results for $p = 2$ of a 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 $n = 4$, $m = 1, 10$, $\boldsymbol{\Sigma} = \mathrm{diag}\,(1, 1)$, $\xi_{1j} = a/3$ and $\xi_{2j} = a$ for $\boldsymbol{\Xi} = (\xi_{ij})$ and $0 \le a \le 8$.

Table 1. Risks of the Estimators UB, HR, 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
HR	.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

The risk performances of estimators of $\boldsymbol{\Sigma}$ are first investigated. For the sake of simplicity, the estimators $\widehat{\boldsymbol{\Sigma}}^{HR}$, $\widehat{\boldsymbol{\Sigma}}^{JS}$, $\widehat{\boldsymbol{\Sigma}}^{TR}$, $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^S]^{TR})$ and $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^H]^{TR})$ with $a_0 = (p-1)/n$, given by (1.4), (1.2), (2.3), (2.27) and (2.29), are denoted by HR, JS, TR, STR and HTR, respectively. Also denote the unbiased estimator $\widehat{\boldsymbol{\Sigma}}^{UB}$ by UB.

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

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 JS, TR, STR and HTR are given in Figure 1 for $0 \le a \le 8$.

Table 1 and Figure 1 reveal that

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

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

(3) the risk gain of TR is not as much as the scale equivariant minimax estimators STR and HTR for $m = 10$, $p = 2$.

The truncated minimax estimator TR is thus recommended when $m < p$. When $m \ge p$, the estimators HTR and STR are recommended for 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, 13680371. The research of the second author was supported in part by Natural Sciences and Engineering Research Council of Canada. The authors thank Mr. M. Ushijima for his help in the simulation experiments and the referee for his helpful comments and suggestions that improved the article.

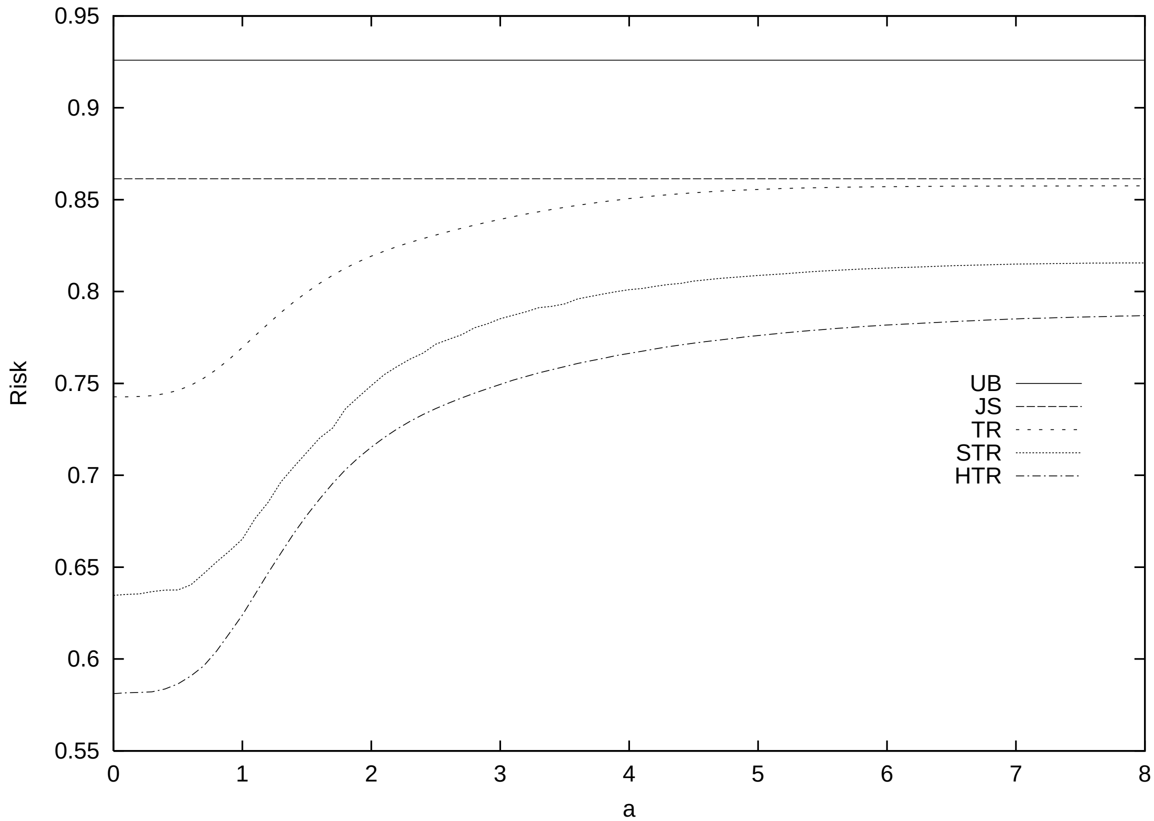

Figure 1. Risks of the Estimator UB, JS, TR, STR and HTR in Estimation of $\boldsymbol{\Sigma}$ for $m = 10$ and $p = 2$

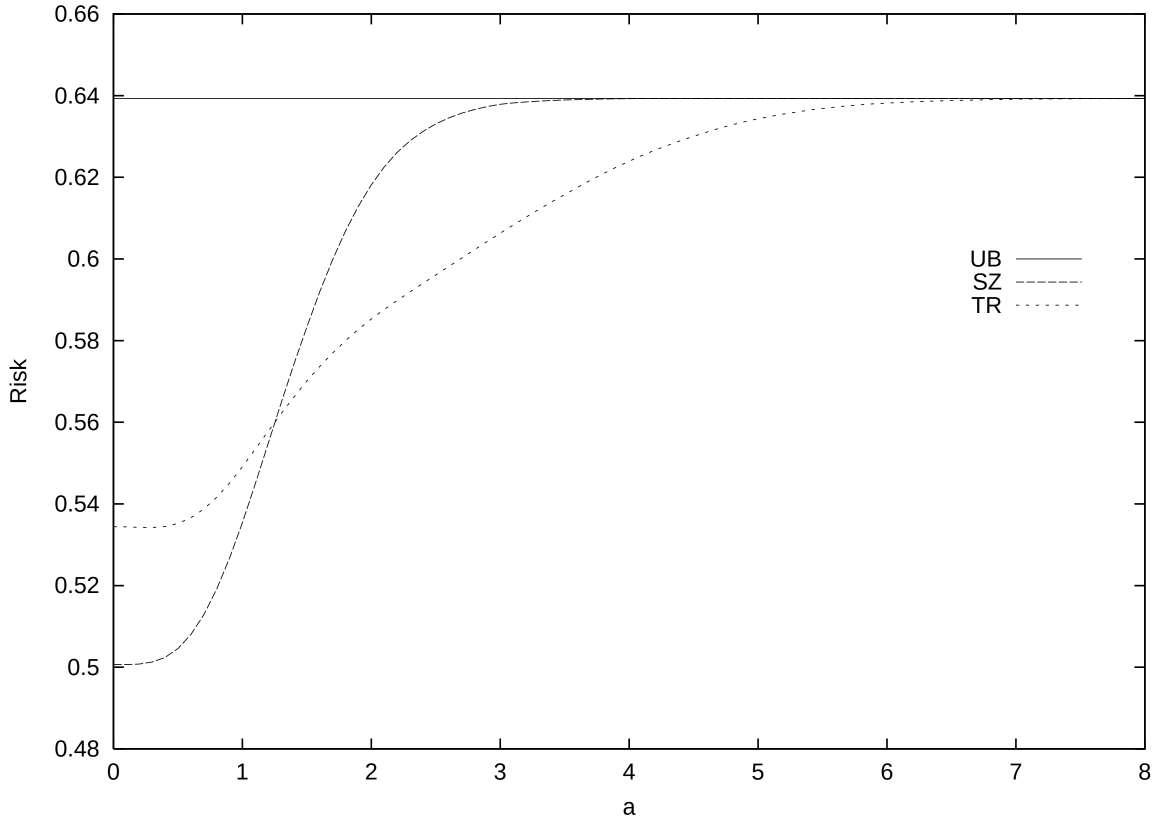

Figure 2. Risks of the Estimators UB, SZ and TR in Estimation of $|\boldsymbol{\Sigma}|$ for $m = 10$ and $p = 2$

REFERENCES

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

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.

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.

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.

Kubokawa, T. and Srivastava, M.S. (1999). Estimating the covariance matrix: A new approach. Discussion Paper CIRJE-F-52, Faculty of Economics, The University of Tokyo.

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.

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. *Ann. Statist.*, **27**, 2008-2032.

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.

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