# A Revisit to Estimation of the Precision Matrix of the Wishart Distribution

Tatsuya Kubokawa*
*University of Tokyo*

March 4, 2004

The estimation of the precision matrix of the Wishart distribution is one of classical problems studied in a decision-theoretic framework and is related to estimation of mean and covariance matrices of a multivariate normal distribution. This paper revisits the estimation problem of the precision matrix and investigates how it connects with the theory of the covariance estimation from a decision-theoretic aspect. To evaluate estimators in terms of risk functions, we employ two kinds of loss functions: the non-scale-invariant loss and the scale-invariant loss functions which are induced from estimation of means. Using the same methods as in the estimation of the covariance matrix, we derive not only the James-Stein type and the Stein type estimators dominating the unbiased estimator, but also a new type of estimators improving on the Stein type one under the non-scale-invariant loss. It is observed that dominance properties given in the estimation of the covariance matrix do not necessarily hold in our setup under the non-scale-invariant loss, but still hold relative to the scale-invariant loss. The simulation studies are given, and estimators having superior risk performances are proposed.

*Key words and phrases:* Covariance matrix, decision theory, empirical Bayes procedure, James-Stein estimator, mean matrix, minimaxity, precision matrix, shrinkage estimation.

## 1 Introduction

In the context of the empirical Bayes estimation, Efron and Morris (1976) showed that the problem of estimating a mean matrix of a multivariate normal distribution can be reduced to that of estimating a precision matrix of a Wishart distribution. The precision is the inverse of covariance, and this fact suggests that the estimation of the mean matrix is related to that of the covariance matrix. For the estimation of the covariance matrix, on the other hand, several decision-theoretic results have been developed in a literature. Of these, James and Stein (1961) established under Stein's loss function, referred to as the entropy loss as well, that the best scalar multiple estimator is not minimax and

*Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan E-Mail: tatsuya@e.u-tokyo.ac.jp

derived a minimax estimator, called the James-Stein estimator, based on the Bartlett decomposition. Stein (1977) showed that the James-Stein minimax estimator is dominated by Stein's orthogonally equivariant estimator, which can be further dominated by the order-preserving estimator as shown by Sheena and Takemura (1992). Our main concern is whether or not these decision-theoretic properties hold in the estimation of the precision matrix under the loss functions induced from the estimation of means.

To describe the problem specifically, let $\boldsymbol{W}$ be a $p \times p$ random matrix having the Wishart distribution $\mathcal{W}_p(m, \boldsymbol{\Sigma})$ with $n$ degrees of freedom and $E[\boldsymbol{W}] = m\boldsymbol{\Sigma}$. Consider the problem of estimating the precision matrix $\boldsymbol{\Sigma}^{-1}$ based on $\boldsymbol{W}$ relative to the loss function

$$\begin{aligned} L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma}) &= \mathrm{tr}\,(\boldsymbol{\delta} - \boldsymbol{\Sigma}^{-1})^2 \boldsymbol{W} \\ &= \mathrm{tr}\,\boldsymbol{W}\boldsymbol{\delta}^2 - 2\mathrm{tr}\,\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\Sigma}^{-1} + \mathrm{tr}\,\boldsymbol{W}\boldsymbol{\Sigma}^{-2} \end{aligned} \tag{1.1}$$

for an estimator $\boldsymbol{\delta}$ of $\boldsymbol{\Sigma}^{-1}$. Efron and Morris (1976) induced this loss function from the estimation of the mean matrix in the context of the empirical Bayes estimation. The derivation of the loss function is explained later through a random effect model. When an estimator is evaluated in terms of the risk function relative to the loss function $L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma})$, the best estimator among multiples of $\boldsymbol{W}$ is given by

$$\boldsymbol{\delta}_0 = a_0 \boldsymbol{W}^{-1}, \quad a_0 = m - p - 1,$$

which is also an unbiased estimator of $\boldsymbol{\Sigma}^{-1}$ with the risk

$$R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}_0) = E[L_1(\boldsymbol{\delta}_0, \boldsymbol{\Sigma})] = (p+1)\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}.$$

Using the similar methods as in the estimation theory of the covariance matrix, we want to construct estimators of $\boldsymbol{\Sigma}^{-1}$ having uniformly smaller risks than $\boldsymbol{\delta}_0$ under the loss $L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma})$.

We begin with addressing the issue of deriving a James-Stein type estimator improving on $\boldsymbol{\delta}_0$. In the estimation of the covariance matrix, the James-Stein estimator is the best within the class of estimators equivariant under transformation with respect to triangular matrices, but in our estimation problem, there does not exist the best, since the risk function of any equivariant estimator depends on unknown parameters. This means that decision-theoretic properties developed in the estimation of the covariance matrix do not necessarily hold in our setup. In Section 2, we derive a James-Stein type estimator within the class of equivariant estimators as a feasible one improving on $\boldsymbol{\delta}_0$. It is shown that the maximum value of the risk of this James-Stein type estimator is equal to that of $\boldsymbol{\delta}_0$, which suggests that the unbiased estimator $\boldsymbol{\delta}_0$ would be minimax, though we could not verify it analytically.

A drawback of the James-Stein type estimator is that it depends on a coordinate system, which leads us to considering orthogonally equivariant estimators. In Section 3, we derive the Stein type orthogonally equivariant estimators improving on $\boldsymbol{\delta}_0$. However, it seems difficult to verify that they dominate the James-Stein type estimator. In Section 3.2, we obtain a new type of orthogonally equivariant estimator which dominates the Stein type estimator. The risk behaviors are numerically investigated in Section 4 to compare the risk behaviors of all the estimators derived in this paper and some estimators given in the literature.

As stated above, the decision-theoretic properties in our estimation problem relative to the non-scale-invariant loss $L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ are slightly different from those in the estimation of the covariance matrix. However, we have a different story when the scale-invariant loss function is employed:

$$
\begin{aligned}
L_2(\boldsymbol{\delta}, \boldsymbol{\Sigma}) &= \mathrm{tr}\,(\boldsymbol{\delta} - \boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}(\boldsymbol{\delta} - \boldsymbol{\Sigma}^{-1})\boldsymbol{W} \\
&= \mathrm{tr}\,\boldsymbol{\delta}\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\Sigma} - 2\mathrm{tr}\,\boldsymbol{\delta}\boldsymbol{W} + \mathrm{tr}\,\boldsymbol{W}\boldsymbol{\Sigma}^{-1}.
\end{aligned}
$$

Section 5 treats the estimation problem of $\boldsymbol{\Sigma}^{-1}$ relative to the scale-invariant loss $L_2(\boldsymbol{\delta}, \boldsymbol{\Sigma})$. The non-minimaxity of $\boldsymbol{\delta}_0$ and the minimaxity of the James-Stein type estimator are demonstrated, and the domination of the Stein type estimator over the James-Stein one is verified for $p = 2$ with a simple method. For $p = 3$, the dominance result follows from Sheena (2003), who recently succeeded in establishing the same property relative to the entropy loss, though it is still open for $p \geq 4$. Finally, we derive an Efron-Morris type estimator superior to $\boldsymbol{\delta}_0$ and investigate their risk behaviors numerically.

We conclude this section with explaining that the loss functions $L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ and $L_2(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ are induced from the following simple prediction problem. Consider a one-way layout random effect model with equal replications:

$$
\boldsymbol{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}, \tag{1.2}
$$

for $i = 1, \ldots, k$ and $j = 1, \ldots, r$, where $p$-dimensional random vectors $\boldsymbol{\epsilon}_{ij}$'s and $\boldsymbol{\alpha}$'s are mutually independently distributed as $\boldsymbol{\epsilon}_{ij} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_E)$ and $\boldsymbol{\alpha}_i \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_A)$. The grand mean $\boldsymbol{\mu}$ and the 'between' component of covariance $\boldsymbol{\Sigma}_A$ are unknown parameters while the 'within' component of covariance $\boldsymbol{\Sigma}_E$ is assumed to be known in this paper for the sake of simplicity. It is supposed that we want to predict the quantities $\boldsymbol{\theta}_i = \boldsymbol{\mu} + \boldsymbol{\alpha}_i$, which are related to the realized means for the individual small areas in the field of small area statistics. The best linear unbiased predictor of $\boldsymbol{\theta}_i$ is given by

$$
\widehat{\boldsymbol{\theta}}_i^B = \overline{\boldsymbol{y}}_i - \boldsymbol{\Sigma}_E\boldsymbol{\Sigma}_2^{-1}(\overline{\boldsymbol{y}}_i - \boldsymbol{\mu})
$$

where $\overline{\boldsymbol{y}}_i = \sum_{j=1}^r \boldsymbol{y}_{ij}/r$ and $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_E + r\boldsymbol{\Sigma}_A$. The grand mean $\boldsymbol{\mu}$ is estimated by $\widehat{\boldsymbol{\mu}} = \sum_{i=1}^k \sum_{j=1}^r \boldsymbol{y}_{ij}/(rk)$, and $\boldsymbol{\Sigma}_2$ is estimated by estimator $\widehat{\boldsymbol{\Sigma}}_2$ based on the statistic $\widetilde{\boldsymbol{W}} = r\sum_{i=1}^k (\overline{\boldsymbol{y}}_i - \widehat{\boldsymbol{\mu}})(\overline{\boldsymbol{y}}_i - \widehat{\boldsymbol{\mu}})^t$, which is distributed as $\mathcal{W}_p(m, \boldsymbol{\Sigma}_2)$ for $m = k - 1$. The resulting predictor is written by

$$
\widehat{\boldsymbol{\theta}}_i^{EB} = \overline{\boldsymbol{y}}_i - \boldsymbol{\Sigma}_E\widehat{\boldsymbol{\Sigma}}_2^{-1}(\overline{\boldsymbol{y}}_i - \widehat{\boldsymbol{\mu}}),
$$

called *the estimated best linear unbiased predictor*, and our concern is how we should estimate $\boldsymbol{\Sigma}_2^{-1}$, which determines how much $\overline{\boldsymbol{y}}_i$ be shrunken towards the total mean $\widehat{\boldsymbol{\mu}}$.

When the predictors are evaluated relative to the loss function

$$
\sum_{i=1}^k (\widehat{\boldsymbol{\theta}}_i^{EB} - \boldsymbol{\theta}_i)^t \boldsymbol{\Sigma}_E^{-1} (\widehat{\boldsymbol{\theta}}_i^{EB} - \boldsymbol{\theta}_i), \tag{1.3}
$$

it can be seen that the risk function is expressed as

$$r^{-1}E\left[\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\Sigma}_2^{-1})\boldsymbol{\Sigma}_E(\widehat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\Sigma}_2^{-1})\boldsymbol{W}\right]+r^{-1}[pk-(k-1)\mathrm{tr}\,\boldsymbol{\Sigma}_E\boldsymbol{\Sigma}_2^{-1}].$$

Let $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_E^{-1/2}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_E^{-1/2}=\boldsymbol{I}+r\boldsymbol{\Sigma}_E^{-1/2}\boldsymbol{\Sigma}_A\boldsymbol{\Sigma}_E^{-1/2}$, $\boldsymbol{W}=\boldsymbol{\Sigma}_E^{-1/2}\widetilde{\boldsymbol{W}}\boldsymbol{\Sigma}_E^{-1/2}$ and $\widehat{\boldsymbol{\Sigma}}=\boldsymbol{\Sigma}_E^{-1/2}\widehat{\boldsymbol{\Sigma}}_2\boldsymbol{\Sigma}_E^{-1/2}$. Then, the problem of predicting $\boldsymbol{\theta}_i$'s with $\widehat{\boldsymbol{\theta}}_i^{EB}$'s can be reduced to that of estimating $\boldsymbol{\Sigma}^{-1}$ relative to the non-scale-invariant loss $L_1(\boldsymbol{\delta},\boldsymbol{\Sigma})$ based on the statistic $\boldsymbol{W}$ having $\mathcal{W}_p(m,\boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}\geq\boldsymbol{I}$. When the predictors $\widehat{\boldsymbol{\theta}}_i^{EB}$'s are evaluated under the loss

$$\sum_{i=1}^k(\widehat{\boldsymbol{\theta}}_i^{EB}-\boldsymbol{\theta}_i)^t\boldsymbol{\Sigma}_E^{-1}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_E^{-1}(\widehat{\boldsymbol{\theta}}_i^{EB}-\boldsymbol{\theta}_i), \tag{1.4}$$

the risk can be written as

$$r^{-1}E\left[\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\Sigma}_2^{-1})\boldsymbol{\Sigma}_2(\widehat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\Sigma}_2^{-1})\boldsymbol{W}\right]+r^{-1}[k\mathrm{tr}\,\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_E^{-1}-p(k-1)],$$

so that the problem is reduced to estimation of $\boldsymbol{\Sigma}^{-1}$ relative to the scale-invariant loss $L_2(\boldsymbol{\delta},\boldsymbol{\Sigma})$.

When we consider a multiple of $\boldsymbol{W}^{-1}$ as an estimator of $\boldsymbol{\Sigma}^{-1}$, the risk function under the loss $L_1(\boldsymbol{\delta},\boldsymbol{\Sigma})$ is written by

$$\begin{aligned}R_1(\boldsymbol{\Sigma},a\boldsymbol{W}^{-1})&=E\left[\mathrm{tr}\,\left(a\boldsymbol{W}^{-1}-\boldsymbol{\Sigma}^{-1}\right)^2\boldsymbol{W}\right]\\&=E\left[a^2\mathrm{tr}\,\boldsymbol{W}^{-1}-2a\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}+\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\boldsymbol{\Sigma}^{-1}\right]\\&=\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\left\{(m-p-1)^{-1}a^2-2a+m\right\},\end{aligned}$$

which is minimized at $a=m-p-1$. For $m>p+1$, let

$$\boldsymbol{\delta}_0=a_0\boldsymbol{W}^{-1},\quad a_0=m-p-1,$$

and it has the risk $R(\boldsymbol{\Sigma},\boldsymbol{\delta}_0)=(p+1)\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}$, which is less than or equal to $m\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}$ for $m>p+1$. This implies that the crude predictor $\overline{\boldsymbol{Y}}=(\overline{\boldsymbol{y}}_1,\ldots,\overline{\boldsymbol{y}}_k)$ is dominated by the shrinkage predictor $\overline{\boldsymbol{Y}}(\boldsymbol{I}-(m-p-1)\boldsymbol{W}^{-1})$ under the loss (1.3) for $m=k-1>p+1$, since $\overline{\boldsymbol{Y}}$ corresponds to $\boldsymbol{\delta}=\boldsymbol{0}$ in the estimation of $\boldsymbol{\Sigma}^{-1}$.

# 2 James-Stein type estimator

James and Stein (1961) provided a method based on the Bartlett decomposition to improve on $\widehat{\boldsymbol{\Sigma}}_0^{-1}$. Let $\mathcal{G}_T^+(p)$ be a set of $p\times p$ lower triangular matrices with positive diagonal elements. By the Bartlett decomposition, we have $\boldsymbol{W}=\boldsymbol{T}\boldsymbol{T}^t$ for $\boldsymbol{T}\in\mathcal{G}_T^+(p)$. Based on the idea of James and Stein (1961), we consider an estimator of the form

$$\boldsymbol{\delta}^{JS}(\boldsymbol{C})=(\boldsymbol{T}^t)^{-1}\boldsymbol{C}\boldsymbol{T}^{-1},\quad \boldsymbol{C}=\mathrm{diag}\,(c_1,\ldots,c_p),$$

where $c_i$'s are positive constants suitably chosen later. The risk function of $\boldsymbol{\delta}^{JS}(\boldsymbol{C})$ and its unbiased estimator are given in the flowing proposition. Let $\boldsymbol{\Sigma}_*^{-1}=(\sigma_*^{ij})=(\widetilde{\boldsymbol{B}}^t\widetilde{\boldsymbol{B}})^{-1}$ for $\widetilde{\boldsymbol{B}}\in\mathcal{G}_T^+(p)$ satisfying $\boldsymbol{\Sigma}=\widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{B}}^t$.

**Proposition 2.1** *The risk function of the estimator $\boldsymbol{\delta}^{JS}(\boldsymbol{C})$ is expressed by*

$$R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}^{JS}(\boldsymbol{C})) - m\mathrm{tr}\,\boldsymbol{\Sigma}^{-1} = \sum_{i=1}^{p} \sigma_*^{ii} \left\{ \sum_{j=i}^{p} \frac{c_j^2}{\prod_{k=i}^{j}(m-k-1)} - 2c_i \right\} \tag{2.1}$$

$$= \sum_{i=1}^{p} \tau_i \left\{ c_i^2 - 2(m-i-1)c_i + 2c_{i+1} \right\}, \tag{2.2}$$

*where $\tau_i = (\tau_{i-1} + \sigma_*^{ii})/(m-i-1)$, $\tau_0 = 0$ and $d_{p+1} = 0$. An unbiased estimator of the risk $R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}^{JS}(\boldsymbol{C}))$ is given by*

$$\widehat{R}_1(\boldsymbol{W}, \boldsymbol{\delta}^{JS}(\boldsymbol{C})) = m(m-p-1)\mathrm{tr}\,\boldsymbol{W}^{-1} + \sum_{i=1}^{p} w_*^{ii} \left\{ c_i^2 - 2(m-i-1)c_i + 2c_{i+1} \right\}, \tag{2.3}$$

*where $w_*^{ii}$ is the $(i,i)$-th element of $(\boldsymbol{T}^t\boldsymbol{T})^{-1}$.*

**Proof.** For the convenience, let $Q(\boldsymbol{\Sigma}, \boldsymbol{\delta}) = R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}) - m\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}$, which is written by

$$Q(\boldsymbol{\Sigma}, \boldsymbol{\delta}) = E_\omega \left[ \mathrm{tr}\,\boldsymbol{\delta}^2\boldsymbol{W} - 2\mathrm{tr}\,\boldsymbol{\delta}\boldsymbol{W}\boldsymbol{\Sigma}^{-1} \right]. \tag{2.4}$$

For the estimator $\boldsymbol{\delta}^{JS} = \boldsymbol{\delta}^{JS}(\boldsymbol{C})$, it is rewritten as

$$\begin{aligned} Q(\boldsymbol{\Sigma}, \boldsymbol{\delta}^{JS}) &= E_\Sigma \left[ \mathrm{tr}\,\boldsymbol{\delta}^{JS}\boldsymbol{W}\boldsymbol{\delta}^{JS} - 2\mathrm{tr}\,\boldsymbol{\delta}^{JS}\boldsymbol{W}\boldsymbol{\Sigma}^{-1} \right] \\ &= E_{\Sigma=I} \left[ \mathrm{tr}\,\boldsymbol{\delta}^{JS}\boldsymbol{W}\boldsymbol{\delta}^{JS}\boldsymbol{\Sigma}_*^{-1} - 2\mathrm{tr}\,\boldsymbol{\delta}^{JS}\boldsymbol{W}\boldsymbol{\Sigma}_*^{-1} \right]. \end{aligned} \tag{2.5}$$

Let us decompose $\boldsymbol{T}$, $\boldsymbol{C}$ and $\boldsymbol{\Sigma}_*$ as

$$\boldsymbol{T} = \begin{pmatrix} t_{11} & \boldsymbol{0}^t \\ \boldsymbol{t}_{21} & \boldsymbol{T}_{22} \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} c_1 & \boldsymbol{0}^t \\ \boldsymbol{0} & \boldsymbol{C}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma}_*^{-1} = \begin{pmatrix} \sigma_*^{11} & \boldsymbol{\sigma}_*^{12} \\ \boldsymbol{\sigma}_*^{21} & \boldsymbol{\Sigma}_*^{22} \end{pmatrix}$$

for scalar $t_{11}$, $c_1$ and $\sigma_*^{11}$. Noted that $t_{11}$, $\boldsymbol{t}_{21}$ and $\boldsymbol{T}_{22}$ are mutually independently distributed as $t_{11}^2 \sim \chi_m^2$ and $\boldsymbol{t}_{21} \sim \mathcal{N}_{p-1}(\boldsymbol{0}, \boldsymbol{I}_{p-1})$. Then,

$$\begin{aligned} &E_I \left[ \mathrm{tr}\,\boldsymbol{\delta}^{JS}\boldsymbol{W}\boldsymbol{\delta}^{JS}\boldsymbol{\Sigma}_*^{-1} \right] \\ =&E_I \left[ \mathrm{tr}\,(\boldsymbol{T}^{-1})^t\boldsymbol{C}^2\boldsymbol{T}^{-1}\boldsymbol{\Sigma}_*^{-1} \right] \\ =&E_I \left[ \mathrm{tr} \begin{pmatrix} t_{11}^{-1} & -t_{11}^{-1}\boldsymbol{t}_{21}^t(\boldsymbol{T}_{22}^{-1})^t \\ \boldsymbol{0} & (\boldsymbol{T}_{22}^{-1})^t \end{pmatrix} \begin{pmatrix} c_1 & \boldsymbol{0}^t \\ \boldsymbol{0} & \boldsymbol{C}_2 \end{pmatrix} \right. \\ &\left. \times \begin{pmatrix} t_{11}^{-1} & \boldsymbol{0}^t \\ -t_{11}^{-1}\boldsymbol{T}_{22}^{-1}\boldsymbol{t}_{21} & \boldsymbol{T}_{22}^{-1} \end{pmatrix} \begin{pmatrix} \sigma_*^{11} & \boldsymbol{\sigma}_*^{12} \\ \boldsymbol{\sigma}_*^{21} & \boldsymbol{\Sigma}_*^{22} \end{pmatrix} \right] \\ =&E_I \left[ \sigma_*^{11}t_{11}^{-2} \left( c_1^2 + \boldsymbol{t}_{21}^t(\boldsymbol{T}_{22}^t)^{-1}\boldsymbol{C}_2^2\boldsymbol{T}_{22}^{-1}\boldsymbol{t}_{21} \right) + \mathrm{tr}\,(\boldsymbol{T}_{22}^{-1})^t\boldsymbol{C}_2^2\boldsymbol{T}_{22}^{-1}\boldsymbol{\Sigma}_*^{22} \right] \\ =&E_I \left[ \frac{\sigma_*^{11}}{m-2} \left( c_1^2 + \boldsymbol{t}_{21}^t(\boldsymbol{T}_{22}^{-1})^t\boldsymbol{C}_2^2\boldsymbol{T}_{22}^{-1}\boldsymbol{t}_{21} \right) + \mathrm{tr}\,(\boldsymbol{T}_{22}^{-1})^t\boldsymbol{C}_2^2\boldsymbol{T}_{22}^{-1}\boldsymbol{\Sigma}_*^{22} \right]. \end{aligned}$$

Repeating this argument gives that

$$
\begin{aligned}
&E_I\left[\operatorname{tr}(\boldsymbol{T}^{-1})^t\boldsymbol{C}^2\boldsymbol{T}^{-1}\boldsymbol{\Sigma}_*^{-1}\right]\\
=&E_I\Bigg[\frac{\sigma_*^{11}}{m-2}\left\{c_1^2+\frac{1}{m-3}\left(c_2^2+\operatorname{tr}(\boldsymbol{T}_{33}^{-1})^t\boldsymbol{C}_3^2\boldsymbol{T}_{33}^{-1}\right)\right\}\\
&\qquad+\frac{\sigma_*^{22}}{m-3}\left(c_2^2+\operatorname{tr}(\boldsymbol{T}_{33}^{-1})^t\boldsymbol{C}_3^2\boldsymbol{T}_{33}^{-1}\right)+\operatorname{tr}(\boldsymbol{T}_{33}^{-1})^t\boldsymbol{C}_3^2\boldsymbol{T}_{33}^{-1}\boldsymbol{\Sigma}_*^{33}\Bigg]\\
=&\cdots\quad\cdots\\
=&\sum_{i=1}^p\sigma_*^{ii}\left\{\prod_{j=i}^p\frac{c_j^2}{\prod_{k=i}^j(m-k-1)}\right\},
\end{aligned}
\tag{2.6}
$$

where $\boldsymbol{T}_{33}$, $\boldsymbol{C}_3$ and $\boldsymbol{\Sigma}_*^{33}$ are the $(p-2)\times(p-2)$ lower right submatrices of $\boldsymbol{T}$, $\boldsymbol{C}$ and $\boldsymbol{\Sigma}_*^{-1}$. Applying the same argument to the term $E_{\Sigma=I}\left[\operatorname{tr}\boldsymbol{\delta}^{JS}\boldsymbol{W}\boldsymbol{\Sigma}_*^{-1}\right]$ yields that

$$
\begin{aligned}
&E_I\left[\operatorname{tr}\boldsymbol{T}\boldsymbol{C}\boldsymbol{T}^{-1}\boldsymbol{\Sigma}_*^{-1}\right]\\
=&E_I\Bigg[\operatorname{tr}\begin{pmatrix} t_{11} & \boldsymbol{0}^t\\ \boldsymbol{t}_{21} & \boldsymbol{T}_{22}\end{pmatrix}\begin{pmatrix} c_1 & \boldsymbol{0}^t\\ \boldsymbol{0} & \boldsymbol{C}_2\end{pmatrix}\\
&\qquad\times\begin{pmatrix} t_{11}^{-1} & \boldsymbol{0}^t\\ -t_{11}^{-1}\boldsymbol{T}_{22}^{-1}\boldsymbol{t}_{21} & \boldsymbol{T}_{22}^{-1}\end{pmatrix}\begin{pmatrix} \sigma_*^{11} & \boldsymbol{\sigma}_*^{12}\\ \boldsymbol{\sigma}_*^{21} & \boldsymbol{\Sigma}_*^{22}\end{pmatrix}\Bigg]\\
=&E_I\left[c_1\sigma_*^{11}+t_{11}^{-1}\left(c_1\boldsymbol{\sigma}_*^{12}\boldsymbol{t}_{21}-\boldsymbol{\sigma}_*^{12}\boldsymbol{T}_{22}\boldsymbol{C}_2\boldsymbol{T}_{22}^{-1}\boldsymbol{t}_{21}\right)+\operatorname{tr}\boldsymbol{T}_{22}\boldsymbol{C}_2\boldsymbol{T}_{22}^{-1}\boldsymbol{\Sigma}_*^{22}\right]\\
=&c_1\sigma_*^{11}+E_I\left[\operatorname{tr}\boldsymbol{T}_{22}\boldsymbol{C}_2\boldsymbol{T}_{22}^{-1}\boldsymbol{\Sigma}_*^{22}\right]\\
=&\sum_{i=1}^p c_i\sigma_*^{ii}.
\end{aligned}
\tag{2.7}
$$

Combining (2.6) and (2.7), we get the expression (2.1) of the risk function in Proposition 2.1. The expression (2.2) can be obtained by putting $\tau_i=(\tau_{i-1}+\sigma_*^{ii})/(m-i-1)$.

For deriving the unbiased estimator (2.3), note that

$$
\begin{aligned}
E_\Sigma[(\boldsymbol{T}^t\boldsymbol{T})^t]&=E_I[\boldsymbol{T}^{-1}\boldsymbol{\Sigma}(\boldsymbol{T}^{-1})^t]\\
&=\begin{pmatrix}\frac{\sigma_*^{11}}{m-2} & *\\ * & E[\frac{\sigma_*^{11}}{m-2}(\boldsymbol{T}_{22}^t\boldsymbol{T}_{22})^{-1}+\boldsymbol{T}_{22}^{-1}\boldsymbol{\Sigma}_*^{22}(\boldsymbol{T}_{22}^{-1})^t]\end{pmatrix}\\
&=\begin{pmatrix}\frac{\sigma_*^{11}}{m-2} & & *\\ & \left(\frac{\sigma_*^{11}}{m-2}+\sigma_*^{22}\right)\frac{1}{m-3} & \\ * & & A_{33}(\boldsymbol{\Sigma}),\end{pmatrix}
\end{aligned}
$$

where

$$
A_{33}(\boldsymbol{\Sigma})=E\left[\left(\frac{\sigma_*^{11}}{m-2}+\sigma_*^{22}\right)\frac{1}{m-3}(\boldsymbol{T}_{33}^t\boldsymbol{T}_{33})^{-1}+\boldsymbol{T}_{33}^{-1}\boldsymbol{\Sigma}_*^{33}(\boldsymbol{T}_{33}^{-1})^t\right].
$$

Repeating this argument, we can notice the equation

$$
E[w_*^{ii}]=\left\{E[w_*^{i-1,i-1}]+\sigma_*^{ii}\right\}/(m-i-1)=\tau_i,
$$

where $w_*^{ii} = [(\boldsymbol{T}^t\boldsymbol{T})^{-1}]_{ii}$, the $(i,i)$ element of $[(\boldsymbol{T}^t\boldsymbol{T})^{-1}]$. Replacing $\tau_i$ with $w_*^{ii}$ in the risk function (2.2), we obtain the unbiased estimator (2.3), and Proposition 2.1 is proved. ■

From this proposition, we see that the optimal $c_1$ is $c_1 = m-2$ while the optimal $c_i$ depends on unknown parameters for $i \geq 2$. A reasonable choice of $c_i$ is given by $c_i = m-i-1$ for $i = 1, \ldots, p$, and hereafter, the notation $c_i$ means $c_i = m-i-1$. Denote the James-Stein type estimator by

$$\boldsymbol{\delta}_c^{JS} = \boldsymbol{\delta}^{JS}(\boldsymbol{C}), \quad \text{for} \quad c_i = m-i-1. \tag{2.8}$$

Noting that for $i = 1, \ldots, p-1$,

$$c_i^2 - 2(m-i-1)c_i + 2c_{i+1} \leq c_p^2 - 2(m-i-1)c_p + 2c_p,$$

we observe that

$$R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}_c^{JS}) \leq \sum_{i=1}^p \tau_i \left\{c_p^2 - 2(m-i-1)c_p + 2c_p\right\} = R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}_0),$$

which implies the domination of $\boldsymbol{\delta}_c^{JS}$ over $\boldsymbol{\delta}_0$.

**Proposition 2.2** *The James-Stein type estimator $\boldsymbol{\delta}_c^{JS}$ dominates $\boldsymbol{\delta}_0$.*

In the context of estimation of the covariance matrix under the Stein loss, it is well known that the James-Stein estimator is minimax with a constant risk, which means that the unbiased estimator is not minimax. However, this decision-theoretic property does not hold in our problem as shown in the following proposition.

**Proposition 2.3** *The estimators $\boldsymbol{\delta}_c^{JS}$ and $\boldsymbol{\delta}_0$ have the same maximum risk under the loss function $L_1^*(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \mathrm{tr}\,(\boldsymbol{\delta} - \boldsymbol{\Sigma}^{-1})^2\boldsymbol{W}/\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}$, which is given by*

$$\sup_{\Sigma}\left\{E[L_1^*(\boldsymbol{\delta}_c^{JS}, \boldsymbol{\Sigma})]\right\} = E[L_1^*(\boldsymbol{\delta}_0, \boldsymbol{\Sigma})] = p(p+1).$$

**Proof**. Note that $\sigma_*^{ii}/\mathrm{tr}\,\boldsymbol{\Sigma}^{-1} \leq 1$. Then from Proposition 2.1, it is seen that

$$\sup_{\Sigma}\left\{E[L_1^*(\boldsymbol{\delta}_c^{JS}, \boldsymbol{\Sigma})]\right\} = p \times \max_i\{A_i\},$$

where

$$A_i = \sum_{j=i}^p \frac{(m-j-1)^2}{\prod_{k=i}^j (m-k-1)} - 2(m-i-1) + m.$$

It suffices to show that $\max_i A_i = p+1$. Clearly, $A_p = p+1$. For $i \leq p-1$, we observe that

$$\begin{aligned} A_i &= \sum_{j=i+1}^p \frac{(m-j-1)^2}{\prod_{k=i}^j (m-k-1)} + i + 1 \\ &= \sum_{j=i+1}^p \left\{\frac{1}{\prod_{k=i}^{j-2}(m-k-1)} - \frac{1}{\prod_{k=i}^{j-1}(m-k-1)}\right\} + i + 1 \\ &= 1 - \frac{1}{\prod_{k=i}^{p-1}(m-k-1)} + i + 1 \leq i + 2, \end{aligned}$$

where $\prod_{k=i}^{i-1}(m-k-1)$ is equal to one. Since $i+2 \leq p+1$ for $i \leq p-1$, it is seen that $\max_i A_i = p+1$, proving Proposition 2.3. ■

# 3 Stein type estimators and further dominance results

A drawback of the James-Stein type estimator $\boldsymbol{\delta}_c^{JS}$ is that it depends on a coordinate system, and it is reasonable to consider orthogonally equivariant estimators. One of them is the Stein type estimator of the form

$$\boldsymbol{\delta}_c^{S} = \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{C}\boldsymbol{H}^t, \quad \boldsymbol{C} = \mathrm{diag}\,(c_1, \dots, c_p), \quad c_i = m - i - 1, \tag{3.1}$$

where $\boldsymbol{H}$ is a $p \times p$ orthogonal matrix and $\boldsymbol{L} = \mathrm{diag}\,(\ell_1, \dots, \ell_p)$, $\ell_1 \geq \dots \geq \ell_p$, such that $\boldsymbol{W} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t$. Although Stein's orthogonally equivariant estimator dominates the James-Stein estimator in the estimation of the covariance matrix, our estimation issue possesses a different story that the Stein type estimator $\boldsymbol{\delta}_c^{S}$ is not always better than the James-Stein type one $\boldsymbol{\delta}_c^{JS}$ because the risk function of $\boldsymbol{\delta}_c^{JS}$ depends on the unknown parameters based on the coordinate system. Thus, we here obtain Stein type estimators improving on the unbiased one $\boldsymbol{\delta}_0$ and develop further dominance results over the Stein type estimators.

## 3.1 Stein type estimators

We begin with deriving conditions under which the unbiased estimator $\boldsymbol{\delta}_0$ is improved on by orthogonally equivariant estimators of the general form

$$\begin{aligned} \boldsymbol{\delta}(\boldsymbol{\Phi}) &= \boldsymbol{H}\boldsymbol{\Phi}(\boldsymbol{\ell})\boldsymbol{H}^t, \quad \boldsymbol{\ell} = (\ell_1, \dots, \ell_p)^t \\ \boldsymbol{\Phi}(\boldsymbol{\ell}) &= \mathrm{diag}\,(\phi_1(\boldsymbol{\ell}), \dots, \phi_p(\boldsymbol{\ell}))\,. \end{aligned} \tag{3.2}$$

The risk function of $\boldsymbol{\delta}(\boldsymbol{\Phi})$ under the loss $L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ is expressed as

$$R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}(\boldsymbol{\Phi})) - m\mathrm{tr}\,\boldsymbol{\Sigma}^{-1} = E\left[\mathrm{tr}\,\{\boldsymbol{\delta}(\boldsymbol{\Phi})\}^2\boldsymbol{W} - 2\mathrm{tr}\,\boldsymbol{\delta}(\boldsymbol{\Phi})\boldsymbol{W}\boldsymbol{\Sigma}^{-1}\right]. \tag{3.3}$$

The Stein-Haff identity given by Stein (1977) and Haff (1979a) shows that

$$E\left[\mathrm{tr}\,\boldsymbol{\delta}(\boldsymbol{\Phi})\boldsymbol{W}\boldsymbol{\Sigma}^{-1}\right] = E\left[(m-p-1)\mathrm{tr}\,\boldsymbol{\delta}(\boldsymbol{\Phi}) + 2\mathrm{tr}\,\boldsymbol{D}_W[\boldsymbol{\delta}(\boldsymbol{\Phi})\boldsymbol{W}]\right], \tag{3.4}$$

where $\boldsymbol{D}_W = (\mathrm{d}_{ij})$ is a $p \times p$ matrix of differential operators $\mathrm{d}_{ij}$'s which are given by $\mathrm{d}_{ij} = 2^{-1}(1 + \delta_{ij})(\partial/\partial w_{ij})$ for the Kronecker delta $\delta_{ij}$ and $\boldsymbol{W} = (w_{ij})$. Following Stein (1977) and Haff (1982), we have that

$$\mathrm{tr}\,\boldsymbol{D}_W[\boldsymbol{H}\boldsymbol{\Phi}(\boldsymbol{\ell})\boldsymbol{L}\boldsymbol{H}^t] = \sum_{i=1}^{p}\left\{\frac{1}{2}\sum_{j\neq i}\frac{\ell_i\phi_i(\boldsymbol{\ell}) - \ell_j\phi(\boldsymbol{\ell})}{\ell_i - \ell_j} + \frac{\partial[\ell_i\phi_i(\boldsymbol{\ell})]}{\partial\ell_i}\right\}. \tag{3.5}$$

Combining (3.3), (3.4) and (3.5), we get the following expression of the risk function.

**Proposition 3.1** *The risk function of the orthogonally equivariant estimator $\boldsymbol{\delta}(\boldsymbol{\Phi})$ is expressed by*

$$\begin{aligned}
&R_1(\boldsymbol{\Sigma},\boldsymbol{\delta}(\boldsymbol{\Phi})) - m\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\\
&\quad= \sum_{i=1}^{p} E\left[\ell_i\phi_i^2 - 2(m-p-1)\phi_i - 2\sum_{j\neq i}\frac{\ell_i\phi_i - \ell_j\phi_j}{\ell_i-\ell_j} - 4\frac{\partial(\ell_i\phi_i)}{\partial\ell_i}\right]\\
&\quad= \sum_{i=1}^{p} E\left[\frac{\psi_i^2}{\ell_i} - 2(m-p-1)\frac{\psi_i}{\ell_i} - 2\sum_{j\neq i}\frac{\psi_i-\psi_j}{\ell_i-\ell_j} - 4\frac{\partial\psi_i}{\partial\ell_i}\right],
\end{aligned}$$

*where $\boldsymbol{\Phi}(\boldsymbol{\ell}) = \boldsymbol{L}^{-1}\boldsymbol{\Psi}(\boldsymbol{\ell}) = \mathrm{diag}\,(\psi_1(\boldsymbol{\ell})/\ell_1, \dots, \psi_p(\boldsymbol{\ell})/\ell_p)$ for $\psi_i = \ell_i\phi_i$.*

From this proposition, we get a sufficient condition for improving on the unbiased estimator $\boldsymbol{\delta}_0$.

**Proposition 3.2** *The estimator $\boldsymbol{\delta}(\boldsymbol{\Phi})$ dominates $\boldsymbol{\delta}_0$ relative to the loss $L_1(\boldsymbol{\Sigma},\boldsymbol{\delta})$ if $\psi_i(\boldsymbol{\ell})$'s satisfy the inequality*

$$\sum_{i=1}^{p}\left\{\frac{\psi_i^2}{\ell_i} - 2(m-p-1)\frac{\psi_i}{\ell_i} - 2\sum_{j\neq i}\frac{\psi_i-\psi_j}{\ell_i-\ell_j} - 4\frac{\partial\psi_i}{\partial\ell_i}\right\} \leq -\sum_{i=1}^{p}\frac{(m-p-1)^2}{\ell_i}.$$

**Proposition 3.3** *Assume that $\boldsymbol{\Psi}(\boldsymbol{\ell}) = \mathrm{diag}\,(\psi_1(\boldsymbol{\ell}), \dots, \psi_p(\boldsymbol{\ell}))$ satisfies the following conditions for $m > p+1$:*

(a) $\partial\psi_i(\boldsymbol{\ell})/\partial\ell_i \geq 0$ *for* $i = 1, \dots, p$.

(b) $\psi_1(\boldsymbol{\ell}) \geq \cdots \geq \psi_p(\boldsymbol{\ell}) = m-p-1$.

(c) $m+p-2i-1 \geq \psi_i(\boldsymbol{\ell})$ *for each* $i$.

*Then the estiamtor $\boldsymbol{\delta}(\boldsymbol{L}^{-1}\boldsymbol{\Psi}) = \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{\Psi}(\boldsymbol{\ell})\boldsymbol{H}^t$ dominates the unbiased estimator $\boldsymbol{\delta}_0$ under the loss $L_1(\boldsymbol{\delta},\boldsymbol{\Sigma})$.*

**Proof.** Note that

$$\begin{aligned}
\sum_i\sum_{j\neq i}\frac{\psi_i-\psi_j}{\ell_i-\ell_j} &= 2\sum_i\sum_{j>i}\frac{\psi_i-\psi_j}{\ell_i-\ell_j}\\
&= 2\sum_i\frac{1}{\ell_i}\sum_{j>i}(\psi_i-\psi_j) + 2\sum_i\frac{1}{\ell_i}\sum_{j>i}\frac{\ell_j(\psi_i-\psi_j)}{\ell_i-\ell_j},
\end{aligned}$$

and that $\sum_{j>i}(\psi_i-\psi_j) = (p-i)\psi_i - \sum_{j>i}\psi_j$. Then, the r.h.s. of the inequality in Proposition 3.2 is expressed by

$$\sum_{i=1}^{p}\frac{1}{\ell_i}\left\{\psi_i^2 - 2(m+p-2i-1)\psi_i + 4\sum_{j>i}\psi_j\right\} - 4\sum_{i=1}^{p}\sum_{j>i}\frac{\ell_j(\psi_i-\psi_j)}{\ell_i(\ell_i-\ell_j)} + \frac{\partial\psi_i}{\partial\ell_i}. \qquad (3.6)$$

The conditions (b) and (c) imply that

$$
\begin{aligned}
&\psi_i^2 - 2(m+p-2i-1)\psi_i + 4\sum_{j>i}\psi_j \\
&\quad \le \psi_{i+1}^2 - 2(m+p-2i-1)\psi_{i+1} + 4\psi_{i+1} + 4\sum_{j>i+1}\psi_j \\
&\quad = \psi_{i+1}^2 - 2\{m+p-2(i+1)-1\}\psi_{i+1} + 4\sum_{j>i+1}\psi_j \\
&\quad \le \cdots \le \psi_p^2 - 2(m-p-1)\psi_p. \qquad (3.7)
\end{aligned}
$$

Combining (3.6) and (3.7), we see that the estimator $\boldsymbol{\delta}(\boldsymbol{L}^{-1}\boldsymbol{\Psi})$ has a uniformly smaller risk than $\boldsymbol{\delta}_0$ under the conditions in Proposition 3.3. ■

For the Stein type estimator $\boldsymbol{\delta}_c^S$ with $c_i = m - i - 1$, it is easily checked to satisfy the conditions in Proposition 3.3. Also, we can consider another estimator with constants $b_i = m + p - 2i - 1$ for $i = 1, \ldots, p$:

$$\boldsymbol{\delta}_b^S = \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{B}\boldsymbol{H}^t, \quad \boldsymbol{B} = \mathrm{diag}\,(b_1, \ldots, b_p), \quad b_i = m + p - 2i - 1, \qquad (3.8)$$

which satisfies the conditions in Proposition 3.3.

**Corollary 3.1** *The Stein type estimators* $\boldsymbol{\delta}_c^S$ *and* $\boldsymbol{\delta}_b^S$ *dominate the unbiased estimator* $\boldsymbol{\delta}_0$ *under the loss* $L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma})$.

The risk performances of the two Stein type estimators $\boldsymbol{\delta}_c^S$ and $\boldsymbol{\delta}_b^S$ are studied numerically in Section 4, which reports that $\boldsymbol{\delta}_b^S$ has much smaller risks than $\boldsymbol{\delta}_c^S$. However, the Stein type orthogonally equivariant estimator $\boldsymbol{\delta}_b^S$ has a shortcoming that the ordered relation that $b_1/\ell_1 \le \ldots \le b_p/\ell_p$ is not preserved. In the estimation of the covariance matrix under the Stein loss, Sheena and Takemura (1992) proved that a non-order-preserving estimator is improved on by the order-preserving methods such as order statistics and isotonic regression. In our setup, however, it is hard to show the similar property, mainly because the loss function $L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ incorporates the random matrix $\boldsymbol{W}$. In the next subsection, we obtain another type of estimators improving on the Stein type estimator $\boldsymbol{\delta}_b^S$.

## 3.2 Further improvement on the Stein type estimator

For the purpose, we consider estimators of the form

$$\boldsymbol{\delta}^{IS}(g) = \boldsymbol{\delta}_b^S + \frac{g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{W}}\boldsymbol{I}_p, \qquad (3.9)$$

where $g(\boldsymbol{\ell})$ is an absolutely continuous function. From Proposition 3.1, it is seen that the risk difference of the two estimators $\boldsymbol{\delta}_b^S$ and $\boldsymbol{\delta}^{IS}(g)$ is given by

$$\begin{aligned}\Delta =&R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}^{IS}(g)) - R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}_b^S)\\ =&\sum_{i=1}^p E\left[\frac{1}{\ell_i}\left(\frac{\ell_i^2 g^2}{(\operatorname{tr}\boldsymbol{W})^2} + 2\frac{b_i \ell_i g}{\operatorname{tr}\boldsymbol{W}}\right) - 2(m-p-1)\frac{g}{\operatorname{tr}\boldsymbol{W}}\right.\\ &\left. -2\sum_{j\neq i}\frac{g}{\operatorname{tr}\boldsymbol{W}} - 4\left(\frac{\sum_{j\neq i}\ell_j}{(\operatorname{tr}\boldsymbol{W})^2}g + \frac{\ell_i}{\operatorname{tr}\boldsymbol{W}}\frac{\partial g}{\partial \ell_i}\right)\right]\\ =&E\left[\frac{1}{\operatorname{tr}\boldsymbol{W}}\left\{g^2 - 4(p-1)g - 4\sum_{i=1}^p \ell_i \frac{\partial g}{\partial \ell_i}\right\}\right],\end{aligned} \tag{3.10}$$

since $\sum_{i=1}^p b_i = p(m-2)$. The expression (3.10) provides the following conditions for $\boldsymbol{\delta}^{IS}(g)$ to dominate $\boldsymbol{\delta}_b^S$.

**Proposition 3.4** *Assume that $g(\boldsymbol{\ell})$ satisfies the conditions:*
(a) $\partial g(\boldsymbol{\ell})/\partial \ell_i \geq 0$ *for* $i = 1, \ldots, p$.
(b) $0 < g(\boldsymbol{\ell}) \leq 4(p-1)$.
*Then the estimator $\boldsymbol{\delta}^{IS}(g)$ dominates the Stein type estimator $\boldsymbol{\delta}_b^S$ under the loss $L_1(\boldsymbol{\Sigma}, \boldsymbol{\delta})$.*

Putting $g(\boldsymbol{\ell}) = 2(p-1)$ in (3.9) gives the improved estimator

$$\boldsymbol{\delta}^{IS} = \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{B}\boldsymbol{H}^t + \frac{2(p-1)}{\operatorname{tr}\boldsymbol{W}}\boldsymbol{I}_p, \tag{3.11}$$

which we shall call *the improved Stein type estimator*. It is noted that $\boldsymbol{\delta}^{IS}$ has a similar form to the Efron-Morris estimator (4.1), which can not dominate $\boldsymbol{\delta}_b^S$, but $\boldsymbol{\delta}_0$.

## 3.3 Truncation rule

All the estimators given so far can be further improved on by use of the information on the restriction of the parameter space that $\boldsymbol{\Sigma}^{-1} \leq \boldsymbol{I}$. Thus, every estimator $\boldsymbol{\delta}$ should be constricted to the restricted space. Let $\boldsymbol{R}$ be a $p \times p$ orthogonal matrix such that $\boldsymbol{\delta} = \boldsymbol{R}\boldsymbol{\Lambda}\boldsymbol{R}^t$ for $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$. Consider the truncation rule defined by

$$[\boldsymbol{\Lambda}]^{TR} = \operatorname{diag}([\lambda_1]^{TR}, \ldots, [\lambda_p]^{TR}) \quad \text{for} \quad [\lambda_i]^{TR} = \min(\lambda_i, 1). \tag{3.12}$$

Then, Efron and Morris (1976) showed the following proposition.

**Proposition 3.5** *If $P[\lambda_i > 1$ for some $i] > 0$, then the estimator $\boldsymbol{\delta} = \boldsymbol{R}\boldsymbol{\Lambda}\boldsymbol{R}^t$ is improved on by the truncated one $\boldsymbol{\delta}^{TR} = \boldsymbol{R}[\boldsymbol{\Lambda}]^{TR}\boldsymbol{R}$ under the loss $L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma})$.*

Applying the truncation rule to the Stein type estimator $\boldsymbol{\delta}_b^S$ and the improved Stein type one $\boldsymbol{\delta}^{IS}$, we get further improved procedures

$$\boldsymbol{\delta}_b^{STR} =\boldsymbol{H}[\boldsymbol{B}\boldsymbol{L}^{-1}]^{TR}\boldsymbol{H}^t, \tag{3.13}$$

$$\boldsymbol{\delta}^{ISTR} =\boldsymbol{H}\left[\boldsymbol{B}\boldsymbol{L}^{-1} + \frac{2(p-1)}{\operatorname{tr}\boldsymbol{L}}\boldsymbol{I}_p\right]^{TR}\boldsymbol{H}^t. \tag{3.14}$$

# 4 Simulation studies

Now we investigate the risk-performances of estimators of $\boldsymbol{\Sigma}^{-1}$ numerically. The estimators we want to investigate are not only the estimators $\boldsymbol{\delta}_0$, $\boldsymbol{\delta}_c^{JS}$, $\boldsymbol{\delta}_c^{S}$, $\boldsymbol{\delta}_b^{S}$, $\boldsymbol{\delta}^{IS}$, $\boldsymbol{\delta}^{STR}$ and $\boldsymbol{\delta}^{ISTR}$ given so far in this paper, but also the following estimators studied in the literature: Efron and Morris (1976) proposed the improved estimator

$$\boldsymbol{\delta}^{EM} = (m-p-1)\boldsymbol{W}^{-1} + \frac{(p-1)(p+2)}{\operatorname{tr}\boldsymbol{W}}\boldsymbol{I}_p, \tag{4.1}$$

called *the Efron-Morris estimator*, and the truncated version

$$\boldsymbol{\delta}^{EMTR} = \boldsymbol{H}\left[(m-p-1)\boldsymbol{L}^{-1} + \frac{(p-1)(p+2)}{\operatorname{tr}\boldsymbol{L}}\boldsymbol{I}_p\right]^{TR}\boldsymbol{H}^t. \tag{4.2}$$

Haff (1979) and Dey *et al*. (1990) have treated more general types of estimators and derived conditions for the improvement over $\boldsymbol{\delta}_0$. Dey *et al*. (1990) proposed the estimator

$$\boldsymbol{\delta}^{DGS} = (m-p-1)\boldsymbol{W}^{-1} + \frac{2(p-1)}{\operatorname{tr}\boldsymbol{W}^2}\boldsymbol{W}, \tag{4.3}$$

and numerically revealed that the risk performance of $\boldsymbol{\delta}^{DGS}$ is comparable to that of the Efron-Morris estimator $\boldsymbol{\delta}^{EM}$.

Every estimator $\boldsymbol{\delta}$ is evaluated by the risk function $R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta})$ under the loss function $L_1(\boldsymbol{\Sigma}, \boldsymbol{\delta})$. The values of the risks of the above estimators are obtained from 5,000 replications through simulation experiments, and the relative efficiencies $R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta})/R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}_0)$ of estimator $\boldsymbol{\delta}$ over $\boldsymbol{\delta}_0$ are reported. The simulation experiments are done in the two cases: (1) $p = 2, 5, 7$, $m = 10$, $\boldsymbol{\Sigma} = \boldsymbol{H}\operatorname{diag}(\sigma_1, \ldots, \sigma_p)\boldsymbol{H}^t$ for $\sigma_i = (i-1)\times k+1$, $k = 0, \ldots, 4$, and some orthogonal matrices $\boldsymbol{H}$ and (2) $p = 5, 10, 15$, $m = 30$, $\boldsymbol{\Sigma} = \boldsymbol{H}\operatorname{diag}(\sigma_1, \ldots, \sigma_p)\boldsymbol{H}^t$ for $\sigma_i = \sqrt{(p-i)k} + 1$, $k = 0, \ldots, 4$, and some orthogonal matrices $\boldsymbol{H}$.

The relative efficiencies of the above estimators for the two cases are given in Tables 1 and 2 where the notations *JS*, *SC*, *SB*, *IS*, *EM*, *DGS*, *SBT*, *IST* and *EMT*, respectively, stand for the estimators $\boldsymbol{\delta}_c^{JS}$, $\boldsymbol{\delta}_c^{S}$, $\boldsymbol{\delta}_b^{S}$, $\boldsymbol{\delta}^{IS}$, $\boldsymbol{\delta}^{EM}$, $\boldsymbol{\delta}^{DSG}$, $\boldsymbol{\delta}^{STR}$, $\boldsymbol{\delta}^{ISTR}$ and $\boldsymbol{\delta}^{EMTR}$.

From these tables, the following conclusions can be drawn.

(1) The truncated improved Stein type estimator $\boldsymbol{\delta}^{ISTR}$ and the truncated Efron-Morris estimator $\boldsymbol{\delta}^{EMTR}$ have the best risk-performances and much smaller risks for higher dimensions $p$ than the unbiased estimator $\boldsymbol{\delta}_0$.

(2) Among the non-truncated estimators, the improved Stein type estimator $\boldsymbol{\delta}^{IS}$ is superior when $p/m$ is small, and the Efron-Morris estimator $\boldsymbol{\delta}^{EM}$ is better for large $p/m$.

(3) The risk gains of the James-Stein type estimator $\boldsymbol{\delta}_c^{JS}$ are quite small. Compared with $\boldsymbol{\delta}_c^{S}$, the Stein type estimator $\boldsymbol{\delta}_b^{S}$ is much better for all the cases. The risk-behaviors of the estimator $\boldsymbol{\delta}^{DGS}$ are worse than $\boldsymbol{\delta}^{EM}$ except the case of small $p/m$.

Table 1: Relative Efficiencies of the Estimators under the Loss $L_1(\boldsymbol{\Sigma}, \boldsymbol{\delta})$ in the Cases of $\boldsymbol{\Sigma} = \boldsymbol{H}\mathrm{diag}\,(\sigma_1, \dots, \sigma_p)\boldsymbol{H}^t$ for $\sigma_i = (i-1) \times k + 1$, $k = 0, \dots, 4$, $m = 10$ and $p = 2, 5, 7$

| $p$ | $k$ | $JS$ | $SC$ | $SB$ | $IS$ | $EM$ | $DSG$ | $SBT$ | $IST$ | $EMT$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 0.973 | 0.869 | 0.766 | 0.730 | 0.851 | 0.818 | 0.438 | 0.298 | 0.334 |
| | 1 | 0.982 | 0.907 | 0.837 | 0.805 | 0.867 | 0.853 | 0.565 | 0.469 | 0.469 |
| $p=2$ | 2 | 0.986 | 0.939 | 0.898 | 0.871 | 0.886 | 0.892 | 0.646 | 0.578 | 0.551 |
| | 3 | 0.989 | 0.957 | 0.930 | 0.906 | 0.902 | 0.918 | 0.683 | 0.629 | 0.593 |
| | 4 | 0.990 | 0.967 | 0.947 | 0.927 | 0.914 | 0.935 | 0.701 | 0.657 | 0.619 |
| | 0 | 0.947 | 0.666 | 0.484 | 0.442 | 0.462 | 0.667 | 0.214 | 0.100 | 0.087 |
| | 1 | 0.968 | 0.762 | 0.642 | 0.611 | 0.604 | 0.788 | 0.481 | 0.433 | 0.406 |
| $p=5$ | 2 | 0.975 | 0.824 | 0.740 | 0.716 | 0.695 | 0.847 | 0.609 | 0.578 | 0.544 |
| | 3 | 0.980 | 0.861 | 0.799 | 0.779 | 0.751 | 0.879 | 0.678 | 0.654 | 0.617 |
| | 4 | 0.983 | 0.886 | 0.837 | 0.821 | 0.790 | 0.900 | 0.721 | 0.701 | 0.663 |
| | 0 | 0.911 | 0.549 | 0.361 | 0.324 | 0.235 | 0.667 | 0.155 | 0.065 | 0.034 |
| | 1 | 0.946 | 0.674 | 0.545 | 0.520 | 0.479 | 0.805 | 0.438 | 0.407 | 0.372 |
| $p=7$ | 2 | 0.958 | 0.748 | 0.651 | 0.632 | 0.604 | 0.858 | 0.566 | 0.545 | 0.519 |
| | 3 | 0.966 | 0.794 | 0.717 | 0.702 | 0.678 | 0.887 | 0.642 | 0.625 | 0.602 |
| | 4 | 0.971 | 0.826 | 0.763 | 0.750 | 0.728 | 0.905 | 0.693 | 0.679 | 0.657 |

Table 2: Relative Efficiencies of the Estimators under the Loss $L_1(\boldsymbol{\Sigma}, \boldsymbol{\delta})$ in the Cases of $\boldsymbol{\Sigma} = \boldsymbol{H}\mathrm{diag}\,(\sigma_1, \dots, \sigma_p)\boldsymbol{H}^t$ for $\sigma_i = \sqrt{(p-i)k}+1$, $k = 0, \dots, 4$, $m = 30$ and $p = 5, 10, 15$

| $p$ | $k$ | $JS$ | $SC$ | $SB$ | $IS$ | $EM$ | $DSG$ | $SBT$ | $IST$ | $EMT$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 0.990 | 0.820 | 0.692 | 0.680 | 0.829 | 0.850 | 0.294 | 0.212 | 0.244 |
| | 1 | 0.988 | 0.881 | 0.806 | 0.795 | 0.853 | 0.882 | 0.663 | 0.636 | 0.650 |
| $p=5$ | 2 | 0.987 | 0.899 | 0.839 | 0.829 | 0.864 | 0.894 | 0.714 | 0.693 | 0.696 |
| | 3 | 0.987 | 0.909 | 0.856 | 0.847 | 0.871 | 0.902 | 0.737 | 0.719 | 0.717 |
| | 4 | 0.987 | 0.916 | 0.868 | 0.859 | 0.877 | 0.908 | 0.751 | 0.734 | 0.729 |
| | 0 | 0.984 | 0.721 | 0.572 | 0.563 | 0.647 | 0.838 | 0.194 | 0.136 | 0.132 |
| | 1 | 0.980 | 0.786 | 0.683 | 0.675 | 0.698 | 0.871 | 0.574 | 0.558 | 0.543 |
| $p=10$ | 2 | 0.980 | 0.807 | 0.718 | 0.710 | 0.718 | 0.882 | 0.637 | 0.625 | 0.607 |
| | 3 | 0.979 | 0.820 | 0.739 | 0.732 | 0.732 | 0.888 | 0.668 | 0.657 | 0.637 |
| | 4 | 0.979 | 0.830 | 0.754 | 0.748 | 0.742 | 0.893 | 0.688 | 0.678 | 0.656 |
| | 0 | 0.978 | 0.642 | 0.500 | 0.493 | 0.473 | 0.846 | 0.143 | 0.100 | 0.066 |
| | 1 | 0.975 | 0.707 | 0.602 | 0.596 | 0.551 | 0.877 | 0.493 | 0.481 | 0.410 |
| $p=15$ | 2 | 0.974 | 0.730 | 0.635 | 0.630 | 0.578 | 0.886 | 0.558 | 0.549 | 0.479 |
| | 3 | 0.974 | 0.744 | 0.657 | 0.651 | 0.596 | 0.891 | 0.593 | 0.585 | 0.515 |
| | 4 | 0.974 | 0.755 | 0.673 | 0.668 | 0.610 | 0.895 | 0.616 | 0.609 | 0.539 |

# 5 Estimation under the scale-invariant loss

Several dominance results have been stated in the previous sections for the non-scale-invariant loss $L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma})$. When a scale-invariant loss is employed, however, we have a different story, that is, the invariant loss allows us to provide the similar decision-theoretic results as in the estimation of the covariance matrix. The loss function we treat in this section is of the form

$$\begin{aligned} L_2(\boldsymbol{\delta}, \boldsymbol{\Sigma}) &= \mathrm{tr}\,(\boldsymbol{\delta} - \boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}(\boldsymbol{\delta} - \boldsymbol{\Sigma}^{-1})\boldsymbol{W} \\ &= \mathrm{tr}\,\boldsymbol{\delta}\boldsymbol{W}\boldsymbol{\delta}\boldsymbol{\Sigma} - 2\mathrm{tr}\,\boldsymbol{\delta}\boldsymbol{W} + \mathrm{tr}\,\boldsymbol{W}\boldsymbol{\Sigma}^{-1}, \end{aligned} \tag{5.1}$$

which is invariant under the scale transformations $\boldsymbol{W} \to \boldsymbol{A}\boldsymbol{W}\boldsymbol{A}^t$, $\boldsymbol{\Sigma} \to \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^t$ and $\boldsymbol{\delta} \to \boldsymbol{A}\boldsymbol{\delta}\boldsymbol{A}^t$ for any $p \times p$ matrix $\boldsymbol{A}$.

## 5.1 James-Stein type minimax estimator

The best scalar multiple of $\boldsymbol{W}^{-1}$ under the loss $L_2(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ is given by $\boldsymbol{\delta}_0 = a_0\boldsymbol{W}^{-1}$ for $a_0 = m - p - 1$. However, it is not minimax under the scale-invariant loss $L_2(\boldsymbol{\delta}, \boldsymbol{\Sigma})$. A minimax estimator with a constant risk is provided by the James-Stein type rule, where the minimaxity follows from Kiefer (1957). Let $\boldsymbol{W} = \boldsymbol{T}\boldsymbol{T}^t$ for $\boldsymbol{T} \in \mathcal{G}_T^+(p)$, and the risk function of the estimator $(\boldsymbol{T}^t)^{-1}\boldsymbol{D}\boldsymbol{T}^{-1}$ for a constant diagonal matrix $\boldsymbol{D} = \mathrm{diag}\,(d_1, \ldots, d_p)$ for positive constants $d_i$'s is written by

$$\begin{aligned} R_2(\boldsymbol{\Sigma}, (\boldsymbol{T}^t)^{-1}\boldsymbol{D}\boldsymbol{T}^{-1}) &= E_\Sigma[L_2((\boldsymbol{T}^t)^{-1}\boldsymbol{D}\boldsymbol{T}^{-1}, \boldsymbol{\Sigma})] \\ &= E_\Sigma\left[\mathrm{tr}\,(\boldsymbol{T}^t)^{-1}\boldsymbol{D}^2\boldsymbol{T}^{-1}\boldsymbol{\Sigma} - 2\mathrm{tr}\,\boldsymbol{D} + mp\right] \\ &= E_I\left[\mathrm{tr}\,\boldsymbol{D}^2\boldsymbol{T}^{-1}(\boldsymbol{T}^t)^{-1}\right] - 2\mathrm{tr}\,\boldsymbol{D} + mp. \end{aligned}$$

Krishnamoorthy and Gupta (1989) calculated the first term in the r.h.s. of the last equality as

$$E_I\left[\mathrm{tr}\,\boldsymbol{D}^2\boldsymbol{T}^{-1}(\boldsymbol{T}^t)^{-1}\right] = \sum_{i=1}^p \frac{m-1}{(m-i-1)(m-i)}d_i^2.$$

Hence the risk is rewritten by

$$R_2(\boldsymbol{\Sigma}, (\boldsymbol{T}^t)^{-1}\boldsymbol{D}\boldsymbol{T}^{-1}) = \sum_{i=1}^p \left\{\frac{m-1}{(m-i-1)(m-i)}d_i^2 - 2d_i + m\right\},$$

which is minimized at $d_i = (m-i-1)(m-i)/(m-1)$ for $i = 1, \ldots, p$, and hereafter, the notation $d_i$ means the constant. Thus, we get the James-Stein type estimator

$$\boldsymbol{\delta}_d^{JS} = (\boldsymbol{T}^t)^{-1}\boldsymbol{D}\boldsymbol{T}^{-1}, \quad \boldsymbol{D} = \mathrm{diag}\,(d_1, \ldots, d_p), \quad d_i = \frac{(m-i-1)(m-i)}{m-1}, \tag{5.2}$$

which is minimax with the constant risk

$$\begin{aligned} R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}_d^{JS}) &= \mathrm{tr}\,\boldsymbol{D} - 2\mathrm{tr}\,\boldsymbol{D} + mp = \sum_{i=1}^p \left\{(2m-1)i - i^2\right\}/(m-1) \\ &= p(p+1)\left[1 - \frac{p-1}{3(m-1)}\right], \end{aligned} \tag{5.3}$$

being less than the risk $R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}_0) = p(p+1)$.

## 5.2 Two dimensional case

For improving on the James-Stein type estimator $\boldsymbol{\delta}_d^{JS}$, consider the Stein type estimator

$$\boldsymbol{\delta}_d^S = \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{D}\boldsymbol{H}^t \tag{5.4}$$

where $\boldsymbol{W}$, $\boldsymbol{L}$ and $\boldsymbol{H}$ are the same notations as used in Section 3 and $\boldsymbol{D}$ is defined by (5.2). The risk function of $\boldsymbol{\delta}_d^S$ is expressed by

$$R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}_d^S) = E\left[\mathrm{tr}\, \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{D}^2\boldsymbol{H}^t\boldsymbol{\Sigma}\right] - 2\mathrm{tr}\, \boldsymbol{D} + mp. \tag{5.5}$$

It is interesting to note that the risk function under the loss $L_2(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ has the similar structure as in the case of the entropy loss function

$$L_e(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \mathrm{tr}\, \boldsymbol{\delta}\boldsymbol{\Sigma} - \log|\boldsymbol{\delta}\boldsymbol{\Sigma}| - p,$$

that is, the risk of $\boldsymbol{\delta}_d^S$ under the $L_e$-loss is given by

$$R_e(\boldsymbol{\Sigma}, \boldsymbol{\delta}_d^S) = E\left[\mathrm{tr}\, \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{D}\boldsymbol{H}^t\boldsymbol{\Sigma} - \log|\boldsymbol{D}| - \log|\boldsymbol{W}^{-1}\boldsymbol{\Sigma}| - p\right].$$

The minimaxity of the Stein estimator $\boldsymbol{\delta}_d^S$ is called the Krishnamoorthy-Gupta conjecture, because it is very hard to evaluate the term $E[\mathrm{tr}\, \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{D}\boldsymbol{H}^t\boldsymbol{\Sigma}]$. Perron (1997) proved the conjecture for $p = 2$, and Sheena (2003) recently succeeded in proving the minimaxity for $p = 3$, which needs a long and hard proof. though the issue is still open for $p \geq 4$. From these results, it follows that the Stein type estimator $\boldsymbol{\delta}_d^S$ dominates $\boldsymbol{\delta}_d^{JS}$ for $p = 2$ and 3. In the case of $p = 2$, we here give a simple proof different from Perron (1997) for the minimaxity of $\boldsymbol{\delta}_d^S$ under the $L_2$-loss.

**Proposition 5.1** *For $p = 2$, the Stein type estimator $\boldsymbol{\delta}_d^S$ dominates the James-Stein type minimax estimator $\boldsymbol{\delta}_d^{JS}$ relative to the scale-invariant loss $L_2(\boldsymbol{\delta}, \boldsymbol{\Sigma})$.*

**Proof.** Without any loss of generality, assume that $\boldsymbol{\Sigma} = \mathrm{diag}\,(\sigma_1, \sigma_2)$, $\sigma_1 \geq \sigma_2$. From (5.3) and (5.5), it suffices to show that

$$g(\boldsymbol{\Sigma}) \equiv E\left[\mathrm{tr}\, \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{D}^2\boldsymbol{H}^t\boldsymbol{\Sigma}\right] \leq \mathrm{tr}\, \boldsymbol{D},$$

or for $\boldsymbol{H} = (h_{ij})$,

$$g(\boldsymbol{\Sigma}) = E\left[\frac{d_1^2}{\ell_1}(\sigma_1 h_{11}^2 + \sigma_2 h_{21}^2) + \frac{d_2^2}{\ell_2}(\sigma_1 h_{12}^2 + \sigma_2 h_{22}^2)\right] \leq d_1 + d_2, \tag{5.6}$$

for $d_1 = m-2$ and $d_2 = (m-2)(m-3)/(m-1)$. Incorporating the term $|\boldsymbol{W}\boldsymbol{\Sigma}^{-1}|^{-1}$ into the Wishart density of $\boldsymbol{W}$ gives that

$$\begin{aligned} g(\boldsymbol{\Sigma}) &= E\left[\frac{\ell_1\ell_2}{\sigma_1\sigma_2}\left\{\frac{d_1^2}{\ell_1}(\sigma_1 h_{11}^2 + \sigma_2 h_{21}^2) + \frac{d_2^2}{\ell_2}(\sigma_1 h_{12}^2 + \sigma_2 h_{22}^2)\right\}\frac{1}{|\boldsymbol{W}\boldsymbol{\Sigma}^{-1}|}\right] \\ &= \frac{1}{(m-2)(m-3)}E\left[\frac{z_1 z_2}{\sigma_1\sigma_2}\left\{\frac{d_1^2}{z_1}(\sigma_1 h_{11}^2 + \sigma_2 h_{21}^2) + \frac{d_2^2}{z_2}(\sigma_1 h_{12}^2 + \sigma_2 h_{22}^2)\right\}\right] \\ &= \frac{1}{(m-2)(m-3)}E\left[d_1^2 z_2\left(\frac{h_{11}^2}{\sigma_2} + \frac{h_{21}^2}{\sigma_1}\right) + d_2^2 z_1\left(\frac{h_{12}^2}{\sigma_2} + \frac{h_{22}^2}{\sigma_1}\right)\right], \end{aligned}$$

where $\mathrm{diag}\,(z_1, z_2) = \boldsymbol{H}^t\boldsymbol{V}\boldsymbol{H}$ and $\boldsymbol{V}$ has $\mathcal{W}_p(m-2, \boldsymbol{\Sigma})$. Since $h_{11}^2 + h_{21}^2 = h_{11}^2 + h_{12}^2 = h_{22}^2 + h_{21}^2 = h_{22}^2 + h_{12}^2 = 1$, we observe that $h_{11}^2 = h_{22}^2$ and $h_{12}^2 = h_{21}^2$, which is used to express that

$$\begin{aligned} g(\boldsymbol{\Sigma}) &= \frac{1}{(m-2)(m-3)} E\left[ d_2^2 z_1 \left( \frac{h_{11}^2}{\sigma_1} + \frac{h_{12}^2}{\sigma_2} \right) + d_1^2 z_2 \left( \frac{h_{12}^2}{\sigma_1} + \frac{h_{22}^2}{\sigma_2} \right) \right] \\ &= \frac{1}{(m-2)(m-3)} E\left[ \mathrm{tr}\, \boldsymbol{H} \mathrm{diag}\,(d_2^2 z_1, d_1^2 z_2) \boldsymbol{H}^t \boldsymbol{\Sigma}^{-1} \right]. \end{aligned}$$

The Stein-Haff identity (3.4) and the equation (3.5) are used to rewrite $g(\boldsymbol{\Sigma})$ as

$$\begin{aligned} g(\boldsymbol{\Sigma}) = & \frac{1}{(m-2)(m-3)} E\left[ (m-5) \mathrm{tr}\, \boldsymbol{H} \mathrm{diag}\,(d_2^2 z_1, d_1^2 z_2) \boldsymbol{H}^t \boldsymbol{V}^{-1} \right. \\ & \left. + 2 \mathrm{tr}\, \boldsymbol{D}_V [\boldsymbol{H} \mathrm{diag}\,(d_2^2 z_1, d_1^2 z_2) \boldsymbol{H}^t] \right] \\ = & \frac{1}{(m-2)(m-3)} E\left[ (m-5)(d_1^2 + d_2^2) + 2 \frac{d_2^2 z_1 - d_1^2 z_2}{z_1 - z_2} + 2(d_1^2 + d_2^2) \right]. \end{aligned}$$

Since $(d_2^2 z_1 - d_1^2 z_2)/(z_1 - z_2) = d_2^2 - (d_1^2 - d_2^2) z_2/(z_1 - z_2)$, $g(\boldsymbol{\Sigma})$ is represented by

$$g(\boldsymbol{\Sigma}) = \frac{1}{(m-2)(m-3)} E\left[ (m-3) d_1^2 + (m-1) d_2^2 - 2(d_1^2 - d_2^2) \frac{z_2}{z_1 - z_2} \right],$$

which is less than $d_1 + d_2$. Therefore, the inequality (5.6) is proved. ■

## 5.3 Other improved estimators

As another orthogonally equivariant estimator, Section 4 has treated the Efron-Morris estimator

$$\boldsymbol{\delta}^{EM} = a_0 \boldsymbol{W}^{-1} + \frac{b}{\mathrm{tr}\, \boldsymbol{W}} \boldsymbol{I},$$

for $a_0 = m - p - 1$ and nonnegative constant $b$. Although it does not dominate the James-Stein type estimator $\boldsymbol{\delta}_d^{JS}$, this type of estimators is one of standard procedures for estimating the precision matrix. We thus obtain a condition on $b$ for $\boldsymbol{\delta}^{EM}$ to improve on the unbiased estimator $\boldsymbol{\delta}_0$ relative to the scale-invariant loss.

**Proposition 5.2** *The Efron-Morris estimator $\boldsymbol{\delta}^{EM}$ dominates the unbiased one $\boldsymbol{\delta}_0$ relative to the scale-invariant loss function $L_2(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ if $0 < b \le 2(p-1)$.*

**Proof.** The risk of the Efron-Morris estimator $\boldsymbol{\delta}^{EB}$ is given by

$$R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}^{EM}) = E\left[ \mathrm{tr}\, \boldsymbol{\delta}^{EM} \boldsymbol{W} \boldsymbol{\delta}^{EM} \boldsymbol{\Sigma} - 2 \mathrm{tr}\, \boldsymbol{\delta}^{EM} \boldsymbol{W} + \mathrm{tr}\, \boldsymbol{W} \boldsymbol{\Sigma}^{-1} \right]. \tag{5.7}$$

Since $\mathrm{tr}\, \boldsymbol{\delta}^{EM} \boldsymbol{W} = a_0 p + b$ and

$$\mathrm{tr}\, \boldsymbol{\delta}^{EM} \boldsymbol{W} \boldsymbol{\delta}^{EM} \boldsymbol{\Sigma} = a_0^2 \mathrm{tr}\, \boldsymbol{W}^{-1} \boldsymbol{\Sigma} + 2 a_0 b \frac{\mathrm{tr}\, \boldsymbol{\Sigma}}{\mathrm{tr}\, \boldsymbol{W}} + b^2 \frac{\mathrm{tr}\, \boldsymbol{W} \boldsymbol{\Sigma}}{(\mathrm{tr}\, \boldsymbol{W})^2},$$

the difference of the risk functions of $\boldsymbol{\delta}_0$ and $\boldsymbol{\delta}^{EM}$ is written by

$$\begin{aligned}\Delta &= R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}^{EM}) - R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}_0)\\ &= -2b\left(1 - a_0 E_\Sigma\left[\frac{\operatorname{tr}\boldsymbol{\Sigma}}{\operatorname{tr}\boldsymbol{W}}\right]\right) + b^2 E_\Sigma\left[\frac{\operatorname{tr}\boldsymbol{W}\boldsymbol{\Sigma}}{(\operatorname{tr}\boldsymbol{W})^2}\right]. \end{aligned} \tag{5.8}$$

We first evaluate the term $E_\Sigma[\operatorname{tr}\boldsymbol{\Sigma}/\operatorname{tr}\boldsymbol{W}]$, which is rewritten by $E_I[\operatorname{tr}\boldsymbol{\Sigma}/\operatorname{tr}\boldsymbol{V}\boldsymbol{\Sigma}]$ where $\boldsymbol{V}$ has $\mathcal{W}_p(m, \boldsymbol{I})$ and $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \dots, \sigma_p)$. Let $v_i$ be the $(i, i)$ diagonal element of $\boldsymbol{V}$ for $i = 1, \dots, p$, and let $\gamma_i = \sigma_i / \sum_{j=1}^p \sigma_j$. Then, $\operatorname{tr}\boldsymbol{\Sigma}/\operatorname{tr}\boldsymbol{V}\boldsymbol{\Sigma} = (\sum_{i=1}^p v_i\gamma_i)^{-1} \le \sum_{i=1}^p v_i^{-1}\gamma_i$, where the inequality follows from Schwarz' inequality. Noting that $E[1/v_i] = 1/(m-2)$ for $i = 1, \dots, p$, we see that

$$E_\Sigma\left[\frac{\operatorname{tr}\boldsymbol{\Sigma}}{\operatorname{tr}\boldsymbol{W}}\right] \le \sum_{i=1}^p E\left[\frac{1}{v_i}\right]\gamma_i = \frac{1}{m-2}\sum_{i=1}^p \gamma_i = \frac{1}{m-2}. \tag{5.9}$$

We next evaluate the term $E_\Sigma[\operatorname{tr}\boldsymbol{W}\boldsymbol{\Sigma}/(\operatorname{tr}\boldsymbol{W})^2]$, which is expressed by

$$E_I\left[\frac{\operatorname{tr}\boldsymbol{V}\boldsymbol{\Sigma}^2}{(\operatorname{tr}\boldsymbol{V}\boldsymbol{\Sigma})^2}\right] = E_I\left[\frac{\sum_{i=1}^p v_i\sigma_i^2}{(\sum_{i=1}^p v_i\sigma_i)^2}\right] \le E_I\left[\frac{\sum_{i=1}^p v_i\eta_i}{\sum_{i=1}^p v_i^2\eta_i}\right],$$

where $\eta_i = \sigma_i^2/\sum_{j=1}^p \sigma_j^2$. Since $\sum_{i=2}^p v_i^2\eta_i \sum_{i=2}^p \eta_i \ge (\sum_{i=2}^p v_i\eta_i)^2$ as checked by using Schwarz' inequality, we observe that

$$\begin{aligned}\frac{\sum_{i=1}^p v_i\eta_i}{\sum_{i=1}^p v_i^2\eta_i} &= \frac{v_1\eta_1 + (1-\eta_1)\sum_{i=2}^p v_i\eta_i/(1-\eta_1)}{v_1^2\eta_1 + (1-\eta_1)\sum_{i=2}^p v_i^2\eta_i/(1-\eta_1)}\\ &\le \frac{v_1\eta_1 + A(1-\eta_1)}{v_1^2\eta_1 + A^2(1-\eta_1)}\\ &\le \frac{1}{v_1}\eta_1 + \frac{1}{A}(1-\eta_1),\end{aligned}$$

where the second inequality is equivalently expressed by $(v_1 - A)^2(v_1 + A) \ge 0$. Recall that $A^{-1} = (\sum_{i=2}^p v_i\theta_i)^{-1} \le \sum_{i=2}^p v_i^{-1}\theta_i$ for $\theta_i = \eta_i/\sum_{j=2}^p \eta_j$. Hence we get the inequality

$$\frac{\sum_{i=1}^p v_i\eta_i}{\sum_{i=1}^p v_i^2\eta_i} \le \frac{1}{v_1}\eta_1 + (1-\eta_1)\sum_{i=2}^p \frac{1}{v_i}\theta_i, \tag{5.10}$$

which implies that

$$E\left[\frac{\sum_{i=1}^p v_i\eta_i}{\sum_{i=1}^p v_i^2\eta_i}\right] \le E[v_1^{-1}]\eta_1 + (1-\eta_1)\sum_{i=2}^p E[v_i^{-1}]\theta_i = \frac{1}{m-2}. \tag{5.11}$$

Combining (5.8), (5.9) and (5.11) gives that

$$\Delta \le -2b\left(1 - \frac{m-p-1}{m-2}\right) + \frac{b^2}{m-2} = \frac{b}{m-2}[-2(p-1) + b],$$

which is less than or equal to zero if $b \le 2(p-1)$. The proof of Proposition 5.2 is therefore complete. ■

We conclude this section with noting that the truncation rule (3.12) gives further improvements under the invariant loss $L_2(\boldsymbol{\delta}, \boldsymbol{\Sigma})$.

**Proposition 5.3** *The estimator* $\boldsymbol{\delta} = \boldsymbol{R}\boldsymbol{\Lambda}\boldsymbol{R}^t$ *is improved on by the truncated one* $\boldsymbol{\delta}^{TR} = \boldsymbol{R}[\boldsymbol{\Lambda}]^{TR}\boldsymbol{R}$ *under the scale-invariant loss* $L_2(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ *if* $P[\lambda_i > 1 \text{ for some } i] > 0$.

This proposition follows from the fact that the risk difference can be written as

$$\begin{aligned} & R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}) - R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}^{TR}) \\ & \quad = E\left[\sum_{i=1}^{p} (\lambda_i - \lambda_i^{TR})\ell_i \left\{(\lambda_i + \lambda_i^{TR})(\boldsymbol{R}^t\boldsymbol{\Sigma}\boldsymbol{R})_{ii} - 2\right\}\right] \\ & \quad \geq E\left[\sum_{i=1}^{p} (\lambda_i - 1)\ell_i \left\{(\lambda_i + 1) - 2\right\} I(\lambda_i > 1)\right], \end{aligned}$$

where $(\boldsymbol{R}^t\boldsymbol{\Sigma}\boldsymbol{R})_{ii}$ is the $(i, i)$ element of $\boldsymbol{R}^t\boldsymbol{\Sigma}\boldsymbol{R}$, and $(\boldsymbol{R}^t\boldsymbol{\Sigma}\boldsymbol{R})_{ii} \geq 1$.

Using Proposition 5.3, we get improved truncated estimators

$$\begin{aligned} \boldsymbol{\delta}_d^{STR} &= \boldsymbol{H}[\boldsymbol{D}\boldsymbol{L}^{-1}]^{TR}\boldsymbol{H}^t, \\ \boldsymbol{\delta}^{EMTR} &= \boldsymbol{H}\left[(m - p - 1)\boldsymbol{L}^{-1} + \frac{2(p-1)}{\operatorname{tr} \boldsymbol{L}}\boldsymbol{I}_p\right]^{TR}\boldsymbol{H}^t. \end{aligned}$$

## 5.4 Simulation results

We now investigate the risk-performances of estimators of $\boldsymbol{\Sigma}^{-1}$ numerically under the loss $L_2(\boldsymbol{\Sigma}, \boldsymbol{\delta})$. The estimators we want here to investigate are $\boldsymbol{\delta}_0$, $\boldsymbol{\delta}_d^{JS}$, $\boldsymbol{\delta}_d^{S}$, $\boldsymbol{\delta}^{EM}$, $\boldsymbol{\delta}_d^{STR}$ and $\boldsymbol{\delta}^{EMTR}$ given above,

The risk functions of the above estimators are obtained from 5,000 replications through simulation experiments, and the relative efficiencies $R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta})/R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}_0)$ of estimator $\boldsymbol{\delta}$ over $\boldsymbol{\delta}_0$ are reported in Tables 3 and 4, where the simulation experiments are done in the same cases as in Tables 1 and 2. The notations $JS$, $ST$, $EM$, $STT$ and $EMT$, respectively, stand for the estimators $\boldsymbol{\delta}_d^{JS}$, $\boldsymbol{\delta}_d^{S}$, $\boldsymbol{\delta}^{EM}$, $\boldsymbol{\delta}_d^{STR}$ and $\boldsymbol{\delta}^{EMTR}$.

These tables show the following conclusions:

(1) Through the numerical results given in Tables 3 and 4, the truncated Stein type estimator $\boldsymbol{\delta}_d^{S}$ has the smallest risks except the case of $p = 2$ in Table 3.

(2) The Stein type estimator $\boldsymbol{\delta}_d^{S}$ is the best of all the non-truncated estimators. The risk-gains of the James-Stein type estimator $\boldsymbol{\delta}_d^{JS}$ is quite small.

# 6 Concluding remarks

In this paper, we have considered the estimation of the precision matrix relative to the non-scale-invariant loss function induced from the estimation of means. Using the methods as in the estimation of the covariance matrix, we have derived not only the James-Stein type and the Stein type estimators, but also a new type estimator, called improved Stein type estimator. However, we have observed that several dominance properties known

Table 3: Relative Efficiencies of the Estimators under the Loss $L_2(\boldsymbol{\Sigma}, \boldsymbol{\delta})$ in the Cases of $\boldsymbol{\Sigma} = \boldsymbol{H}\text{diag}\,(\sigma_1, \ldots, \sigma_p)\boldsymbol{H}^t$ for $\sigma_i = (i-1) \times k + 1$, $k = 0, \ldots, 4$, $m = 10$ and $p = 2, 5, 7$

| $p$ | $k$ | $JS$ | $ST$ | $EM$ | $STT$ | $EMT$ |
|---|---|---|---|---|---|---|
| | 0 | 0.9556 | 0.7806 | 0.8883 | 0.5725 | 0.4852 |
| | 1 | 0.9555 | 0.8261 | 0.8954 | 0.6621 | 0.5817 |
| $p = 2$ | 2 | 0.9554 | 0.8696 | 0.9063 | 0.7285 | 0.6467 |
| | 3 | 0.9554 | 0.8940 | 0.9157 | 0.7640 | 0.6821 |
| | 4 | 0.9554 | 0.9085 | 0.9234 | 0.7840 | 0.7048 |
| | 0 | 0.8489 | 0.5511 | 0.7352 | 0.5114 | 0.4962 |
| | 1 | 0.8486 | 0.6047 | 0.7467 | 0.5825 | 0.5905 |
| $p = 5$ | 2 | 0.8486 | 0.6321 | 0.7520 | 0.6154 | 0.6321 |
| | 3 | 0.8485 | 0.6472 | 0.7546 | 0.6327 | 0.6513 |
| | 4 | 0.8485 | 0.6568 | 0.7561 | 0.6435 | 0.6618 |
| | 0 | 0.7786 | 0.5145 | 0.6969 | 0.5070 | 0.5698 |
| | 1 | 0.7788 | 0.5513 | 0.7071 | 0.5469 | 0.6125 |
| $p = 7$ | 2 | 0.7788 | 0.5651 | 0.7103 | 0.5616 | 0.6316 |
| | 3 | 0.7789 | 0.5721 | 0.7117 | 0.5689 | 0.6419 |
| | 4 | 0.7789 | 0.5765 | 0.7125 | 0.5736 | 0.6484 |

Table 4: Relative Efficiencies of the Estimators under the Loss $L_2(\boldsymbol{\Sigma}, \boldsymbol{\delta})$ in the Cases of $\boldsymbol{\Sigma} = \boldsymbol{H}\text{diag}\,(\sigma_1, \ldots, \sigma_p)\boldsymbol{H}^t$ for $\sigma_i = \sqrt{(p-i)k}+1$, $k = 0, \ldots, 4$, $m = 30$ and $p = 5, 10, 15$

| $p$ | $k$ | $JS$ | $ST$ | $EM$ | $STT$ | $EMT$ |
|---|---|---|---|---|---|---|
| | 0 | 0.9546 | 0.6685 | 0.9149 | 0.4817 | 0.5065 |
| | 1 | 0.9545 | 0.7452 | 0.9168 | 0.6969 | 0.7833 |
| $p = 5$ | 2 | 0.9545 | 0.7639 | 0.9175 | 0.7261 | 0.8134 |
| | 3 | 0.9544 | 0.7730 | 0.9179 | 0.7392 | 0.8252 |
| | 4 | 0.9544 | 0.7786 | 0.9181 | 0.7470 | 0.8318 |
| | 0 | 0.8977 | 0.5421 | 0.8916 | 0.4649 | 0.5806 |
| | 1 | 0.8978 | 0.5953 | 0.8926 | 0.5835 | 0.7936 |
| $p = 10$ | 2 | 0.8978 | 0.6081 | 0.8928 | 0.6004 | 0.8240 |
| | 3 | 0.8978 | 0.6149 | 0.8930 | 0.6087 | 0.8360 |
| | 4 | 0.8978 | 0.6193 | 0.8931 | 0.6139 | 0.8424 |
| | 0 | 0.8405 | 0.4898 | 0.8831 | 0.4617 | 0.6501 |
| | 1 | 0.8405 | 0.5242 | 0.8837 | 0.5207 | 0.7895 |
| $p = 15$ | 2 | 0.8405 | 0.5321 | 0.8838 | 0.5300 | 0.8175 |
| | 3 | 0.8405 | 0.5365 | 0.8839 | 0.5349 | 0.8305 |
| | 4 | 0.8405 | 0.5394 | 0.8840 | 0.5381 | 0.8381 |

in the estimation of the covariance matrix do not necessarily hold under the non-scale-invariant loss, but still hold relative to the scale-invariant loss. The simulation studies under the non-scale-invariant loss show that the truncated improved Stein type estimator $\boldsymbol{\delta}^{ISTR}$ and the truncated Efron-Morris estimator have the best risk performances among the competitors and much smaller risks for high dimension $p$ than the unbiased estimator $\boldsymbol{\delta}_0$.

Although the 'within' component of covariance $\boldsymbol{\Sigma}_E$ in the model (1.2) is assumed to be known in this paper, this assumption needs to be removed in more practical situations studied by Kubokawa and Srivastava (2003). Some results given in the previous sections can be extended to the model with unknown error covariance matrix $\boldsymbol{\Sigma}_E$, and the same method as in the derivation of $\boldsymbol{\delta}^{ISTR}$ can apply to the model for providing a superior estimator.

Finally, we give some conjectures which we could not show here. For the non-scale-invariant loss $L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma})$, we conjecture that the unbiased estimator $\boldsymbol{\delta}_0$ should be minimax. It is also interesting to investigate whether the Stein type estimator is improved on by the order-preserving estimator introduced by Sheena and Takemura (1992). For the scale-invariant loss $L_2(\boldsymbol{\delta}, \boldsymbol{\Sigma})$, we have the Krishnamoorthy-Gupta conjecture that the James-Stein type estimator is improved on the Stein type estimator for $p \geq 4$, and the conjecture that the Stein type estimator is dominated by the order-preserving one. It could be interesting to show that the Stein type estimator dominates $\boldsymbol{\delta}_0$.

The simulation studies demonstrates that the truncated procedures are much better than the non-truncated, but applying the truncation rule given in Section 3 results in non-smooth estimators. From Bayesian perspective, it is the most interesting issue to find smooth or Bayesian estimators which exist on the parameter space of $\boldsymbol{\Sigma}^{-1} \leq \boldsymbol{I}_p$ and dominate $\boldsymbol{\delta}_0$ under the loss $L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma})$. The estimators treated by Zheng (1986a,b) may be helpful for the purpose. For a positive valued and absolutely continuous function $f(\boldsymbol{\ell})$, the Zheng type estimator is given by

$$\boldsymbol{\delta}^Z(f) = \boldsymbol{H}\mathrm{diag}\left(\frac{\partial}{\partial \ell_1}\log f(\boldsymbol{\ell}), \ldots, \frac{\partial}{\partial \ell_p}\log f(\boldsymbol{\ell})\right)\boldsymbol{H}^t,$$

which, for instance, includes the Stein type estimator $\boldsymbol{\delta}_b^S$ and the Efron-Morris type one $\boldsymbol{\delta}^{EM}$, derived by putting $f(\boldsymbol{\ell}) = \prod_{j=1}^p \ell_j^{-b_j}$ and $f(\boldsymbol{\ell}) = (\prod_{j=1}^p \ell_j^{-(m-p-1)})(\sum_{j=1}^p \ell_j)^{-(p-1)(p+2)}$, respectively. The interesting issue is how to find the function $f(\boldsymbol{\ell})$ which yields a superior estimator with the above requirements.

**Acknowledgments.** The research of the author was supported in part by grants from the Ministry of Education, Japan, Nos. 11680320, 13680371, 15200021 and 15200022 and in part by a grant from COE-Economics, University of Tokyo.

## REFERENCES

Dey, D.K., Ghosh, M. and Srinivasan, C. (1990). A new class of improved estimators of a multinormal precision matrix. *Statist. and Decision*, **8**, 141-151.

Efron, B. and Morris, C. (1976). Multivariate empirical Bayes estimation of covariance matrices. *Ann. Statist.*, **4**, 22-32.

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

Haff, L.R. (1979b). Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity. *Ann. Statist.*, **7**, 1264-1276.

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.

Kiefer, J. (1957). Invariant, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.*, **28**, 573-601.

Krishnamoorthy, K. and Gupta, A.K. (1989). Improved minimax estimation of a normal precision matrix. *Canadian J. Statist.*, **17**, 91-102.

Kubokawa, T. and Srivastava, M.S. (2003). Prediction in multivariate mixed linear models. *J. Japan Statist. Soc.*, **33**, 245-270.

Perron, F. (1997). On a conjecture of Krishnamoorthy and Gupta. *J. Multivariate Anal.*, **62**, 110-120.

Sheena, Y. (1995). Unbiased estimator of risk for an orthogonally invariant estimator of a covariance matrix. *J. Japan Statist. Soc.*, **25**, 35-48.

Sheena, Y. (2003). On minimaxity of the normal precision matrix estimator of Krishnamoorthy and Gupta. *Statistics*, **37**, 387-399.

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.

Stein, C. (1977). Lectures on multivariate estimation theory. (In Russian.) In *Investigation on Statistical Estimation Theory* I, 4-65. Zapiski Nauchych Seminarov LOMI im. V.A. Steklova AN SSSR vol. 74, Leningrad.(In Russian)

Zheng, Z. (1986a). On estimation of matrix of normal mean. *J. Multivariate Anal.*, **18**, 70-82.

Zheng, Z. (1986b). Selecting a minimax estimator doing well at a point. *J. Multivariate Anal.*, **19**, 14-23.