Another Class of Minimax Estimators of A Variance Covariance Matrix in Multivariate Normal Distribution

Hisayuki Hara

METR 98-08 September 1998

Communicated by Professor C.Hirotsu

Another Class of Minimax Estimators of A Variance Covariance Matrix in Multivariate Normal Distribution

Hisayuki Hara

Department of Mathematical Engineering and Information physics
University of Tokyo,

Summary

It is well known that the best equivariant estimator of a variance covariance matrix of multivariate normal distribution with respect to the full affine group of transformation is not even minimax. Some minimax estimators have been proposed. Here we treat this problem in the framework of a multivariate analysis of variance(MANOVA) model and give other classes of minimax estimators.

Key words and phrases: Conditional risk difference, entropy loss, inadmissibility, MANOVA, minimax estimation, Wishart matrix.

AMS 1985 subject classifications: primary 62F10; secondary 62H99.

Address: Department of Mathematical Engineering and Information physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

e-mail: hara@stat.t.u-tokyo.ac.jp

1 Introduction

A canonical form of a normal MANOVA model is expressed as

$$\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k) : \quad p \times k \sim \mathrm{N}(\boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_p),$$

$$\boldsymbol{S} : \quad p \times p \sim \mathrm{W}_p(n, \boldsymbol{\Sigma}) \quad (n \geq p), \tag{1}$$

$$\boldsymbol{X} \text{ and } \boldsymbol{S} \text{ are independent},$$

where $\mathrm{N}(\boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_p)$ denotes that $\boldsymbol{X}_i$ are independently distributed to the multivariate normal distribution with mean $\boldsymbol{\mu}_i$, the corresponding column of $\boldsymbol{M} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k)$, and covariance matrix $\boldsymbol{\Sigma}$. Let $\mathrm{W}_p(n, \boldsymbol{\Sigma})$ denote the Wishart distribution with parameter $\boldsymbol{\Sigma}$ and degrees of freedom n. Here we consider the problem of estimating $\boldsymbol{\Sigma}$ with unknown $\boldsymbol{M}$. We use the Stein's loss

$$L(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{tr}(\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}) - \log|\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p. \tag{2}$$

as a criterion.

This problem remains invariant under the full affine group acting as

$$(\boldsymbol{X}, \boldsymbol{S}) \to (\boldsymbol{A}\boldsymbol{X} + \boldsymbol{B}, \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}'), \quad (\boldsymbol{M}, \boldsymbol{\Sigma}) \to (\boldsymbol{A}\boldsymbol{M} + \boldsymbol{B}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')$$

for $p \times p$ nonsingular matrix $\boldsymbol{A}$ and $p \times k$ matrix $\boldsymbol{B}$. The best affine equivariant estimator is given by

$$\delta_0 = \frac{1}{n}\boldsymbol{S}, \tag{3}$$

which is also the uniformly minimum variance unbiased(UMVU) estimator. It is well known that δ_0 is not even minimax.

Let $\boldsymbol{T}$ be the lower triangular matrix satisfying $\boldsymbol{S} = \boldsymbol{T}\boldsymbol{T}'$. James and Stein[4] showed the estimator

$$\delta^{JS} = \boldsymbol{T}\boldsymbol{L}\boldsymbol{T}', \quad \boldsymbol{L} = \mathrm{diag}\{(n + p + 1 - 2i)^{-1}\} \tag{4}$$

dominates δ_0 and is a miminax estimator under the loss (2) with constant risk. But δ^{JS} is inadmissible and various estimators improving on δ^{JS} can be conceivable.

Let the spectrum decomposition of $\boldsymbol{S}$ be expressed as $\boldsymbol{S} = \boldsymbol{H\Lambda H}'$, where

$$\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda}), \quad \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p), \quad \lambda_1 \geq \cdots \geq \lambda_p.$$

Stein[11] and Dey and Srinivasan[1] showed that the estimator

$$\delta^{DS} = \boldsymbol{H\Lambda}^{1/2}\boldsymbol{L\Lambda}^{1/2}\boldsymbol{H}', \quad \boldsymbol{\Lambda}^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_p^{1/2}) \tag{5}$$

improves δ^{JS} and then is minimax.

In general the estimator of form

$$\delta^{OI} = \boldsymbol{H\Lambda}^{\phi}\boldsymbol{H}', \quad \boldsymbol{\Lambda}^{\phi} = \mathrm{diag}\left\{\phi_i(\boldsymbol{\lambda})\right\}, \tag{6}$$

is called orthogonal invariant. Obviously δ^{DS} is in this class. Dey and Srinivasan[1] also proposed other orthogonal invariant estimators improving on δ^{DS}.

Since $\lambda_1 \geq \cdots \geq \lambda_p$, it seems preferable to take $\phi_i(\boldsymbol{\lambda})$ satifying

$$\phi_1(\boldsymbol{\lambda}) \geq \cdots \geq \phi_p(\boldsymbol{\lambda}).$$

with probability 1. The estimator of Dey and Srinivasan[1] does not satisfy this condition. Sheena and Takemura[12] showed that any orthogonal invariant estimator which does not preserve the order of $\phi_i(\boldsymbol{\lambda})$ is dominated by some modified estimators preserving the order and proposed two methods to modify non-orderpreserving estimators.

In this way only the information of $\boldsymbol{S}$ has been used to obtain minimax estimators. Besides these results, there are several work on these approach in Takemura[13], Haff[2], Perron[8] etc.

On the other hand, there is another approach to obtain the estimator improving on δ_0, which is the one to use the information of not only $\boldsymbol{S}$ but $\boldsymbol{X}$. When

$p = 1$, it is well known that Stein[10] presented a truncated estimator dominating δ_0. Sinha and Ghosh[9] extended the Stein's results to the MANOVA model (1) and obtained the estimator improving on δ_0 under the loss (2). However their estimator is not continuous. For $k = 1$ and $p \geq 2$, Perron[7], Kubokawa et al.[5] and Kubokawa et al.[6] proposed continuous estimators which dominates Sinha-Ghosh's estimator. Recently Hara[3] generalize Perron's result and derive the continuous estimator improving on Sinha-Ghosh's estimator for $k \geq 2$. Hara[3] also extended the Kubokawa's results to $k \geq 2$ and obtained some new estimators improving on δ_0. But these improved estimators which use the information of $\boldsymbol{X}$ are not shown to be minimax.

These two approaches have developed independently. Main purpose of this paper is to give new classes of minimax estimators which also use the information of $\boldsymbol{X}$ by using the argument in Hara[3]. In section 2 we obtain some minimax estimators with the information of $\boldsymbol{X}$ explicitly which dominate δ^{DS}. In section 3 we show the inadmissibility of order-preserving minimax estimators proposed by Sheena and Takemura[12]. Section 4 gives some Monte Carlo studies to show the performance of the estimator proposed in section 2 and section 3.

2 Another Class of Minimax Estimators

Letting $\boldsymbol{U}$ be $\boldsymbol{U} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$, δ^{DS} can be rewritten by

$$\delta^{DS} = \boldsymbol{U}\boldsymbol{L}\boldsymbol{U}'.$$

Let the spectrum decomposition of $\boldsymbol{U}^{-1}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{U}^{-1\prime}$ be written by $\boldsymbol{U}^{-1}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{U}^{-1\prime} = \boldsymbol{H}^*\boldsymbol{\Lambda}^*\boldsymbol{H}^{*\prime}$ with

$$\boldsymbol{\Lambda}^* = \mathrm{diag}(\boldsymbol{\lambda}^*), \quad \boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_p^*).$$

First we consider to improve δ^{DS} by the estimator in the following class,

$$\delta_1^{\phi} = \phi(\boldsymbol{\lambda}^*)\boldsymbol{U}\boldsymbol{L}\boldsymbol{U}', \tag{7}$$

where $\phi(\boldsymbol{\lambda}^*)$ is a scalar function of $\boldsymbol{\lambda}^*$ satisfying $\phi(\boldsymbol{\lambda}^*) - 1 \leq 0$..

Let $\boldsymbol{W}$ be a $p \times p$ random matrix whose probability density function with respect to the Lebesgue measure d$\boldsymbol{W}$ is

$$\text{Const.}|\boldsymbol{W}\boldsymbol{W}'|^{(n-p)/2} \exp\left(-\frac{1}{2}\text{tr}\boldsymbol{W}\boldsymbol{W}'\right). \tag{8}$$

It is easy to show that $\boldsymbol{W}\boldsymbol{W}'$ and $\boldsymbol{W}'\boldsymbol{W}$ follow $\text{W}_p(n, \boldsymbol{I}_p)$. Then we can set

$$\boldsymbol{W}\boldsymbol{W}' = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}'^{-1/2} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{U}\boldsymbol{U}'\boldsymbol{\Sigma}'^{-1/2},$$

where $\boldsymbol{\Sigma}^{-1/2}$ is a $p \times p$ constatnt matrix satisfying $\boldsymbol{\Sigma}^{-1/2'}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$. Letting $\boldsymbol{P}$ be $\boldsymbol{P} = \boldsymbol{U}^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{W}$ with $\boldsymbol{\Sigma}^{1/2} = (\boldsymbol{\Sigma}^{-1/2})^{-1}$, we can rewrite $\boldsymbol{W}$ as

$$\boldsymbol{W} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{U}\boldsymbol{P}. \tag{9}$$

We can easily see that $\boldsymbol{P}$ is an orthogonal matrix. In the following argument we define $\boldsymbol{W}$ as (9) and let $\tilde{\boldsymbol{X}}$ be $\tilde{\boldsymbol{X}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}$.

To compare δ^{DS} and δ_1^{ϕ}, we consider the conditional risk difference with $\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}} = \boldsymbol{P}\boldsymbol{U}^{-1}\boldsymbol{X}$ fixed, i.e.,

$$\text{E}[L(\delta_1^{\phi}, \boldsymbol{\Sigma}) - L(\delta^{DS}, \boldsymbol{\Sigma})|\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}].$$

Since the spectrum decomposition of $\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}'\boldsymbol{W}^{-1'}$ is expressed as

$$\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}'\boldsymbol{W}' = \boldsymbol{P}'\boldsymbol{H}^*\boldsymbol{\Lambda}^*\boldsymbol{H}^{*\prime}\boldsymbol{P},$$

$\phi(\boldsymbol{\lambda}^*)$ is constant with $\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}$ given. Then

$$\begin{aligned}
\text{E}[\text{tr}(\delta_1^{\phi} - \delta^{DS})\boldsymbol{\Sigma}^{-1}|\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}] &= \text{E}[(\phi(\boldsymbol{\lambda}^*) - 1)\text{tr}\boldsymbol{L}\boldsymbol{P}'\boldsymbol{U}'\boldsymbol{\Sigma}^{-1}\boldsymbol{U}\boldsymbol{P}\ |\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}] \\
&\leq \frac{\phi(\boldsymbol{\lambda}^*) - 1}{n + p - 1}\text{trE}[\boldsymbol{W}'\boldsymbol{W}|\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}] \\
&\leq \frac{(n+k)(\phi(\boldsymbol{\lambda}^*) - 1)}{n + p - 1}\text{tr}(\boldsymbol{I}_p + \boldsymbol{W}^{-1}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{W}^{-1'})^{-1} \\
&= \frac{n+k}{n+p-1}\sum_{i=1}^{p}\frac{\phi(\boldsymbol{\lambda}^*) - 1}{1 + \lambda_i^*},
\end{aligned} \tag{10}$$

$$\begin{aligned}
&\log|\delta_1^\phi \boldsymbol{\Sigma}^{-1}| - \log|\delta^{DS}\boldsymbol{\Sigma}^{-1}| \\
&\quad = \log|\boldsymbol{L}^{1/2}\boldsymbol{U}'\boldsymbol{U}\boldsymbol{L}^{1/2}||\phi(\boldsymbol{\lambda})\boldsymbol{I}_p||\boldsymbol{\Sigma}^{-1}| - \log|\boldsymbol{L}^{1/2}\boldsymbol{U}'\boldsymbol{U}\boldsymbol{L}^{1/2}||\boldsymbol{\Sigma}^{-1}| \\
&\quad = \log|\phi(\boldsymbol{\lambda}^*)\boldsymbol{I}_p| \\
&\quad = p\log\phi(\boldsymbol{\lambda}^*). \qquad (11)
\end{aligned}$$

The second inequality in (10) is by Sinha and Ghosh[9, Lemma 4],

$$\mathrm{E}[\boldsymbol{W}'\boldsymbol{W}|\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}] \geq (n+k)(\boldsymbol{I}_p + \boldsymbol{W}^{-1}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{W}^{-1\prime})^{-1}.$$

Combining (10) and (11),

$$\begin{aligned}
\mathrm{E}[L(\delta_1^\phi,\boldsymbol{\Sigma}) - L(\delta^{DS},\boldsymbol{\Sigma})|\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}] &\leq \frac{n+k}{n+p-1}\sum_{i=1}^p \frac{\phi(\boldsymbol{\lambda}^*)-1}{1+\lambda_i^*} - p\log\phi(\boldsymbol{\lambda}) \\
&\leq \frac{p(n+k)}{n+p-1}\frac{\phi(\boldsymbol{\lambda}^*)-1}{1+\bar{\lambda}_\cdot^*} - p\log\phi(\boldsymbol{\lambda}),
\end{aligned}$$

where $\bar{\lambda}_\cdot^* = (1/p)\sum_{i=1}^p \lambda_i^*$. Then if

$$\frac{n+k}{n+p-1}\cdot\frac{\phi(\boldsymbol{\lambda}^*)-1}{1+\bar{\lambda}_\cdot^*} - \log\phi(\boldsymbol{\lambda}) \leq 0, \qquad (12)$$

δ_1^ϕ dominates δ^{DS}.

Theorem 2.1 *If $\phi(\boldsymbol{\lambda})$ satisfies the condition (12), δ_1^ϕ dominates δ^{DS}.*

Suppose $k \geq p$. Letting $\phi^{TR1}(\boldsymbol{\lambda})$ be

$$\phi^{TR1}(\boldsymbol{\lambda}) = \min\left(1, \frac{n+p-1}{n+k}(1+\bar{\lambda}_\cdot^*)\right),$$

it is easy to show that $\phi^{TR}(\boldsymbol{\lambda})$ satisfies (12). Let the corresponding esimator be expressed as δ^{TR1}.

Corollary 2.1 *δ^{TR1} dominates δ^{DS} when $k \geq p$.*

Next we consider the class of estimators

$$\delta_2^\phi = \boldsymbol{U}\boldsymbol{L}^\phi\boldsymbol{U}', \quad \boldsymbol{L}^\phi = \mathrm{diag}\left\{\frac{\phi_i(\boldsymbol{\lambda}^*)}{n+p+1-2i}\right\}.$$

Assume that $\phi_i(\boldsymbol{\lambda}^*)$ satisfies $\phi_i(\boldsymbol{\lambda}^*) - 1 \leq 0$ and

$$\begin{aligned}\frac{\phi_i(\boldsymbol{\lambda}^*) - 1}{n+p+1-2i} = \frac{\phi_1(\boldsymbol{\lambda}^*) - 1}{n+p-1} \quad &\Leftrightarrow \quad \boldsymbol{L}^\phi - \boldsymbol{L} = \frac{\phi_1(\boldsymbol{\lambda}^*) - 1}{n+p-1}\boldsymbol{I}_p \\ &\Leftrightarrow \quad \phi_i(\boldsymbol{\lambda}^*) = \alpha_i \phi_1(\boldsymbol{\lambda}) + (1-\alpha_i), \end{aligned} \tag{13}$$

where $\alpha_i = (n+p+1-2i)/(n+p-1)$.

In the same way as (10) and (11), we have

$$\begin{aligned}\mathrm{E}[\mathrm{tr}(\delta_2^\phi - \delta^{DS})\boldsymbol{\Sigma}^{-1}|\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}] &= \mathrm{E}[\mathrm{tr}(\boldsymbol{L}^\phi - \boldsymbol{L})\boldsymbol{U}'\boldsymbol{\Sigma}^{-1}\boldsymbol{U}\ |\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}] \\ &\leq \frac{\phi_1(\boldsymbol{\lambda}^*) - 1}{n+p-1}\mathrm{tr}\mathrm{E}[\boldsymbol{W}'\boldsymbol{W}|\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}] \\ &\leq \frac{(n+k)(\phi_1(\boldsymbol{\lambda}^*) - 1)}{n+p-1}\mathrm{tr}(\boldsymbol{I}_p + \boldsymbol{W}^{-1}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{W}^{-1\prime})^{-1} \\ &= \frac{n+k}{n+p-1}\sum_{i=1}^p \frac{\phi_1(\boldsymbol{\lambda}^*) - 1}{1+\lambda_i^*}, \end{aligned} \tag{14}$$

$$\begin{aligned}\log|\delta_2^\phi \boldsymbol{\Sigma}^{-1}| - \log|\delta^{DS}\boldsymbol{\Sigma}^{-1}| &= \log|\mathrm{diag}\{\phi_i(\boldsymbol{\lambda}^*)\}| \\ &= \sum_{i=1}^p \log(\alpha_i\phi_1(\boldsymbol{\lambda}^*) + (1-\alpha_i)) \\ &\geq \log\phi_1(\boldsymbol{\lambda})\sum_{i=1}^p \alpha_i \\ &= \frac{np}{n+p-1}\log\phi_1(\boldsymbol{\lambda}). \end{aligned} \tag{15}$$

Combining (14) and (15),

$$\begin{aligned}&\mathrm{E}[L(\delta_2^\phi, \boldsymbol{\Sigma}) - L(\delta^{DS}, \boldsymbol{\Sigma})|\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}] \\ &\quad \leq \frac{n+k}{n+p-1}\sum_{i=1}^p \frac{\phi_1(\boldsymbol{\lambda}^*) - 1}{1+\lambda_i^*} - \frac{np}{n+p-1}\log\phi_1(\boldsymbol{\lambda}) \\ &\quad \leq \frac{p(n+k)}{n+p-1}\frac{\phi(\boldsymbol{\lambda}^*) - 1}{1+\bar{\lambda}_\cdot^*} - \frac{np}{n+p-1}\log\phi_1(\boldsymbol{\lambda}). \end{aligned}$$

Then with respect to $\phi_i(\boldsymbol{\lambda}^*)$ in (13) such that $\phi_1(\boldsymbol{\lambda}^*)$ satisfies

$$(n+k)\frac{\phi_1(\boldsymbol{\lambda}^*) - 1}{1+\bar{\lambda}_\cdot^*} - n\log\phi_1(\boldsymbol{\lambda}) \leq 0, \tag{16}$$

δ_2^ϕ dominates δ^{DS}.

Theorem 2.2 *If $\phi_1(\boldsymbol{\lambda})$ satisfies the condition (16), δ_2^{ϕ} dominates δ^{DS}.*

Letting $\phi^{TR2}(\boldsymbol{\lambda})$ be

$$\phi^{TR2}(\boldsymbol{\lambda}) = \min\left(1, \frac{n}{n+k}(1+\bar{\lambda}^*_\cdot)\right),$$

$\phi^{TR2}(\boldsymbol{\lambda})$ satisfies (16). Let the corresponding esimator be expressed as δ^{TR2}.

Corollary 2.2 *δ^{TR2} dominates δ^{DS} for all k and $p \le n$.*

3 Inadmissibility of other minimax estimators

Using the argument in the previous section we can show the inadmissibility of order-preserving estimators modifying δ^{DS} proposed by Sheena and Takemura[12] when $k \ge p$. Sheena and Takemura[12] proposed two methods of modifying δ^{DS}.

One is the method using order statistics. The estimator is

$$\delta^{OS} = \boldsymbol{H}\boldsymbol{L}^{OS}\boldsymbol{H}', \quad \boldsymbol{L}^{OS} = \mathrm{diag}(\psi_1^{OS}(\boldsymbol{\lambda}), \ldots, \psi_p^{OS}(\boldsymbol{\lambda})),$$

where $\psi_j^{OS}(\boldsymbol{\lambda})$ are the jth largest element in $\tilde{\lambda}_i = \lambda_i/(n+p+1-2i)$, $i = 1, \ldots, p$.

The other is the isotonic regression of $\tilde{\lambda}_i$. The estimator is

$$\delta^{IR} = \boldsymbol{H}\boldsymbol{L}^{IR}\boldsymbol{H}', \quad \boldsymbol{L}^{IR} = \mathrm{diag}(\psi_1^{IR}(\boldsymbol{\lambda}), \ldots, \psi_p^{IR}(\boldsymbol{\lambda})),$$

where $\psi_j^{IR}(\boldsymbol{\lambda})$ are the solutions of

$$\min_{\psi^{IR} \in \mathcal{F}} \sum_{i=1}^{p} (\psi_i^{IR}(\boldsymbol{\lambda}) - \tilde{\lambda}_i)^2,$$

where $\mathcal{F} = \{\psi^{IR} = (\psi_1^{IR}(\boldsymbol{\lambda}), \ldots, \psi_p^{IR}(\boldsymbol{\lambda})) | \psi_1^{IR}(\boldsymbol{\lambda}) \ge \cdots \ge \psi_p^{IR}(\boldsymbol{\lambda})\}$. Both of δ^{OS} and δ^{IR} was shown to improve on δ^{DS} under the loss (2). On the analogy of (7) we consider the classes of estimators

$$\delta_\phi^{OS} = \phi^{OS}(\boldsymbol{\lambda}^*)\boldsymbol{H}\boldsymbol{L}^{OS}\boldsymbol{H}',$$

$$\delta_\phi^{IR} = \phi^{IR}(\boldsymbol{\lambda}^*)\boldsymbol{H}\boldsymbol{L}^{IR}\boldsymbol{H}'.$$

We suppose that $\phi^{OS}(\boldsymbol{\lambda}^*) \le 1$ and $\phi^{IR}(\boldsymbol{\lambda}^*) \le 1$.

Rewrite δ^{OS} and δ^{IR} as

$$\delta^{OS} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}\boldsymbol{L}^{OS*}\boldsymbol{\Lambda}^{1/2}\boldsymbol{H}', \quad \boldsymbol{L}^{OS*} = \boldsymbol{\Lambda}^{-1}\boldsymbol{L}^{OS} = \mathrm{diag}\{\psi_i^{OS}(\boldsymbol{\lambda})/\lambda_i\},$$

$$\delta^{IR} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}\boldsymbol{L}^{IR*}\boldsymbol{\Lambda}^{1/2}\boldsymbol{H}', \quad \boldsymbol{L}^{IR*} = \boldsymbol{\Lambda}^{-1}\boldsymbol{L}^{IR} = \mathrm{diag}\{\psi_i^{IR}(\boldsymbol{\lambda})/\lambda_i\},$$

respectively. Now we present the following lemma.

Lemma 3.1

$$\boldsymbol{L}^{OS*} \ge \frac{1}{n+p-1}\boldsymbol{I}_p, \quad \boldsymbol{L}^{IR*} \ge \frac{1}{n+p-1}\boldsymbol{I}_p.$$

(Proof)

The $\psi_i^0(\boldsymbol{\lambda})/\lambda_i$ can be expressed as

$$\frac{\psi_i^0(\boldsymbol{\lambda})}{\lambda_i} = \frac{1}{n+p+1-2j}\frac{\lambda_j}{\lambda_i}, \quad 1 \le j \le p.$$

When $i \ge j$, $\lambda_i \le \lambda_j$. Therefore

$$\frac{1}{n+p+1-2j}\frac{\lambda_j}{\lambda_i} \ge \frac{1}{n+p+1-2j} \ge \frac{1}{n+p-1}. \tag{17}$$

When $i \le j$, λ_i and λ_j satisfy

$$\frac{\lambda_j}{\lambda_i} \ge \frac{n+1+p-2j}{n+1+p-2i}$$

by the definition of $\boldsymbol{L}^{OS}$, which also implies (17). Then $\boldsymbol{L}^{OS*} \ge (1/(n+p-1))\boldsymbol{I}_p$ is proved.

According to Sheena and Takemura[12], $\psi_i^{IR}(\boldsymbol{\lambda})$ can be expressed with some constants a, b, $1 \le a \le b \le p$ as

$$\psi_i^{IR}(\boldsymbol{\lambda}) = \frac{1}{b-a-1}\sum_{j=a}^{b}\frac{\lambda_j}{n+p+1-2j}.$$

Using the above argument we can prove the latter one similarly. □

In the same way as (10) and (11) we have by using the Lemma 3.1

$$\mathrm{E}[\mathrm{tr}(\delta_\phi^{OS} - \delta^{OS})\boldsymbol{\Sigma}^{-1}|\boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}] \le \sum_{i=1}^{p}\frac{(n+k)}{n+p-1}\frac{\phi^{OS}(\boldsymbol{\lambda}^*)-1}{1+\lambda_i^*},$$

$$\log |\delta_\phi^{OS} \boldsymbol{\Sigma}^{-1}| - \log |\delta^{OS} \boldsymbol{\Sigma}^{-1}| = p \log \phi^{OS}(\boldsymbol{\lambda}^*),$$

$$\mathrm{E}[\mathrm{tr}(\delta_\phi^{IR} - \delta^{IR})\boldsymbol{\Sigma}^{-1} | \boldsymbol{W}^{-1}\tilde{\boldsymbol{X}}] \leq \sum_{i=1}^{p} \frac{(n+k)}{n+p-1} \frac{\phi^{IR}(\boldsymbol{\lambda}^*) - 1}{1 + \lambda_i^*}.$$

$$\log |\delta_\phi^{IR} \boldsymbol{\Sigma}^{-1}| - \log |\delta^{OS} \boldsymbol{\Sigma}^{-1}| = p \log \phi^{IR}(\boldsymbol{\lambda}^*).$$

Therefore we can obtain the following results.

Theorem 3.1 *When $\phi^{OS}(\boldsymbol{\lambda}^*)$ satisfy the condition (12), δ_ϕ^{OS} dominates δ^{OS}. Similarly when $\phi^{IR}(\boldsymbol{\lambda}^*)$ satisfy the condition (12), δ_ϕ^{IR} dominates δ^{IR}.*

Corollary 3.1 *Assume $k \geq p$. Then $\delta^{TR3} = \phi^{TR1}(\boldsymbol{\lambda}^*)\delta^{OS}$ dominates δ^{OS} and $\delta^{TR4} = \phi^{TR1}(\boldsymbol{\lambda}^*)\delta^{IR}$ dominates δ^{IR}.*

4 Monte Carlo Study

We study the risk performance of the proposed estimators with some Monte Carlo studies. We compare the average losses of δ^{DS}, δ^{TR1} and δ^{TR2} of Section2, δ^{OS}, δ^{TR3}, δ^{IR} and δ^{TR4} of Section 3. We present in Table 1 to 4 the average losses of the seven estimators over 100000 replications for $p = 3$ and some combinations of $(n, k, \|\boldsymbol{\mu}\|, \boldsymbol{\Sigma})$. Since the risks of the seven estimators depend only on the eigenvalue of $\boldsymbol{\Sigma}$, we set $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. We note the risks of δ^{DS} and δ^{TR1} are identical with those of δ^{OS} and δ^{TR3}, respectively, when $\boldsymbol{\Sigma} = \boldsymbol{I}_p$.

The summary of the experiment is as follows.

- Although the improvement is not large, we can see the dominance of the estimator proposed in this article over δ^{DS}, δ^{OS},δ^{IR}.

- When σ_1^2, σ_2^2 and σ_3^2 are close together, the proposed estimator, especially δ^{TR2}, save much risk.

- The improvement is on the whole in proportion to degrees of freedom of $\boldsymbol{S}$ and in inverse proportion to $\|\boldsymbol{\mu}\|$ except δ^{TR2}.

- The improvement is not always in proportion to k. In this sense the estimator proposed here may not use the information of $\boldsymbol{X}$ effectively.

Acknowledgement

The author wishes to thank Professor Tatsuya Kubokawa of University of Tokyo for his helpful comments and advice.

References

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

[2] L. R. Haff, The variational form of certain Bayes estimators, *Ann. Statist.* **19** (1991), 1163-1190.

[3] H. Hara, Estimating a covariance matrix in MANOVA model, Technical Report of Depertment of Mathematical Engineering and Information Physics. (METR 98-7), University of Tokyo, 1998.

[4] W. James and C. Stein, Estimation with quadratic less. in *Proc. 4th Berkeley Symp. Math. Statist.* (J. Neyman, ed.), Vol. 1, pp. 361-379, Univ. of Carfornia Press, 1961.

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

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

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

[8] F. Perron, Minimax estimators of a Covariance Matrix, *J. Multivariate Annal.* **43** (1992) 16-28.

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

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

[11] C. Stein, Unpublished notes on estimating the covariance matrix, 1977.

[12] Y. Sheena and A. Takemura, Inadmissibility of non-order-preserving orthogonally invariant estimators of the covariance matrix in the case of Stein's loss, *J. Multivariate Annal.* **41** (1992), 117-131.

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

Table 1. Average losses forδ^{TR1} for $p = 3$.

n	σ_1^2	σ_2^2	σ_3^2	k	$\|\boldsymbol{\mu}_i\|^2$ 0.0	1.0	2.0
5	1.00	1.00	1.00	5	0.99915	0.99955	0.99967
				7	0.99875	0.99952	0.99968
	$R(\delta^{DS}) = 0.99972$			10	0.99847	0.99942	0.99965
	1.00	0.90	0.80	5	1.00502	1.00529	1.00538
				7	1.00480	1.00525	1.00538
	$R(\delta^{DS}) = 1.00541$			10	1.00466	1.00528	1.00538
	1.00	0.70	0.50	5	1.02976	1.02985	1.02987
				7	1.02970	1.02984	1.02987
	$R(\delta^{DS}) = 1.02988$			10	1.02973	1.02986	1.02987
10	1.00	1.00	1.00	5	0.45876	0.45928	0.45942
				7	0.45824	0.45917	0.45941
	$R(\delta^{DS}) = 0.45946$			10	0.45755	0.45909	0.45940
	1.00	0.90	0.80	5	0.46006	0.46041	0.46050
				7	0.45976	0.46034	0.46048
	$R(\delta^{DS}) = 0.46052$			10	0.45948	0.46032	0.46050
	1.00	0.70	0.50	5	0.48126	0.48136	0.48140
				7	0.48117	0.48135	0.48140
	$R(\delta^{DS}) = 0.48140$			10	0.48115	0.48138	0.48140

Table 2. Average losses forδ^{TR2} for $p = 3$.

n	σ_1^2	σ_2^2	σ_3^2	k	$\|\boldsymbol{\mu}_i\|^2$ 0.0	1.0	2.0
5	1.00	1.00	1.00	5	0.98956	0.97817	0.97775
				7	0.98447	0.97801	0.97794
	$R(\delta^{DS}) = 0.99972$			10	0.97965	0.97785	0.97819
	1.00	0.90	0.80	5	0.99707	0.98640	0.98474
				7	0.99302	0.98590	0.98475
	$R(\delta^{DS}) = 1.00541$			10	0.98821	0.98525	0.98468
	1.00	0.70	0.50	5	1.02545	1.02007	1.01883
				7	1.02372	1.01971	1.01878
	$R(\delta^{DS}) = 1.02988$			10	1.02136	1.01925	1.01867
10	1.00	1.00	1.00	5	0.45946	0.44933	0.44677
				7	0.45946	0.44933	0.44677
	$R(\delta^{DS}) = 0.45946$			10	0.44932	0.44677	0.44608
	1.00	0.90	0.80	5	0.46052	0.45331	0.45146
				7	0.46052	0.45331	0.45146
	$R(\delta^{DS}) = 0.46052$			10	0.45331	0.45146	0.45103
	1.00	0.70	0.50	5	0.48140	0.47865	0.47815
				7	0.48140	0.47865	0.47815
	$R(\delta^{DS}) = 0.48140$			10	0.47865	0.47815	0.47807

Table 3. Average losses forδ^{TR3} for $p = 3$.

n	σ_1^2	σ_2^2	σ_3^2	k	$\|\|\boldsymbol{\mu}_i\|\|^2$ 0.0	1.0	2.0
5	1.00	0.90	0.80	5	1.00493	1.00520	1.00528
				7	1.00471	1.00516	1.00529
	$R(\delta^{OS}) = 1.00531$			10	1.00456	1.00519	1.00520
	1.00	0.70	0.50	5	1.02908	1.02917	1.02918
				7	1.02902	1.02916	1.02919
	$R(\delta^{OS}) = 1.02919$			10	1.02906	1.02918	1.02919
10	1.00	0.90	0.80	5	0.46004	0.46038	0.46048
				7	0.45974	0.46031	0.46046
	$R(\delta^{OS}) = 0.46050$			10	0.45945	0.46029	0.46047
	1.00	0.70	0.50	5	0.48109	0.48119	0.48123
				7	0.48100	0.48118	0.48123
	$R(\delta^{OS}) = 0.48123$			10	0.48098	0.48121	0.48123

Table 4. Average losses forδ^{TR4} for $p = 3$.

n	σ_1^2	σ_2^2	σ_3^2	k	$\lVert\boldsymbol{\mu}_i\rVert^2$		
					0.0	1.0	2.0
5	1.00	1.00	1.00	5	0.99739	0.99779	0.99791
				7	0.99699	0.99776	0.99792
	$R(\delta^{IR}) = 0.99972$			10	0.99671	0.99766	0.99789
	1.00	0.90	0.80	5	1.00319	1.00345	1.00354
				7	1.00296	1.00341	1.00354
	$R(\delta^{IR}) = 1.00357$			10	1.00282	1.00531	1.00354
	1.00	0.70	0.50	5	1.02782	1.02791	1.02793
				7	1.02777	1.02790	1.02793
	$R(\delta^{IR}) = 1.02794$			10	1.02780	1.02792	1.02793
10	1.00	1.00	1.00	5	0.45853	0.45906	0.45919
				7	0.45801	0.45895	0.45919
	$R(\delta^{IR}) = 0.45924$			10	0.45733	0.45887	0.45917
	1.00	0.90	0.80	5	0.45983	0.46018	0.46027
				7	0.45953	0.46011	0.46025
	$R(\delta^{IR}) = 0.46029$			10	0.45924	0.46008	0.46026
	1.00	0.70	0.50	5	0.48099	0.48109	0.48113
				7	0.48090	0.48109	0.48113
	$R(\delta^{IR}) = 0.48113$			10	0.48088	0.48111	0.48113