Simultaneous estimation of normal precision matrices

Hisayuki Tsukuma* and Tatsuya Kubokawa†
Toho University and University of Tokyo

December 27, 2006

Abstract

This paper treats the problem of simultaneously estimating the precision matrices in multivariate normal distributions. A condition for improvement on the unbiased estimators of the precision matrices is derived under a quadratic loss function. The improvement condition is similar to the superharmonic condition established by Stein (1981). The condition allows us not only to provide various alternative estimators such as shrinkage type and enlargement type estimators for the unbiased estimators, but also to present a condition on a prior density under which the resulting generalized Bayes estimators dominate the unbiased estimators. Also, a unified method improving upon both the shrinkage and the enlargement type estimators is discussed.

Key words and phrases: Bayes estimation, common mean, decision theory, James-Stein estimator, risk function, simultaneous estimation, superharmonic function.

1 Introduction

There have been many papers to treat the problem of estimating the precision matrix in a multivariate normal distribution and proposed various types of estimators for some loss functions. These papers include Efron and Morris (1976), Haff (1977, 1979), Dey (1987), Krishnamoorshy and Gupta (1989), and Kubokawa (2005). For the motivation of the problem of estimating the precision matrix, see Efron and Morris (1976), Haff (1986) and Kubokawa (2005). In this paper we treat an extended model to a k-sample problem and consider simultaneous estimation of the precision matrices under a quadratic loss function. The main aim is to derive estimators of the precision matrices by means of applying Stein (1981)'s idea to our problem.

*Faculty of Medicine, Toho University, 5-21-16 Omori-nishi, Ota-ku, Tokyo 143-8540, JAPAN, E-Mail: tsukuma@med.toho-u.ac.jp
†Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: tatsuya@e.u-tokyo.ac.jp

To specify the problem considered here, let $\boldsymbol{S}_1, \dots, \boldsymbol{S}_k$ be mutually independent random matrices such that

$$\boldsymbol{S}_i \sim \mathcal{W}_{p_i}(\boldsymbol{\Sigma}_i, n_i), \quad i = 1, \dots, k, \tag{1.1}$$

where for $i = 1, \dots, k$, $\boldsymbol{S}_i$ is a $p_i \times p_i$ matrix, $\boldsymbol{\Sigma}_i$ is a $p_i \times p_i$ unknown positive-definite matrix and $n_i - p_i - 3 > 0$. Consider simultaneous estimation of the precision matrices, denoted by $\boldsymbol{\Sigma}^{-1} = (\boldsymbol{\Sigma}_1^{-1}, \dots, \boldsymbol{\Sigma}_k^{-1})$, under the quadratic loss function

$$L(\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = \sum_{i=1}^{k} \mathrm{tr}\,(\boldsymbol{\delta}_i - \boldsymbol{\Sigma}_i^{-1})^2, \tag{1.2}$$

where $\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_k)$ is an estimator of $\boldsymbol{\Sigma}^{-1}$. Every estimator is evaluated by the risk function $R(\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = E[L(\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1})]$.

An ordinary estimator of $\boldsymbol{\Sigma}^{-1}$ is $\boldsymbol{\delta}^c = (\boldsymbol{\delta}_1^c, \dots, \boldsymbol{\delta}_k^c)$, where $\boldsymbol{\delta}_i^c = c_i \boldsymbol{S}_i^{-1}$ and the c_i's are positive constants. Using Theorem 3.2 of Haff (1979), we can write the risk of $\boldsymbol{\delta}^c$ as

$$\begin{aligned} R(\boldsymbol{\delta}^c, \boldsymbol{\Sigma}^{-1}) &= \sum_{i=1}^{k} E[\mathrm{tr}\,(\boldsymbol{\delta}_i^c - \boldsymbol{\Sigma}_i^{-1})^2] \\ &= \sum_{i=1}^{k} \{a_{1,i} c_i^2 (\mathrm{tr}\,\boldsymbol{\Sigma}_i^{-1})^2 + a_{2,i} c_i^2 \mathrm{tr}\,\boldsymbol{\Sigma}_i^{-2} - 2a_{3,i} c_i \mathrm{tr}\,\boldsymbol{\Sigma}_i^{-2} + \mathrm{tr}\,\boldsymbol{\Sigma}_i^{-2}\}. \end{aligned}$$

where $a_{1,i} = \{(n_i - p_i)(n_i - p_i - 1)(n_i - p_i - 3)\}^{-1}$, $a_{2,i} = \{(n_i - p_i)(n_i - p_i - 3)\}^{-1}$, and $a_{3,i} = (n_i - p_i - 1)^{-1}$. Since the constants c_i's minimizing the risk depend on the unknown parameters $\boldsymbol{\Sigma}_i$'s, there are no optimal constants c_i's. A natural choice of c_i is $c_i = n_i - p_i - 1$, which leads to the unbiased estimator of $\boldsymbol{\Sigma}^{-1}$, given by

$$\boldsymbol{\delta}^{UB} = (\boldsymbol{\delta}_1^{UB}, \dots, \boldsymbol{\delta}_k^{UB}) = ((n_1 - p_1 - 1)\boldsymbol{S}_1^{-1}, \dots, (n_k - p_k - 1)\boldsymbol{S}_k^{-1}).$$

In this paper, we consider the problem of constructing estimators improving on the unbiased estimator $\boldsymbol{\delta}^{UB}$. Especially, we develop an interesting dominance condition corresponding to the superharmonic condition given by Stein (1981), who derived it in simultaneous estimation of a multivariate normal mean vector. More specifically, let $f(\boldsymbol{S})$ be a scalar-valued function of $\boldsymbol{S} = (\boldsymbol{S}_1, \dots, \boldsymbol{S}_k)$, where f is a twice differentiable function and $f(\boldsymbol{S}) > 0$. For $i = 1, \dots, k$, let $\mathcal{D}_i$ be a $p_i \times p_i$ matrix of differential operators with respect to $\boldsymbol{S}_i = (s_{i\cdot ab})$ such as

$$\mathcal{D}_i = \left(\frac{1}{2}(1 + \delta_{ab})\frac{\partial}{\partial s_{i\cdot ab}}\right),$$

where $\delta_{ab} = 1$ for $a = b$ and $\delta_{ab} = 0$ for $a \neq b$. The estimator considered in this paper is $\boldsymbol{\delta}^f = (\boldsymbol{\delta}_1^f, \dots, \boldsymbol{\delta}_k^f)$ with the form

$$\boldsymbol{\delta}_i^f = \boldsymbol{\delta}_i^{UB} - 4\mathcal{D}_i \log f(\boldsymbol{S}) = \boldsymbol{\delta}_i^{UB} - \frac{4}{f(\boldsymbol{S})}\mathcal{D}_i f(\boldsymbol{S}). \tag{1.3}$$

An interesting fact is that the same idea and arguments as in Stein (1981) can be applied to evaluate the risk function of the estimator $\boldsymbol{\delta}^f$. In Section 2, we derive the condition

$$\sum_{i=1}^{k} \operatorname{tr} \mathcal{D}_i \mathcal{D}_i^t f(\boldsymbol{S}) < 0, \tag{1.4}$$

under which $\boldsymbol{\delta}^f$ dominates $\boldsymbol{\delta}^{UB}$. The condition (1.4) corresponds to Stein's superharmonic condition as noted in Remark 2.1, while it does not imply that $f(\boldsymbol{S})$ is superharmonic. Hence, the improvement over the unbiased estimator can be shown by checking the condition (1.4), and various types of improved estimators are developed in Sections 3, 4 and 5. In Section 3, four kinds of shrinkage and enlargement estimators improving on $\boldsymbol{\delta}^{UB}$ are presented. In Section 4, we handle the generalized Bayes procedure and provide a condition on a prior distribution of $\boldsymbol{\Sigma}^{-1}$ under which the resulting generalized Bayes estimator dominates $\boldsymbol{\delta}^{UB}$. An empirical Bayes method is discussed in Section 5 and it is shown that an Efron and Morris (1976)-type estimators are characterized as the empirical Bayes estimators. Section 6 gives the unified dominance result of both shrinkage and enlargement estimators which improve upon $\boldsymbol{\delta}^{UB}$. Section 7 presents the numerical comparison of the risk behavior of alternative estimators and shows that certain alternative estimator substantially reduces risk over $\boldsymbol{\delta}^{UB}$ in case that the precision matrices $\boldsymbol{\Sigma}_i^{-1}$'s are near the identity matrices.

Finally, it may be noted that the simultaneous estimation of $\boldsymbol{\Sigma}^{-1}$ is involved in the following estimation problems: (i) Consider the common mean of multivariate normal distributions, $\boldsymbol{X}_i \sim \mathcal{N}_p(\boldsymbol{\theta}, \boldsymbol{\Sigma}_i)$, for $i = 1, \ldots, k$ where the Wishart matrices (1.1) with $p_1 = \cdots = p_k = p$ are available. If the $\boldsymbol{\Sigma}_i$'s are known, the best linear unbiased estimator of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}}_0 = \Big(\sum_{i=1}^{k} \boldsymbol{\Sigma}_i^{-1}\Big)^{-1} \Big(\sum_{i=1}^{k} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{X}_i\Big).$$

Hence it is necessary to replace the $\boldsymbol{\Sigma}_i^{-1}$'s with their estimators when the $\boldsymbol{\Sigma}_i^{-1}$'s are unknown; (ii) Consider the k-sample problem of simultaneously estimating the normal mean matrices, $\boldsymbol{\Theta}_i$'s, with the identity covariance matrices under the loss $\sum_{i=1}^{k} \operatorname{tr} (\widehat{\boldsymbol{\Theta}}_i - \boldsymbol{\Theta}_i)^t (\widehat{\boldsymbol{\Theta}}_i - \boldsymbol{\Theta}_i)$. From the arguments of Efron and Morris (1976), the problem resolves itself into that of estimating $\boldsymbol{\Sigma}^{-1}$ under a quadratic loss function. Since we need to consider loss functions different from (1.2) to handle the problems (i) and (ii), the results given in this paper can not be directly applied to these problems. However, the ideas and methods used here will help us develop improved estimators in the problems.

2 Condition of dominance over the unbiased estimator

In this section we derive the condition (1.4) under which $\boldsymbol{\delta}^f$ dominates $\boldsymbol{\delta}^{UB}$ relative to the loss (1.2).

We begin with describing some matrix operations. Let $\boldsymbol{A} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k)$ and $\boldsymbol{B} = (\boldsymbol{B}_1, \ldots, \boldsymbol{B}_k)$, where $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are $p_i \times p_i$ squared matrices, respectively, for $i = 1, \ldots, k$.

Denote $\phi \boldsymbol{A} = (\phi \boldsymbol{A}_1, \dots, \phi \boldsymbol{A}_k)$ for a scalar ϕ and $\boldsymbol{A} - \boldsymbol{B} = (\boldsymbol{A}_1 - \boldsymbol{B}_1, \dots, \boldsymbol{A}_k - \boldsymbol{B}_k)$. Define the notations $\boldsymbol{A} \circ \boldsymbol{B}$ and $\|\boldsymbol{A}\|_M$ by

$$\boldsymbol{A} \circ \boldsymbol{B} = \sum_{i=1}^{k} \mathrm{tr}\, \boldsymbol{A}_i \boldsymbol{B}_i^t, \qquad \|\boldsymbol{A}\|_M = \sqrt{\boldsymbol{A} \circ \boldsymbol{A}}. \tag{2.1}$$

Then the quadratic loss function (1.2) is written as

$$L(\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = \|\boldsymbol{\delta} - \boldsymbol{\Sigma}^{-1}\|_M^2. \tag{2.2}$$

Let $h(\boldsymbol{S})$ be a scalar-valued function of $\boldsymbol{S}$ and $\boldsymbol{H}_i(\boldsymbol{S}) = (h_{i \cdot ab})$ a matrix-valued function of $\boldsymbol{S}$. The actions of $\mathcal{D}_i$ on $h(\boldsymbol{S})$ and on $\boldsymbol{H}_i(\boldsymbol{S})$ are defined as, respectively,

$$\mathcal{D}_i h(\boldsymbol{S}) = \Big(\frac{1}{2}(1 + \delta_{ab}) \frac{\partial h(\boldsymbol{S})}{\partial s_{i \cdot ab}} \Big), \qquad \mathcal{D}_i \boldsymbol{H}_i(\boldsymbol{S}) = \Big(\sum_{c=1}^{p_i} \frac{1 + \delta_{ac}}{2} \frac{\partial h_{i \cdot cb}}{\partial s_{i \cdot ac}} \Big),$$

where $\mathcal{D}_i h(\boldsymbol{S})$ and $\mathcal{D}_i \boldsymbol{H}_i(\boldsymbol{S})$ are $p_i \times p_i$ matrices. Also, the actions of $\mathcal{D} = (\mathcal{D}_1, \dots, \mathcal{D}_k)$ on $h(\boldsymbol{S})$ and on $\boldsymbol{H}(\boldsymbol{S}) = (\boldsymbol{H}_1(\boldsymbol{S}), \dots, \boldsymbol{H}_k(\boldsymbol{S}))$ are defined as, respectively,

$$\mathcal{D} h(\boldsymbol{S}) = (\mathcal{D}_1 h(\boldsymbol{S}), \dots, \mathcal{D}_k h(\boldsymbol{S})), \qquad \mathcal{D} \circ \boldsymbol{H}(\boldsymbol{S}) = \sum_{i=1}^{k} \mathrm{tr}\, \mathcal{D}_i \boldsymbol{H}_i(\boldsymbol{S}).$$

Then the estimator (1.3) is written as

$$\boldsymbol{\delta}^f = \boldsymbol{\delta}^{UB} - 4\mathcal{D} \log f(\boldsymbol{S}). \tag{2.3}$$

To evaluate the risk, we use the Wishart identity given by

$$E_i[\mathrm{tr}\, \boldsymbol{\Sigma}_i^{-1} \boldsymbol{H}_i(\boldsymbol{S})] = E_i[(n_i - p_i - 1) \mathrm{tr}\, \boldsymbol{S}_i^{-1} \boldsymbol{H}_i(\boldsymbol{S}) + 2 \mathrm{tr}\, \mathcal{D}_i \boldsymbol{H}_i(\boldsymbol{S})],$$

provided both expectations exist. Here E_i denotes conditional expectation of $\boldsymbol{S}_i$ given $\boldsymbol{S}_1, \dots, \boldsymbol{S}_{i-1}, \boldsymbol{S}_{i+1}, \dots, \boldsymbol{S}_k$. The Wishart identity is equivalent to

$$E_i[\mathrm{tr}\, (\boldsymbol{\delta}_i^{UB} - \boldsymbol{\Sigma}_i^{-1}) \boldsymbol{H}_i(\boldsymbol{S})] = E_i[-2 \mathrm{tr}\, \mathcal{D}_i \boldsymbol{H}_i(\boldsymbol{S})].$$

Using this identity gives that

$$\begin{aligned} E\Big[\sum_{i=1}^{k} \mathrm{tr}\, (\boldsymbol{\delta}_i^{UB} - \boldsymbol{\Sigma}_i^{-1}) \boldsymbol{H}_i(\boldsymbol{S}) \Big] &= \sum_{i=1}^{k} E_{(i)} \Big[E_i \Big[\mathrm{tr}\, (\boldsymbol{\delta}_i^{UB} - \boldsymbol{\Sigma}_i^{-1}) \boldsymbol{H}_i(\boldsymbol{S}) \Big] \Big] \\ &= \sum_{i=1}^{k} E_{(i)} \Big[E_i \Big[-2 \mathrm{tr}\, \mathcal{D}_i \boldsymbol{H}_i(\boldsymbol{S}) \Big] \Big] \\ &= E \Big[-2 \sum_{i=1}^{k} \mathrm{tr}\, \mathcal{D}_i \boldsymbol{H}_i(\boldsymbol{S}) \Big], \end{aligned}$$

where $E_{(i)}$ denotes expectation of $\boldsymbol{S}_1, \dots, \boldsymbol{S}_{i-1}, \boldsymbol{S}_{i+1}, \dots, \boldsymbol{S}_k$. Similar to the notation (2.1), we have the extended Wishart identity

$$E[(\boldsymbol{\delta}^{UB} - \boldsymbol{\Sigma}^{-1}) \circ \boldsymbol{H}(\boldsymbol{S})] = -2E[\mathcal{D} \circ \boldsymbol{H}(\boldsymbol{S})]. \tag{2.4}$$

Then the extended Wishart identity (2.4) is used to get our main result.

Theorem 2.1 *The risk function of $\boldsymbol{\delta}^f = \boldsymbol{\delta}^{UB} - 4\mathcal{D}\log f(\boldsymbol{S})$ is expressed as*

$$R(\boldsymbol{\delta}^f, \boldsymbol{\Sigma}^{-1}) = R(\boldsymbol{\delta}^{UB}, \boldsymbol{\Sigma}^{-1}) + E\Big[\frac{16}{f(\boldsymbol{S})}\mathcal{D}\circ\mathcal{D}f(\boldsymbol{S})\Big].$$

Hence, if

$$\mathcal{D}\circ\mathcal{D}f(\boldsymbol{S}) = \sum_{i=1}^{k} \operatorname{tr}\mathcal{D}_i\mathcal{D}_i^t f(\boldsymbol{S}) < 0, \tag{2.5}$$

then $\boldsymbol{\delta}^f$ dominates $\boldsymbol{\delta}^{UB}$ relative to the loss (2.2).

Proof. It is observed that

$$\begin{aligned} R(\boldsymbol{\delta}^f, \boldsymbol{\Sigma}^{-1}) &= R(\boldsymbol{\delta}^{UB}, \boldsymbol{\Sigma}^{-1}) + E[-8(\boldsymbol{\delta}^{UB} - \boldsymbol{\Sigma}^{-1})\circ\mathcal{D}\log f(\boldsymbol{S}) + 16\|\mathcal{D}\log f(\boldsymbol{S})\|_M^2] \\ &= R(\boldsymbol{\delta}^{UB}, \boldsymbol{\Sigma}^{-1}) + E[16\mathcal{D}\circ\mathcal{D}\log f(\boldsymbol{S}) + 16\|\mathcal{D}\log f(\boldsymbol{S})\|_M^2], \end{aligned} \tag{2.6}$$

where the last equality follows from the Wishart identity (2.4). Note that $\mathcal{D}\log f(\boldsymbol{S}) = \{f(\boldsymbol{S})\}^{-1}\mathcal{D}f(\boldsymbol{S})$ and that

$$\mathcal{D}\circ\mathcal{D}\log f(\boldsymbol{S}) = \frac{\mathcal{D}\circ\mathcal{D}f(\boldsymbol{S})}{f(\boldsymbol{S})} - \frac{\|\mathcal{D}f(\boldsymbol{S})\|_M^2}{f^2(\boldsymbol{S})} = \frac{\mathcal{D}\circ\mathcal{D}f(\boldsymbol{S})}{f(\boldsymbol{S})} - \|\mathcal{D}\log f(\boldsymbol{S})\|_M^2.$$

Therefore combining the above facts and (2.6) completes the proof. ■

Remark 2.1 Theorem 2.1 is motivated by Stein (1981), and his result is stated here briefly. Let $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\theta}, \boldsymbol{I}_p)$, where $\boldsymbol{\theta}$ is an unknown mean vector and $\boldsymbol{I}_p$ denotes the identity matrix. Consider the estimation of $\boldsymbol{\theta}$ under the quadratic loss function $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2$ where $\widehat{\boldsymbol{\theta}}$ is an estimator of $\boldsymbol{\theta}$. Let $g(\boldsymbol{x})$ be a real-valued and twice differentiable function of $\boldsymbol{x} \in \mathbb{R}^p$ and also let ∇ be the vector differential operator of first partial derivatives with i-th coordinate $\partial/\partial x_i$. Then the estimator of the form

$$\boldsymbol{X} + \nabla\log g(\boldsymbol{X})$$

dominates the maximum likelihood estimator $\boldsymbol{X}$ under the quadratic loss if

$$\nabla^t\nabla\sqrt{g(\boldsymbol{x})} = \sum_{i=1}^{p}\frac{\partial^2\sqrt{g(\boldsymbol{x})}}{\partial x_i^2} < 0,$$

which is equivalent to the function $\sqrt{g(\boldsymbol{X})}$ being superharmonic. Although the condition (2.5) corresponds to this superharmonic condition, it does not imply that $f(\boldsymbol{S})$ is superharmonic, since the (a,b)-element of $\mathcal{D}_i$ is given by $(1/2)(1+\delta_{ab})\partial/\partial s_{i\cdot ab}$. ■

Remark 2.2 We can treat the estimation problem of $\boldsymbol{\Sigma}^{-1}$ under the Kullback-Leibler type loss function, namely, the so-called Stein loss function

$$L_S(\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) = \sum_{i=1}^{k}\{\operatorname{tr}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\delta}_i - \log|\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\delta}_i| - p_i\}.$$

However the dominance condition, such as (2.5), for this loss function can not be derived since it is hard to evaluate the trace and logarithmic terms in the loss function. ■

3 Alternative estimators for improvement

3.1 Example of alternative estimators

In this subsection we shall apply Theorem 2.1 to some functions for f and give alternative estimators of $\boldsymbol{\Sigma}^{-1}$ for improving on the unbiased estimator $\boldsymbol{\delta}^{UB}$. The detailed derivations of the alternative estimators and the proofs for results given in this subsection are put in Subsection 3.2.

Let g_1, g_2 and g_3 be, respectively, scalar-valued functions and assume that g_j, $j = 1, 2, 3$, are positive and twice differentiable functions. Denote by g_j' and g_j'', respectively, the first and second derivatives of g_j. Also let α, β and γ be constants. Consider the following functions:

(1) $f_{EM}(\boldsymbol{S}) = t^{\alpha} g_1(t)$, where $t = \sum_{i=1}^{k} \operatorname{tr} \boldsymbol{S}_i$,

(2) $f_{JS}(\boldsymbol{S}) = u^{-\beta} g_2(u^2)$, where $u = \|\boldsymbol{S}\|_M$,

(3) $f_{US}(\boldsymbol{S}) = v^{\gamma} g_3(v)$, where $v = \prod_{i=1}^{k} |\boldsymbol{S}_i|$.

Then from Lemma 3.2 given below, the corresponding estimators can be expressed as, respectively,

$$\begin{aligned}
\boldsymbol{\delta}^{EM} &= \boldsymbol{\delta}^{UB} - 4\mathcal{D} \log f_{EM}(\boldsymbol{S}) = \boldsymbol{\delta}^{UB} - 4\Big(\frac{\alpha}{t} + \frac{g_1'(t)}{g_1(t)}\Big)\boldsymbol{I}, \\
\boldsymbol{\delta}^{JS} &= \boldsymbol{\delta}^{UB} - 4\mathcal{D} \log f_{JS}(\boldsymbol{S}) = \boldsymbol{\delta}^{UB} + 4\Big(\frac{\beta}{u^2} - \frac{2g_2'(u^2)}{g_2(u)}\Big)\boldsymbol{S}, \\
\boldsymbol{\delta}^{US} &= \boldsymbol{\delta}^{UB} - 4\mathcal{D} \log f_{US}(\boldsymbol{S}) = \boldsymbol{\delta}^{UB} - 4\Big(\gamma + \frac{v g_3'(v)}{g_3(v)}\Big)\boldsymbol{S}^{-1},
\end{aligned}$$

where $\boldsymbol{I} = (\boldsymbol{I}_{p_1}, \ldots, \boldsymbol{I}_{p_k})$ and $\boldsymbol{S}^{-1} = (\boldsymbol{S}_1^{-1}, \ldots, \boldsymbol{S}_k^{-1})$. Using Theorem 2.1, we can get

Theorem 3.1 *The following dominance results hold relative to the loss* (2.2)*.*

(1) *If* $\alpha(\alpha-1)g_1(t) + 2\alpha t g_1'(t) + t^2 g_1''(t) < 0$*, then* $\boldsymbol{\delta}^{EM}$ *dominates* $\boldsymbol{\delta}^{UB}$*.*

(2) *If* $\beta(\beta+2-p_0)g_2(u^2) - 2(2\beta-p_0)u^2 g_2'(u^2) + 4u^4 g_2''(u^2) < 0$ *for* $p_0 = \sum_{i=1}^{k} p_i(p_i+1)/2$*, then* $\boldsymbol{\delta}^{JS}$ *dominates* $\boldsymbol{\delta}^{UB}$*.*

(3) *If* $(\gamma^2-\gamma)g_3(v) + 2\gamma v g_3'(v) + v^2 g_3''(v) < 0$ *and* $\gamma g_3(v) + v g_3'(v) \geq 0$*, then* $\boldsymbol{\delta}^{US}$ *dominates* $\boldsymbol{\delta}^{UB}$*.*

For the g_j's satisfying the conditions of (1), (2) and (3) of Theorem 3.1, we can choose $g_j(x) = 1$, $g_j(x) = \log(1+x)$ and $g_j(x) = (1+x)^{-b}$, $b \geq 0$, for any j.

When we consider the special case of $k = 1$, the above functions have the simple forms $f_{EM}^*(\boldsymbol{S}) = (\operatorname{tr} \boldsymbol{S}_1)^{\alpha}$, $f_{JS}^*(\boldsymbol{S}) = (\operatorname{tr} \boldsymbol{S}_1^2)^{-\beta/2}$ and $f_{US}^*(\boldsymbol{S}) = |\boldsymbol{S}_1|^{\gamma}$, which result in the estimators

$$\begin{aligned}
\boldsymbol{\delta}_1^{EM} &= \boldsymbol{\delta}_1^{UB} - \frac{4\alpha}{\mathrm{tr}\,\boldsymbol{S}_1}\boldsymbol{I}_{p_1} \qquad \text{with } 0 < \alpha < 1, \\
\boldsymbol{\delta}_1^{JS} &= \boldsymbol{\delta}_1^{UB} + \frac{4\beta}{\mathrm{tr}\,\boldsymbol{S}_1^2}\boldsymbol{S}_1 \qquad \text{with } 0 < \beta < p_1(p_1+1)/2 - 2, \\
\boldsymbol{\delta}_1^{US} &= \boldsymbol{\delta}_1^{UB} - 4\gamma\boldsymbol{S}_1^{-1} = (n_1 - p_1 - 1 - 4\gamma)\boldsymbol{S}_1^{-1} \qquad \text{with } 0 < \gamma < 1.
\end{aligned}$$

The estimator $\boldsymbol{\delta}_1^{EM}$ is the similar type to that of Efron and Morris (1976) and $\boldsymbol{\delta}_1^{JS}$ is like the James and Stein (1961) estimator for means of normal distributions. The estimator $\boldsymbol{\delta}_1^{US}$ is probably a usual and natural estimator of $\boldsymbol{\Sigma}_1^{-1}$ because the form of $\boldsymbol{\delta}_1^{US}$ is a constant multiplier of $\boldsymbol{S}_1^{-1}$. See also Dey (1987) and Tsukuma and Konno (2006). It can be seen that $\boldsymbol{\delta}_1^{EM} < \boldsymbol{\delta}_1^{UB}$ and $\boldsymbol{\delta}_1^{US} < \boldsymbol{\delta}_1^{UB}$, namely, $\boldsymbol{\delta}_1^{UB} - \boldsymbol{\delta}_1^{EM}$ and $\boldsymbol{\delta}_1^{UB} - \boldsymbol{\delta}_1^{US}$ are positive definite matrices, respectively. Thus $\boldsymbol{\delta}_1^{EM}$ and $\boldsymbol{\delta}_1^{US}$ are called the shrinkage estimators. On the other hand, $\boldsymbol{\delta}_1^{JS} > \boldsymbol{\delta}_1^{UB}$ and hence $\boldsymbol{\delta}_1^{JS}$ is called the enlargement estimator.

In the special case of $p_1 = \cdots = p_k = 1$, the Wishart distribution degenerates the chi-squared distribution. Thus the model (1.1) is rewritten as $s_i \sim \sigma_i^2\chi_{n_i}^2$ for $i = 1, \ldots, k$ and the loss function becomes $\sum_{i=1}^k (\delta_i - \sigma_i^{-2})^2 = L(\boldsymbol{\delta}, \boldsymbol{\sigma}^{-2})$, say. Noting that $p_0 = k$ in (2) of Theorem 3.1, we can see that $\boldsymbol{\delta}^{JS}$ dominates $\boldsymbol{\delta}^{UB}$ relative to the loss $L(\boldsymbol{\delta}, \boldsymbol{\sigma}^{-2})$ if $0 < \beta < k - 2$ and $k \geq 3$, namely, it is necessary for k to be greater than or equal to three. The Stein phenomenon is also revealed in simultaneous estimation of the precisions (reciprocal of variances).

We next consider an estimator in the special case of $p_1 = \cdots = p_k = p$, say. Let $w = |\boldsymbol{S}_1 + \cdots + \boldsymbol{S}_k|$ and let $g_4(x)$ be a twice differentiable function. Let us define $f_{AM}(\boldsymbol{S}) = w^\varepsilon g_4(w)$, where ε is a constant, and consider the estimator

$$\boldsymbol{\delta}^{AM} = \boldsymbol{\delta}^{UB} - 4\mathcal{D}\log f_{AM}(\boldsymbol{S}).$$

Using Lemma 3.3 (3) given below, we can rewrite $\boldsymbol{\delta}^{AM} = (\boldsymbol{\delta}_1^{AM}, \ldots, \boldsymbol{\delta}_k^{AM})$ as

$$\boldsymbol{\delta}_i^{AM} = \boldsymbol{\delta}_i^{UB} - 4\Big(\varepsilon + \frac{wg_4'(w)}{g_4(w)}\Big)(\boldsymbol{S}_1 + \cdots + \boldsymbol{S}_k)^{-1}.$$

Thus, applying Theorem 2.1 to $f_{AM}(\boldsymbol{S})$, we obtain the following.

Theorem 3.2 *Let $p_1 = \cdots = p_k$. If $(\varepsilon^2 - \varepsilon)g_4(w) + 2\varepsilon wg_4'(w) + w^2 g_4''(w) < 0$ and $\varepsilon g_4(w) + wg_4'(w) \geq 0$, then $\boldsymbol{\delta}^{AM}$ dominates $\boldsymbol{\delta}^{UB}$ relative to the loss* (2.2).

It is noted that $\boldsymbol{\delta}_i^{AM} < \boldsymbol{\delta}_i^{UB}$ for $i = 1, \ldots, k$ since $\varepsilon + wg_4'(w)/g_4(w) \geq 0$. Thus $\boldsymbol{\delta}^{AM}$ is regarded as a shrinkage estimator.

3.2 Proofs

The following lemmas are useful for calculation with respect to the matrix differential operator $\mathcal{D}_i$.

Lemma 3.1 (Haff (1981)) *Let* $\boldsymbol{G}_1$ *and* $\boldsymbol{G}_2$ *be* $p_i \times p_i$ *symmetric matrices whose elements are functions of* $\boldsymbol{S}_i$. *Then* $\mathcal{D}_i(\boldsymbol{G}_1\boldsymbol{G}_2) = [\mathcal{D}_i\boldsymbol{G}_1]\boldsymbol{G}_2 + (\boldsymbol{G}_1\mathcal{D}_i)^t\boldsymbol{G}_2$, *where* $[\mathcal{D}_i\boldsymbol{G}_1]$ *means that* $\mathcal{D}_i$ *acts only on* $\boldsymbol{G}_1$.

Lemma 3.2 *Let* $\boldsymbol{C}$ *be a* $p_i \times p_i$ *symmetric matrix of constants. Then it holds that* (1) $\mathcal{D}_i \mathrm{tr}\, \boldsymbol{S}_i\boldsymbol{C} = \boldsymbol{C}$, (2) $\mathcal{D}_i \mathrm{tr}\, \boldsymbol{S}_i^2 = 2\boldsymbol{S}_i$, (3) $\mathcal{D}_i\boldsymbol{S}_i^{-1} = -(1/2)\{(\mathrm{tr}\, \boldsymbol{S}_i^{-1})\boldsymbol{S}_i^{-1} + \boldsymbol{S}_i^{-2}\}$ *and* (4) $\mathcal{D}_i|\boldsymbol{S}_i| = |\boldsymbol{S}_i|\boldsymbol{S}_i^{-1}$.

Proof. Since the equalities (1), (2) and (3) are due to Haff (1982), we shall prove (4). For convenience of notation, denote $\boldsymbol{S}_i = (s_{i\cdot ab})$ by $\boldsymbol{S} = (s_{ab})$ and p_i by p only in this proof. Note that $|\boldsymbol{S}| = \sum_{l=1}^p s_{al}\Delta_{al}$, where Δ_{al} is the cofactor of s_{al}. Here, Δ_{al} is equivalent to the determinant of the matrix obtained from $\boldsymbol{S}$ by, in the a-th row and the l-th column of $\boldsymbol{S}$, replacing the (a, l)-element with one and the others with zeros, namely,

$$\Delta_{al} = \begin{vmatrix} s_{11} & \cdots & s_{1,l-1} & 0 & s_{1,l+1} & \cdots & s_{1p} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{a-1,1} & \cdots & s_{a-1,l-1} & 0 & s_{a-1,l+1} & \cdots & s_{a-1,p} \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ s_{a+1,1} & \cdots & s_{a+1,l-1} & 0 & s_{a+1,l+1} & \cdots & s_{a+1,p} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_{p,l-1} & 0 & s_{p,l+1} & \cdots & s_{pp} \end{vmatrix} = |\boldsymbol{S}_{(al)}|, \quad \text{say.}$$

If $a = b$, then it obviously follows that

$$\frac{\partial}{\partial s_{aa}}|\boldsymbol{S}| = \frac{\partial}{\partial s_{aa}}\sum_{l=1}^p s_{al}\Delta_{al} = \Delta_{aa}.$$

When $a \neq b$, we can see that

$$\frac{\partial}{\partial s_{ab}}|\boldsymbol{S}| = \Delta_{ab} + \sum_{l\neq a} s_{al}\frac{\partial}{\partial s_{ab}}\Delta_{al}.$$

Note that the cofactor expansion of Δ_{al} along the b-th row is given by

$$\Delta_{al} = \sum_{m\neq l} s_{bm}\Delta_{al}(bm),$$

where $\Delta_{al}(bm)$ is the cofactor with respect to $\boldsymbol{S}_{(al)}$, namely, the determinant of matrix obtained from $\boldsymbol{S}_{(al)}$ by, in the b-th row and the m-th column of $\boldsymbol{S}$, replacing the (b, m)-element with one and the others with zeros. Thus we get

$$\frac{\partial}{\partial s_{ab}}\Delta_{al} = \Delta_{al}(ba).$$

Noting that $\boldsymbol{S}$ is symmetric and $\Delta_{al}(ba) = \Delta_{ab}(la)$, we have

$$\sum_{l\neq a} s_{al}\frac{\partial}{\partial s_{ab}}\Delta_{al} = \sum_{l\neq a} s_{la}\Delta_{ab}(la) = \Delta_{ab},$$

which implies that for $a \neq b$

$$\frac{\partial}{\partial s_{ab}}|\boldsymbol{S}| = 2\Delta_{ab}.$$

Hence $\mathcal{D}_i \boldsymbol{S} = (\Delta_{ab}) = |\boldsymbol{S}|\boldsymbol{S}^{-1}$ and we proved (4). ■

Proof of Theorem 3.1. We first prove (1). Note that from Lemma 3.2 (1),

$$\mathcal{D}_i\, t = \mathcal{D}_i \mathrm{tr}\, \boldsymbol{S}_i = \boldsymbol{I}_{p_i},$$

which gives that

$$\begin{aligned}
\mathcal{D} \circ \mathcal{D} f_{EM}(\boldsymbol{S}) &= \sum_{i=1}^{k} \mathrm{tr}\, \mathcal{D}_i \mathcal{D}_i t^{\alpha} g_1(t) \\
&= \sum_{i=1}^{k} \mathrm{tr}\, \mathcal{D}_i \{\alpha t^{\alpha-1} g_1(t) + t^{\alpha} g_1'(t)\} \\
&= \{\alpha(\alpha-1) g_1(t) + 2\alpha t g_1'(t) + t^2 g_1''(t)\} t^{\alpha-2} p_*,
\end{aligned}$$

where $p_* = \sum_{i=1}^{k} p_i$. Hence, from Theorem 2.1, $\boldsymbol{\delta}^{EM}$ dominates $\boldsymbol{\delta}^{UB}$ if the last right hand-side in the above equation is negative.

For the proof of (2), it is seen that from Lemma 3.2 (2),

$$\mathcal{D}_i\, u = \mathcal{D}_i (u^2)^{1/2} = (1/2)(u^2)^{-1/2} \mathcal{D}_i\, u^2 = (1/2) u^{-1} \mathcal{D}_i \mathrm{tr}\, \boldsymbol{S}_i^2 = u^{-1} \boldsymbol{S}_i.$$

Noting that $\mathcal{D}_i \boldsymbol{S}_i = (p_i + 1)\boldsymbol{I}_{p_i}/2$, we get

$$\begin{aligned}
\mathcal{D} \circ \mathcal{D} f_{JS}(\boldsymbol{S}) &= \sum_{i=1}^{k} \mathrm{tr}\, \mathcal{D}_i \mathcal{D}_i u^{-\beta} g_2(u^2) \\
&= \sum_{i=1}^{k} \mathrm{tr}\, \mathcal{D}_i \{-\beta u^{-\beta-2} g_2(u^2) + 2u^{-\beta} g_2'(u^2)\} \boldsymbol{S}_i \\
&= \sum_{i=1}^{k} \mathrm{tr}\, \{\beta(\beta+2) u^{-\beta-4} g_2(u^2) \boldsymbol{S}_i^2 - 4\beta u^{-\beta-2} g_2'(u^2) \boldsymbol{S}_i^2 + 4u^{-\beta} g_2''(u^2) \boldsymbol{S}_i^2 \\
&\qquad - \beta(p_i+1) u^{-\beta-2} g_2(u^2) \boldsymbol{I}_{p_i}/2 + (p_i+1) u^{-\beta} g_2'(u^2) \boldsymbol{I}_{p_i}\} \\
&= u^{-\beta-2} \{\beta(\beta+2-p_0) g_2(u^2) - 2(2\beta - p_0) u^2 g_2'(u^2) + 4u^4 g_2''(u^2)\},
\end{aligned}$$

where $p_0 = \sum_{i=1}^{k} p_i(p_i+1)/2$. The proof of (2) is completed.

Finally we give the proof of (3). Note that from Lemma 3.2 (4),

$$\mathcal{D}_i\, v^{\gamma} = \Big(\prod_{j \neq i} |\boldsymbol{S}_j|^{\gamma}\Big) \mathcal{D}_i |\boldsymbol{S}_i|^{\gamma} = \gamma |\boldsymbol{S}_i|^{\gamma-1} \Big(\prod_{j \neq i} |\boldsymbol{S}_j|^{\gamma}\Big) \mathcal{D}_i |\boldsymbol{S}_i| = \gamma v^{\gamma} \boldsymbol{S}_i^{-1}.$$

We also use Lemma 3.2 (3) to give

$$\begin{aligned}
\mathcal{D}\circ\mathcal{D}f_{US}(\boldsymbol{S}) &= \sum_{i=1}^k \operatorname{tr}\mathcal{D}_i\mathcal{D}_i v^\gamma g_3(v)\\
&= \sum_{i=1}^k \operatorname{tr}\mathcal{D}_i\{\gamma v^\gamma g_3(v) + v^{\gamma+1}g_3'(v)\}\boldsymbol{S}_i^{-1}\\
&= \sum_{i=1}^k \operatorname{tr}\{\gamma^2 v^\gamma g_3(v)\boldsymbol{S}_i^{-2} + (2\gamma+1)v^{\gamma+1}g_3'(v)\boldsymbol{S}_i^{-2} + v^{\gamma+2}g_3''(v)\boldsymbol{S}_i^{-2}\\
&\qquad - (1/2)(\gamma v^\gamma g_3(v) + v^{\gamma+1}g_3'(v))((\operatorname{tr}\boldsymbol{S}_i^{-1})\boldsymbol{S}_i^{-1} + \boldsymbol{S}_i^{-2})\}.
\end{aligned}$$

The fact that $(\operatorname{tr}\boldsymbol{S}_i^{-1})^2 \geq \operatorname{tr}\boldsymbol{S}_i^{-2}$ implies that if $\gamma g_3(v) + v g_3'(v) \geq 0$, then

$$\mathcal{D}\circ\mathcal{D}f_{US}(\boldsymbol{S}) \leq v^\gamma\{(\gamma^2-\gamma)g_3(v) + 2\gamma v g_3'(v) + v^2 g_3''(v)\}\sum_{i=1}^k \operatorname{tr}\boldsymbol{S}_i^{-2},$$

which proves (3). ■

To derive the dominance result with respect to $\boldsymbol{\delta}^{AM}$, we use the following lemma.

Lemma 3.3 *Let $p = p_1 = \cdots = p_k$ and $\boldsymbol{T} = \boldsymbol{S}_1 + \cdots + \boldsymbol{S}_k$ and let $\boldsymbol{C}$ be a $p \times p$ symmetric matrix of constants. Then for $i = 1, \ldots, k$, we have*

(1) $(\boldsymbol{C}\mathcal{D}_i)^t\boldsymbol{S}_i = (1/2)\{\operatorname{tr}\boldsymbol{C}\}\boldsymbol{I}_p + (1/2)\boldsymbol{C}$,

(2) $\mathcal{D}_i\boldsymbol{T}^{-1} = -(1/2)\{\operatorname{tr}\boldsymbol{T}^{-1}\}\boldsymbol{T}^{-1} - (1/2)\boldsymbol{T}^{-2}$,

(3) $\mathcal{D}_i|\boldsymbol{T}| = |\boldsymbol{T}|\boldsymbol{T}^{-1}$.

Proof. The expression (1) is due to Haff (1981). From (1), it is observed that

$$\begin{aligned}
\boldsymbol{0}_{p\times p} &= \mathcal{D}_i(\boldsymbol{T}^{-1}\boldsymbol{T}) = [\mathcal{D}_i\boldsymbol{T}^{-1}]\boldsymbol{T} + (\boldsymbol{T}^{-1}\mathcal{D}_i)^t\boldsymbol{T}\\
&= [\mathcal{D}_i\boldsymbol{T}^{-1}]\boldsymbol{T} + (\boldsymbol{T}^{-1}\mathcal{D}_i)^t\boldsymbol{S}_i\\
&= [\mathcal{D}_i\boldsymbol{T}^{-1}]\boldsymbol{T} + (1/2)\{\operatorname{tr}\boldsymbol{T}^{-1}\}\boldsymbol{I}_p + (1/2)\boldsymbol{T}^{-1},
\end{aligned}$$

giving (2). The expression (3) can be verified by the same argument as in the proof of Lemma 3.2 (3). ■

Proof of Theorem 3.2. Using Lemma 3.3 (2) and (3), we can prove Theorem 3.2 based on the same argument as in Theorem 3.1 (3). ■

4 Generalized Bayes estimation

In this section we consider the Bayes procedures for estimation of the precision matrices and establish a condition of prior distributions such that the resulting Bayes estimator dominates the unbiased estimator relative to the loss (2.2).

Let $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_k) = (\boldsymbol{\Sigma}_1^{-1}, \ldots, \boldsymbol{\Sigma}_k^{-1})$. Denote by $\pi(\boldsymbol{\Lambda}) \prod_{i=1}^k |\boldsymbol{\Lambda}_i|^{-p_i-1}$ a prior density of $\boldsymbol{\Lambda}$. The resulting generalized Bayes estimator is expressed as $\boldsymbol{\delta}^{GB} = (\boldsymbol{\delta}_1^{GB}, \ldots, \boldsymbol{\delta}_k^{GB})$ with

$$\boldsymbol{\delta}_i^{GB} = \frac{\int \boldsymbol{\Lambda}_i \{\prod_{i=1}^k |\boldsymbol{\Lambda}_i|^{n_i/2-p_i-1} \exp(-\mathrm{tr}\, \boldsymbol{\Lambda}_i \boldsymbol{S}_i/2)\} \pi(\boldsymbol{\Lambda}) d\boldsymbol{\Lambda}}{\int \{\prod_{i=1}^k |\boldsymbol{\Lambda}_i|^{n_i/2-p_i-1} \exp(-\mathrm{tr}\, \boldsymbol{\Lambda}_i \boldsymbol{S}_i/2)\} \pi(\boldsymbol{\Lambda}) d\boldsymbol{\Lambda}}.$$

It is noted from Lemma 3.2 that $\mathcal{D}_i \log |\boldsymbol{S}_i| = \boldsymbol{S}_i^{-1}$ and $\mathcal{D}_i \mathrm{tr}\, \boldsymbol{\Lambda}_i \boldsymbol{S}_i = \boldsymbol{\Lambda}_i$. The generalized Bayes estimator can be rewritten as

$$\begin{aligned} \boldsymbol{\delta}^{GB} &= \boldsymbol{\delta}^{UB} + (\boldsymbol{\delta}^{GB} - \boldsymbol{\delta}^{UB}) \\ &= \boldsymbol{\delta}^{UB} - 4\mathcal{D} \log f_\pi(\boldsymbol{S}), \end{aligned} \tag{4.1}$$

where

$$f_\pi(\boldsymbol{S}) = \Big[\Big(\prod_{i=1}^k |\boldsymbol{S}_i|^{(n_i-p_i-1)/2} \Big) \int \Big\{ \prod_{i=1}^k |\boldsymbol{\Lambda}_i|^{n_i/2-p_i-1} \exp(-\mathrm{tr}\, \boldsymbol{\Lambda}_i \boldsymbol{S}_i/2) \Big\} \pi(\boldsymbol{\Lambda}) d\boldsymbol{\Lambda} \Big]^{1/2}.$$

For $i = 1, \ldots, k$, let $\boldsymbol{\Xi}_i = \boldsymbol{S}_i^{1/2} \boldsymbol{\Lambda}_i \boldsymbol{S}_i^{1/2}$ where $\boldsymbol{S}_i^{1/2}$ is a symmetric half matrix of $\boldsymbol{S}_i$, namely, $\boldsymbol{S}_i = \boldsymbol{S}_i^{1/2} \boldsymbol{S}_i^{1/2}$. Since the Jacobian of the transformation from $\boldsymbol{\Lambda}_i$ to $\boldsymbol{\Xi}_i$ is given by $|\boldsymbol{S}_i|^{-(p_i+1)/2}$, it follows that

$$f_\pi(\boldsymbol{S}) = \Big[\int \Big\{ \prod_{i=1}^k |\boldsymbol{\Xi}_i|^{n_i/2-p_i-1} \exp(-\mathrm{tr}\, \boldsymbol{\Xi}_i/2) \Big\} \pi(\boldsymbol{S}^{-1/2} \boldsymbol{\Xi} \boldsymbol{S}^{-1/2}) d\boldsymbol{\Xi} \Big]^{1/2},$$

where $\boldsymbol{S}^{-1/2} \boldsymbol{\Xi} \boldsymbol{S}^{-1/2}$ means $(\boldsymbol{S}_1^{-1/2} \boldsymbol{\Xi}_1 \boldsymbol{S}_1^{-1/2}, \ldots, \boldsymbol{S}_k^{-1/2} \boldsymbol{\Xi}_k \boldsymbol{S}_k^{-1/2})$. Hence, the following dominance property can be established by applying Theorem 2.1.

Theorem 4.1 *If $\mathcal{D} \circ \mathcal{D} f_\pi(\boldsymbol{S})$ is negative, then the generalized Bayes estimator $\boldsymbol{\delta}^{GB}$ dominates $\boldsymbol{\delta}^{UB}$ relative to the loss* (2.2).

It can easily be observed that

$$\begin{aligned} \mathcal{D} \circ \mathcal{D} f_\pi(\boldsymbol{S}) = &\frac{1}{2} \frac{\int \{\prod_{i=1}^k |\boldsymbol{\Xi}_i|^{n_i/2-p_i-1} \exp(-\mathrm{tr}\, \boldsymbol{\Xi}_i/2)\} [\mathcal{D} \circ \mathcal{D} \pi(\boldsymbol{S}^{-1/2} \boldsymbol{\Xi} \boldsymbol{S}^{-1/2})] d\boldsymbol{\Xi}}{\{\int \{\prod_{i=1}^k |\boldsymbol{\Xi}_i|^{n_i/2-p_i-1} \exp(-\mathrm{tr}\, \boldsymbol{\Xi}_i/2)\} \pi(\boldsymbol{S}^{-1/2} \boldsymbol{\Xi} \boldsymbol{S}^{-1/2}) d\boldsymbol{\Xi}\}^{1/2}} \\ &- \frac{1}{4} \frac{\|\mathcal{D} \int \{\prod_{i=1}^k |\boldsymbol{\Xi}_i|^{n_i/2-p_i-1} \exp(-\mathrm{tr}\, \boldsymbol{\Xi}_i/2)\} \pi(\boldsymbol{S}^{-1/2} \boldsymbol{\Xi} \boldsymbol{S}^{-1/2}) d\boldsymbol{\Xi}\|_M^2}{\{\int \{\prod_{i=1}^k |\boldsymbol{\Xi}_i|^{n_i/2-p_i-1} \exp(-\mathrm{tr}\, \boldsymbol{\Xi}_i/2)\} \pi(\boldsymbol{S}^{-1/2} \boldsymbol{\Xi} \boldsymbol{S}^{-1/2}) d\boldsymbol{\Xi}\}^{3/2}}. \end{aligned}$$

Since the second term in the right hand-side of the above expression is nonpositive, we get a sufficient condition on the prior $\pi(\cdot)$.

Corollary 4.1 *If $\mathcal{D} \circ \mathcal{D} \pi(\boldsymbol{S}^{-1/2} \boldsymbol{\Xi} \boldsymbol{S}^{-1/2})$ is negative, then the generalized Bayes estimator $\boldsymbol{\delta}^{GB}$ dominates $\boldsymbol{\delta}^{UB}$ relative to the loss* (2.2).

It is noted that the condition $\mathcal{D} \circ \mathcal{D} \pi(\boldsymbol{S}^{-1/2} \boldsymbol{\Xi} \boldsymbol{S}^{-1/2}) < 0$ provides a characterization of the prior distribution $\pi(\cdot)$ for the resulting generalized Bayes estimator to dominate

$\boldsymbol{\delta}^{UB}$. However, we may obtain a better condition by evaluating $\mathcal{D}\circ\mathcal{D}f_\pi(\boldsymbol{S})$ than $\mathcal{D}\circ\mathcal{D}\pi(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2})$. For instance, let us consider the prior distribution

$$\pi_B(\boldsymbol{\Lambda}) = \left(\prod_{i=1}^k |\boldsymbol{\Lambda}_i^{-1}|\right)^a \left(\sum_{i=1}^k \mathrm{tr}\,\boldsymbol{\Lambda}_i^{-1}\right)^b, \tag{4.2}$$

namely,

$$\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2}) = (|\boldsymbol{S}_1\boldsymbol{\Xi}_1^{-1}|\cdots|\boldsymbol{S}_k\boldsymbol{\Xi}_k^{-1}|)^a(\mathrm{tr}\,\boldsymbol{S}_1\boldsymbol{\Xi}_1^{-1}+\cdots+\mathrm{tr}\,\boldsymbol{S}_k\boldsymbol{\Xi}_k^{-1})^b,$$

where a and b are positive constants. Let $V=|\boldsymbol{S}_1\boldsymbol{\Xi}_1^{-1}|\cdots|\boldsymbol{S}_k\boldsymbol{\Xi}_k^{-1}|$ and $T=\mathrm{tr}\,\boldsymbol{S}_1\boldsymbol{\Xi}_1^{-1}+\cdots+\mathrm{tr}\,\boldsymbol{S}_k\boldsymbol{\Xi}_k^{-1}$. It is observed from Lemma 3.2 that $\mathcal{D}_iV^a=aV^a\boldsymbol{S}_i^{-1}$ and $\mathcal{D}_iT^b=bT^{b-1}\boldsymbol{\Xi}_i^{-1}$, which yield that

$$\begin{aligned}\mathcal{D}_i\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2}) &= aV^aT^b\boldsymbol{S}_i^{-1}+bV^aT^{b-1}\boldsymbol{\Xi}_i^{-1}\\ &= (a\boldsymbol{S}_i^{-1}+bT^{-1}\boldsymbol{\Xi}_i^{-1})\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2}) \end{aligned}\tag{4.3}$$

and

$$\begin{aligned}&\mathrm{tr}\,\mathcal{D}_i\mathcal{D}_i\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2})\\ &= a^2V^aT^b\mathrm{tr}\,\boldsymbol{S}_i^{-2}+2abV^aT^{b-1}\mathrm{tr}\,\boldsymbol{S}_i^{-1}\boldsymbol{\Xi}_i^{-1}+b(b-1)V^aT^{b-2}\mathrm{tr}\,\boldsymbol{\Xi}_i^{-2}\\ &\qquad -\frac{a}{2}V^aT^b(\{\mathrm{tr}\,\boldsymbol{S}_i^{-1}\}^2+\mathrm{tr}\,\boldsymbol{S}_i^{-2})\\ &\le (a^2-a)V^aT^b\mathrm{tr}\,\boldsymbol{S}_i^{-2}+2abV^aT^{b-1}\mathrm{tr}\,\boldsymbol{S}_i^{-1}\boldsymbol{\Xi}_i^{-1}+b(b-1)V^aT^{b-2}\mathrm{tr}\,\boldsymbol{\Xi}_i^{-2}\\ &= \{(a^2-a)\mathrm{tr}\,\boldsymbol{S}_i^{-2}+2abT^{-1}\mathrm{tr}\,\boldsymbol{S}_i^{-1}\boldsymbol{\Xi}_i^{-1}+b(b-1)T^{-2}\mathrm{tr}\,\boldsymbol{\Xi}_i^{-2}\}\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2}). \end{aligned}\tag{4.4}$$

Letting

$$g_B(\boldsymbol{S}) = \int\Big\{\prod_{i=1}^k|\boldsymbol{\Xi}_i|^{n_i/2-p_i-1}\exp(-\mathrm{tr}\,\boldsymbol{\Xi}_i/2)\Big\}\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2})d\boldsymbol{\Xi},$$

we can write $\mathcal{D}\circ\mathcal{D}f_{\pi_B}(\boldsymbol{S})$ as

$$\mathcal{D}\circ\mathcal{D}f_{\pi_B}(\boldsymbol{S}) = \frac{1}{2}\frac{\mathcal{D}\circ\mathcal{D}g_B(\boldsymbol{S})}{\{g_B(\boldsymbol{S})\}^{1/2}}-\frac{1}{4}\frac{\|\mathcal{D}g_B(\boldsymbol{S})\|_M^2}{\{g_B(\boldsymbol{S})\}^{3/2}}. \tag{4.5}$$

From (4.3), it is seen that

$$\begin{aligned}&\|\mathcal{D}g_B(\boldsymbol{S})\|_M^2\\ &= \sum_{i=1}^k\mathrm{tr}\Big[\int\Big\{\prod_{i=1}^k|\boldsymbol{\Xi}_i|^{n_i/2-p_i-1}\exp(-\mathrm{tr}\,\boldsymbol{\Xi}_i/2)\Big\}\mathcal{D}_i\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2})d\boldsymbol{\Xi}\Big]^2\\ &= g_B(\boldsymbol{S})\int\Big\{\prod_{i=1}^k|\boldsymbol{\Xi}_i|^{n_i/2-p_i-1}\exp(-\mathrm{tr}\,\boldsymbol{\Xi}_i/2)\Big\}\sum_{i=1}^k\{a^2\mathrm{tr}\,\boldsymbol{S}_i^{-2}+2abT^{-1}\mathrm{tr}\,\boldsymbol{S}_i^{-1}\boldsymbol{\Xi}_i^{-1}\}\\ &\qquad\times\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2})d\boldsymbol{\Xi}\\ &\quad+\sum_{i=1}^k b^2\mathrm{tr}\Big[\int T^{-1}\boldsymbol{\Xi}_i^{-1}\Big\{\prod_{i=1}^k|\boldsymbol{\Xi}_i|^{n_i/2-p_i-1}\exp(-\mathrm{tr}\,\boldsymbol{\Xi}_i/2)\Big\}\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2})d\boldsymbol{\Xi}\Big]^2,\end{aligned}$$

which is evaluated as

$$
\begin{aligned}
&\|\mathcal{D}g_B(\boldsymbol{S})\|_M^2 \\
&\ge g_B(\boldsymbol{S})\int\Big\{\prod_{i=1}^k |\boldsymbol{\Xi}_i|^{n_i/2-p_i-1}\exp(-\mathrm{tr}\,\boldsymbol{\Xi}_i/2)\Big\}\sum_{i=1}^k\{a^2\mathrm{tr}\,\boldsymbol{S}_i^{-2}+2abT^{-1}\mathrm{tr}\,\boldsymbol{S}_i^{-1}\boldsymbol{\Xi}_i^{-1}\} \\
&\qquad\times\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2})d\boldsymbol{\Xi}. \qquad (4.6)
\end{aligned}
$$

Also we use (4.4) to evaluate $\mathcal{D}\circ\mathcal{D}g_B(\boldsymbol{S})$ as

$$
\begin{aligned}
&\mathcal{D}\circ\mathcal{D}g_B(\boldsymbol{S}) \\
&=\int\Big\{\prod_{i=1}^k |\boldsymbol{\Xi}_i|^{n_i/2-p_i-1}\exp(-\mathrm{tr}\,\boldsymbol{\Xi}_i/2)\Big\}\sum_{i=1}^k\mathrm{tr}\,\mathcal{D}_i\mathcal{D}_i\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2})d\boldsymbol{\Xi} \\
&\le\int G_0(\boldsymbol{S},\boldsymbol{\Xi})\Big\{\prod_{i=1}^k |\boldsymbol{\Xi}_i|^{n_i/2-p_i-1}\exp(-\mathrm{tr}\,\boldsymbol{\Xi}_i/2)\Big\}\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2})d\boldsymbol{\Xi}, \qquad (4.7)
\end{aligned}
$$

where

$$
G_0(\boldsymbol{S},\boldsymbol{\Xi})=(a^2-a)\sum_{i=1}^k\mathrm{tr}\,\boldsymbol{S}_i^{-2}+2abT^{-1}\sum_{i=1}^k\mathrm{tr}\,\boldsymbol{S}_i^{-1}\boldsymbol{\Xi}_i^{-1}+b(b-1)T^{-2}\sum_{i=1}^k\mathrm{tr}\,\boldsymbol{\Xi}_i^{-2}.
$$

Combining (4.6) and (4.7) gives that

$$
\begin{aligned}
&\mathcal{D}\circ\mathcal{D}f_{\pi_B}(\boldsymbol{S}) \\
&=\frac{1}{2}\frac{\mathcal{D}\circ\mathcal{D}g_B(\boldsymbol{S})}{\{g_B(\boldsymbol{S})\}^{1/2}}-\frac{1}{4}\frac{\|\mathcal{D}g_B(\boldsymbol{S})\|_M^2}{\{g_B(\boldsymbol{S})\}^{3/2}} \\
&\le\frac{1}{4}\{g_B(\boldsymbol{S})\}^{-1/2}\int G(\boldsymbol{S},\boldsymbol{\Xi})\Big\{\prod_{i=1}^k |\boldsymbol{\Xi}_i|^{n_i/2-p_i-1}\exp(-\mathrm{tr}\,\boldsymbol{\Xi}_i/2)\Big\}\pi_B(\boldsymbol{S}^{-1/2}\boldsymbol{\Xi}\boldsymbol{S}^{-1/2})d\boldsymbol{\Xi},
\end{aligned}
$$

where

$$
G(\boldsymbol{S},\boldsymbol{\Xi})=(a^2-2a)\sum_{i=1}^k\mathrm{tr}\,\boldsymbol{S}_i^{-2}+2abT^{-1}\sum_{i=1}^k\mathrm{tr}\,\boldsymbol{S}_i^{-1}\boldsymbol{\Xi}_i^{-1}+2b(b-1)T^{-2}\sum_{i=1}^k\mathrm{tr}\,\boldsymbol{\Xi}_i^{-2}.
$$

Hence, if $G(\boldsymbol{S},\boldsymbol{\Xi})$ is negative, then $\mathcal{D}\circ\mathcal{D}f_{\pi_B}(\boldsymbol{S})$ is negative. Furthermore, applying the fact that $T^{-1}\mathrm{tr}\,\boldsymbol{S}_i^{-1}\boldsymbol{\Xi}_i^{-1}<\mathrm{tr}\,\boldsymbol{S}_i^{-2}$ to $G(\boldsymbol{S},\boldsymbol{\Xi})$, we get

$$
G(\boldsymbol{S},\boldsymbol{\Xi})<(a^2-2a+2ab)\sum_{i=1}^k\mathrm{tr}\,\boldsymbol{S}_i^{-2}+2b(b-1)T^{-2}\sum_{i=1}^k\mathrm{tr}\,\boldsymbol{\Xi}_i^{-2}.
$$

Then we get the following theorem.

Theorem 4.2 *The generalized Bayes estimator $\boldsymbol{\delta}^{GB}$ with the prior π_B dominates $\boldsymbol{\delta}^{UB}$ relative to the loss (2.2) if $a^2-2a+2ab\le 0$ and $b(b-1)\le 0$.*

It is noted that $\boldsymbol{\delta}^{GB}$ with the prior π_B is regarded as a shrinkage estimator since each component of $-\mathcal{D}\log f_{\pi_B}(\boldsymbol{S})$ is negative definite.

5 Empirical Bayes method

In this section we consider the empirical Bayes method for estimation of the precision matrices. The findings of this section is that $\boldsymbol{\delta}^{EM}$ and $\boldsymbol{\delta}^{AM}$ given in Section 3 are characterized as empirical Bayes estimators.

Letting $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_k)$ with $\boldsymbol{\Lambda}_i = \boldsymbol{\Sigma}_i^{-1}$ for $i = 1, \ldots, k$, we can write the likelihood of $\boldsymbol{S} = (\boldsymbol{S}_1, \ldots, \boldsymbol{S}_k)$ as

$$p(\boldsymbol{S}|\boldsymbol{\Lambda}) \propto \prod_{i=1}^{k} |\boldsymbol{\Lambda}_i|^{n_i/2} |\boldsymbol{S}_i|^{(n_i-p_i-1)/2} e^{-\mathrm{tr}\, \boldsymbol{\Lambda}_i \boldsymbol{S}_i/2}.$$

For $i = 1, \ldots, k$, suppose that $\boldsymbol{\Lambda}_i = \boldsymbol{\Xi}_i + \gamma \boldsymbol{I}_{p_i}$ and that the $\boldsymbol{\Xi}_i$'s are independently distributed as the matrix-variate t-distributions having the joint density

$$p(\boldsymbol{\Xi}|\gamma) \propto \prod_{i=1}^{k} |\boldsymbol{\Xi}_i|^{m_i/2} |\boldsymbol{I}_{p_i} + \boldsymbol{\Xi}_i/\gamma|^{-n_i/2},$$

where $n_i > m_i + 2p_i$. Then the posterior density of $\boldsymbol{\Xi}$ and the marginal density of $\boldsymbol{S}$ are given by

$$p(\boldsymbol{\Xi}|\boldsymbol{S}) \propto \prod_{i=1}^{k} |\boldsymbol{\Xi}_i|^{m_i/2} e^{-\mathrm{tr}\, \boldsymbol{\Xi}_i \boldsymbol{S}_i/2},$$
$$p(\boldsymbol{S}|\gamma) \propto \prod_{i=1}^{k} \gamma^{(n_i-m_i-p_i-1)p_i/2} |\boldsymbol{S}_i|^{(n_i-m_i-2p_i-2)/2} e^{-\gamma \mathrm{tr}\, \boldsymbol{S}_i/2},$$

both of which are Wishart distributions. Hence, the posterior mean of $\boldsymbol{\Lambda}_i$ is

$$E[\boldsymbol{\Lambda}_i|\gamma] = E[\boldsymbol{\Xi}_i + \gamma \boldsymbol{I}_{p_i}|\gamma] = (m_i + p_i + 1)\boldsymbol{S}_i^{-1} + \gamma \boldsymbol{I}_{p_i}.$$

From the marginal density of $\boldsymbol{S}$, the maximum likelihood estimator of γ is given by

$$\hat{\gamma} = \frac{c}{\mathrm{tr}\, \boldsymbol{S}_1 + \cdots + \mathrm{tr}\, \boldsymbol{S}_k},$$

where c is a constant. Thus the resulting empirical Bayes estimator of $\boldsymbol{\Lambda}_i = \boldsymbol{\Sigma}_i^{-1}$ is

$$\boldsymbol{\delta}_i^{EB} = (m_i + p_i + 1)\boldsymbol{S}_i^{-1} + \frac{c}{\mathrm{tr}\, \boldsymbol{S}_1 + \cdots + \mathrm{tr}\, \boldsymbol{S}_k} \boldsymbol{I}_{p_i}$$

for $i = 1, \ldots, k$. This estimator is the same type as $\boldsymbol{\delta}^{EM}$ considered in Section 3.

Next, we consider an empirical Bayes estimator in the case of $p = p_1 = \cdots = p_k$. Suppose that $\boldsymbol{\Lambda}_i = \boldsymbol{\Xi}_i + \boldsymbol{\Gamma}$ for $i = 1, \ldots, k$ and that

$$p(\boldsymbol{\Xi}|\boldsymbol{\Gamma}) \propto \prod_{i=1}^{k} |\boldsymbol{\Xi}_i|^{m_i/2} |\boldsymbol{I}_p + \boldsymbol{\Xi}_i \boldsymbol{\Gamma}^{-1}|^{-n_i/2}.$$

Then the posterior density of $\boldsymbol{\Xi}$ and the marginal density of $\boldsymbol{S}$ are given by, respectively,

$$p(\boldsymbol{\Xi}|\boldsymbol{S}) \propto \prod_{i=1}^{k} |\boldsymbol{\Xi}_i|^{m_i/2} e^{-\mathrm{tr}\,\boldsymbol{\Xi}_i \boldsymbol{S}_i/2},$$
$$p(\boldsymbol{S}|\boldsymbol{\Gamma}) \propto \prod_{i=1}^{k} |\boldsymbol{\Gamma}|^{(n_i-m_i-p-1)/2} |\boldsymbol{S}_i|^{(n_i-m_i-2p-2)/2} e^{-\mathrm{tr}\,\boldsymbol{\Gamma}\boldsymbol{S}_i/2}.$$

Therefore the posterior mean of $\boldsymbol{\Lambda}_i$ is $E[\boldsymbol{\Lambda}_i|\gamma] = (m_i+p+1)\boldsymbol{S}_i^{-1} + \boldsymbol{\Gamma}$. From the marginal density of $\boldsymbol{S}$, the maximum likelihood estimator of $\boldsymbol{\Gamma}$ is given by $\widehat{\boldsymbol{\Gamma}} = c(\boldsymbol{S}_1 + \cdots + \boldsymbol{S}_k)^{-1}$, where c is a constant. Thus the resulting empirical Bayes estimator of $\boldsymbol{\Sigma}_i^{-1}$ is

$$\boldsymbol{\delta}_i^{EB*} = (m_i + p + 1)\boldsymbol{S}_i^{-1} + c(\boldsymbol{S}_1 + \cdots + \boldsymbol{S}_k)^{-1}.$$

Therefore $\boldsymbol{\delta}_i^{EB*}$ with $m_i = n_i - 2p - 2$ is equivalent to $\boldsymbol{\delta}^{AM}$ given in Section 3.

The dominance results of $\boldsymbol{\delta}^{EB}$ and $\boldsymbol{\delta}^{EB*}$ over the unbiased estimator $\boldsymbol{\delta}^{UB}$ are given as in Section 3.

6 Further dominance results

6.1 Unified improvement upon both shrinkage and enlargement estimators

As seen in Section 3, the estimators $\boldsymbol{\delta}^{EM}$, $\boldsymbol{\delta}^{AM}$ and $\boldsymbol{\delta}^{US}$ are shrinkage type estimators for the unbiased estimator $\boldsymbol{\delta}^{UB}$, while $\boldsymbol{\delta}^{JS}$ is an enlargement type estimator. Furthermore, from result of Section 4, we can see that the generalized Bayes estimator $\boldsymbol{\delta}^{GB}$ given by (4.1) is a shrinkage type estimator if $\mathcal{D}_i f_\pi(\boldsymbol{S})$ is positive definite. This section concerns unified improvement methods on both shrinkage and enlargement estimators.

First, define a shrinkage estimator as $\boldsymbol{\delta}^{SH} = \boldsymbol{\delta}^{UB} - 4\mathcal{D}\log f_{SH}(\boldsymbol{S})$ and an enlargement estimator as $\boldsymbol{\delta}^{EN} = \boldsymbol{\delta}^{UB} - 4\mathcal{D}\log f_{EN}(\boldsymbol{S})$, where $f_{SH}(\boldsymbol{S})$ and $f_{EN}(\boldsymbol{S})$ are positive and scalar-valued functions of $\boldsymbol{S}$. We note that each component of $\mathcal{D}f_{SH}(\boldsymbol{S})$ is positive definite, namely, $\mathcal{D}_i f_{SH}(\boldsymbol{S})$ is positive definite for $i = 1, \ldots, k$, and that $\boldsymbol{\delta}^{SH}$ is the shrinkage type estimator for the unbiased estimator $\boldsymbol{\delta}^{UB}$ since $\boldsymbol{\delta}^{SH} = \boldsymbol{\delta}^{UB} - (4/f_{SH}(\boldsymbol{S}))\mathcal{D}f_{SH}(\boldsymbol{S})$. Similarly we see that each component of $\mathcal{D}f_{EN}(\boldsymbol{S})$ is negative definite.

Let us consider an improved estimator of the form

$$\boldsymbol{\delta}^{I} = \boldsymbol{\delta}^{UB} - 4\mathcal{D}\log(f_{SH}(\boldsymbol{S})f_{EN}(\boldsymbol{S})),$$

which can be expressed as

$$\begin{aligned}\boldsymbol{\delta}^{I} &= \boldsymbol{\delta}^{SH} - 4\mathcal{D}\log f_{EN}(\boldsymbol{S}) \\ &= \boldsymbol{\delta}^{EN} - 4\mathcal{D}\log f_{SH}(\boldsymbol{S}).\end{aligned}$$

Then we get the interesting result.

Theorem 6.1 *Assume that* $\mathcal{D}\circ\mathcal{D}f_{SH}(\boldsymbol{S})<0$ *and* $\mathcal{D}\circ\mathcal{D}f_{EN}(\boldsymbol{S})<0$. *Then* $\boldsymbol{\delta}^I$ *dominates both* $\boldsymbol{\delta}^{SH}$ *and* $\boldsymbol{\delta}^{EN}$ *relative to the loss* (2.2).

Proof. By applying Theorem 2.1 to the estimators $\boldsymbol{\delta}^I$ and $\boldsymbol{\delta}^{SH}$, the difference between the risk functions of $\boldsymbol{\delta}^I$ and $\boldsymbol{\delta}^{SH}$ can be expressed as

$$\begin{aligned}R(\boldsymbol{\delta}^I,\boldsymbol{\Sigma}^{-1})-R(\boldsymbol{\delta}^{SH},\boldsymbol{\Sigma}^{-1})&=16E\Big[\frac{\mathcal{D}\circ\mathcal{D}(f_{SH}(\boldsymbol{S})f_{EN}(\boldsymbol{S}))}{f_{SH}(\boldsymbol{S})f_{EN}(\boldsymbol{S})}\Big]-16E\Big[\frac{\mathcal{D}\circ\mathcal{D}f_{SH}(\boldsymbol{S})}{f_{SH}(\boldsymbol{S})}\Big]\\&=32E\Big[\frac{[\mathcal{D}f_{SH}(\boldsymbol{S})]\circ[\mathcal{D}f_{EN}(\boldsymbol{S})]}{f_{SH}(\boldsymbol{S})f_{EN}(\boldsymbol{S})}\Big]+16E\Big[\frac{\mathcal{D}\circ\mathcal{D}f_{EN}(\boldsymbol{S})}{f_{EN}(\boldsymbol{S})}\Big].\end{aligned}\tag{6.1}$$

It is observed that

$$[\mathcal{D}f_{SH}(\boldsymbol{S})]\circ[\mathcal{D}f_{EN}(\boldsymbol{S})]=\sum_{i=1}^k\mathrm{tr}\,[\mathcal{D}_if_{SH}(\boldsymbol{S})][\mathcal{D}_if_{EN}(\boldsymbol{S})].$$

Hence, the first expectation in the last right hand-side of (6.1) is negative since $\mathcal{D}_if_{SH}(\boldsymbol{S})$ is positive definite and $\mathcal{D}_if_{EN}(\boldsymbol{S})$ is negative definite for $i=1,\ldots,k$. Therefore the assumption on $f_{EN}(\boldsymbol{S})$ gives that $R(\boldsymbol{\delta}^I,\boldsymbol{\Sigma}^{-1})-R(\boldsymbol{\delta}^{SH},\boldsymbol{\Sigma}^{-1})<0$. Similarly, we can show that $R(\boldsymbol{\delta}^I,\boldsymbol{\Sigma}^{-1})<R(\boldsymbol{\delta}^{EN},\boldsymbol{\Sigma}^{-1})$, and the proof of Theorem 6.1 is complete. ■

For instance, the estimators $\boldsymbol{\delta}^{EM}$ with $f^*_{EM}(\boldsymbol{S})=(\mathrm{tr}\,\boldsymbol{S}_1+\cdots+\mathrm{tr}\,\boldsymbol{S}_k)^\alpha$ and $\boldsymbol{\delta}^{JS}$ with $f^*_{JS}(\boldsymbol{S})=\|\boldsymbol{S}\|_M^{-\beta}$ are dominated by

$$\begin{aligned}\boldsymbol{\delta}^{IM}&=\boldsymbol{\delta}^{UB}-4\mathcal{D}\log(f^*_{EM}(\boldsymbol{S})f^*_{JS}(\boldsymbol{S}))\\&=\boldsymbol{\delta}^{EM}+\frac{4\beta}{\|\boldsymbol{S}\|_M^2}\boldsymbol{S}\\&=\boldsymbol{\delta}^{JS}-\frac{4\alpha}{\sum_{i=1}^k\mathrm{tr}\,\boldsymbol{S}_i}\boldsymbol{I},\end{aligned}$$

where $0<\alpha<1$ and $0<\beta<p_0-2$ for $p_0=\sum_{i=1}^k p_i(p_i+1)/2$.

Remark 6.1 We now consider geometric interpretation of the above result. Let $\boldsymbol{F}=(\boldsymbol{F}_1,\ldots,\boldsymbol{F}_k)$ and $\boldsymbol{G}=(\boldsymbol{G}_1,\ldots,\boldsymbol{G}_k)$, where $\boldsymbol{F}_i$'s and $\boldsymbol{G}_i$'s are, respectively, $p_i\times p_i$ matrices and the elements of $\boldsymbol{F}_i$'s and $\boldsymbol{G}_i$'s are functions of $\boldsymbol{S}$. Define the inner product of $\boldsymbol{F}$ and $\boldsymbol{G}$ as $\langle\boldsymbol{F},\boldsymbol{G}\rangle=E[\boldsymbol{F}\circ\boldsymbol{G}]$ and denote the norm of $\boldsymbol{F}$ by $\|\boldsymbol{F}\|_E=\sqrt{\langle\boldsymbol{F},\boldsymbol{F}\rangle}$. It is noted that $\langle\boldsymbol{\delta}^{UB}-\boldsymbol{\Sigma}^{-1},\boldsymbol{\Sigma}^{-1}\rangle=0$, namely, $\boldsymbol{\delta}^{UB}-\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma}^{-1}$ intersect orthogonally. Also, note that both norms $\|\boldsymbol{\delta}^{UB}\|_E$ and $\|\boldsymbol{\delta}^{UB}-\boldsymbol{\Sigma}^{-1}\|_E$ depend on $\boldsymbol{\Sigma}^{-1}$.

Let S_O be the open sphere centered at the origin $\boldsymbol{O}$ with radius $\|\boldsymbol{\delta}^{UB}\|_E$ and also let S_Σ be the open sphere centered at $\boldsymbol{\Sigma}^{-1}$ with radius $\|\boldsymbol{\delta}^{UB}-\boldsymbol{\Sigma}^{-1}\|_E$. Note that all shrinkage type estimators improving upon the unbiased estimator $\boldsymbol{\delta}^{UB}$ belong to $S_O\cap S_\Sigma$ and that all enlargement type estimators improving upon $\boldsymbol{\delta}^{UB}$ belong to $S_O^c\cap S_\Sigma$, where S_O^c denotes the complement of S_O.

Figure 1 shows positions of the estimators $\boldsymbol{\delta}^{UB}$, $\boldsymbol{\delta}^{EM}$, $\boldsymbol{\delta}^{JS}$ and $\boldsymbol{\delta}^{IM}$ in the metric space with norm $\|\cdot\|_E$. The notation B_O and B_Σ denote the boundaries of S_O and S_Σ, respectively. $\boldsymbol{\delta}^{EM}$ is a shrinkage type estimator and belongs to $S_O \cap S_\Sigma$. Also $\boldsymbol{\delta}^{JS}$ is an enlargement type estimator and belongs to $S_O^c \cap S_\Sigma$. Moreover $\boldsymbol{\delta}^{IM}$ is given by parallel translation from $\boldsymbol{\delta}^{EM}$ along $\boldsymbol{\delta}^{JS} - \boldsymbol{\delta}^{UB}$ or from $\boldsymbol{\delta}^{JS}$ along $\boldsymbol{\delta}^{EM} - \boldsymbol{\delta}^{UB}$. ■

6.2 Improvement upon the usual estimator

As seen from Theorem 3.1, the usual estimator $\boldsymbol{\delta}^{US} = \boldsymbol{\delta}^{UB} - 4\gamma \boldsymbol{S}^{-1}$ with $0 < \gamma < 1$ improves upon the unbiased estimator $\boldsymbol{\delta}^{UB}$ relative to the loss (2.2). From the result in the preceding subsection, we can improve upon the usual estimator $\boldsymbol{\delta}^{US}$ by adding an enlargement factor such as $-4\mathcal{D}\log f_{JS}^*(\boldsymbol{S})$ since $\boldsymbol{\delta}^{US}$ is a shrinkage estimator. In this subsection we consider improvement upon $\boldsymbol{\delta}^{US}$ by an estimator having another enlargement factor than $-4\mathcal{D}\log f_{JS}^*(\boldsymbol{S})$.

The estimator considered here is of the form

$$\boldsymbol{\delta}^{IU} = \boldsymbol{\delta}^{US} - 4\mathcal{D}\log f(\boldsymbol{S}),$$

whose risk function can be expressed as

$$\begin{aligned} R(\boldsymbol{\delta}^{IU}, \boldsymbol{\Sigma}^{-1}) &= R(\boldsymbol{\delta}^{US}, \boldsymbol{\Sigma}^{-1}) - 8E\Big[\sum_{i=1}^{k} \mathrm{tr}\,\{(n_i - p_i - 1 - 4\gamma)\boldsymbol{S}_i^{-1} - \boldsymbol{\Sigma}_i^{-1}\}\mathcal{D}_i \log f(\boldsymbol{S})\Big] \\ &\qquad + 16E[\|\mathcal{D}\log f(\boldsymbol{S})\|_M^2] \\ &= R(\boldsymbol{\delta}^{US}, \boldsymbol{\Sigma}^{-1}) + 16E[\mathcal{D}\circ\mathcal{D}\log f(\boldsymbol{S})] + 32\gamma E[\boldsymbol{S}^{-1}\circ\mathcal{D}\log f(\boldsymbol{S})] \\ &\qquad + 16E[\|\mathcal{D}\log f(\boldsymbol{S})\|_M^2] \\ &= R(\boldsymbol{\delta}^{US}, \boldsymbol{\Sigma}^{-1}) + E\Big[\frac{16}{f(\boldsymbol{S})}(\mathcal{D}\circ\mathcal{D}f(\boldsymbol{S}) + 2\gamma\boldsymbol{S}^{-1}\circ\mathcal{D}f(\boldsymbol{S}))\Big]. \end{aligned}$$

Then we get the following theorem.

Theorem 6.2 *The estimator* $\boldsymbol{\delta}^{IU}$ *dominates* $\boldsymbol{\delta}^{US}$ *relative to the loss* (2.2) *if*

$$\mathcal{D}\circ\mathcal{D}f(\boldsymbol{S}) + 2\gamma\boldsymbol{S}^{-1}\circ\mathcal{D}f(\boldsymbol{S}) < 0. \tag{6.2}$$

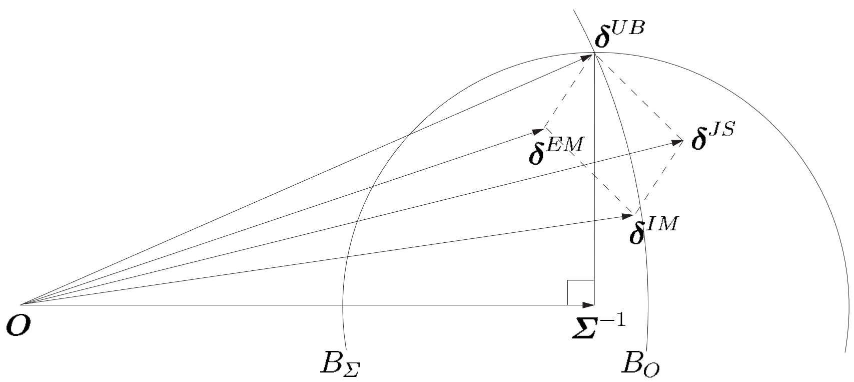

Figure 1: Geometric interpretation with respect to the estimators $\boldsymbol{\delta}^{UB}$, $\boldsymbol{\delta}^{EM}$, $\boldsymbol{\delta}^{JS}$ and $\boldsymbol{\delta}^{IM}$.

We shall seek the function $f(\boldsymbol{S})$ satisfying (6.2). First, we look into $f^*_{EM}(\boldsymbol{S}) = (\mathrm{tr}\,\boldsymbol{S}_1 + \cdots + \mathrm{tr}\,\boldsymbol{S}_k)^\alpha$. For $f^*_{EM}(\boldsymbol{S})$, assume that $-\mathcal{D}_i \log f^*_{EM}(\boldsymbol{S})$ is positive definite, namely, α is negative constant. Noting that

$$\Big(\sum_{i=1}^k \mathrm{tr}\,\boldsymbol{S}_i\Big)\Big(\sum_{i=1}^k \mathrm{tr}\,\boldsymbol{S}_i^{-1}\Big) \geq p_*^2$$

for $p_* = \sum_{i=1}^k p_i$ and that $\gamma > 0$, we can evaluate the left hand-side of (6.2) for $f^*_{EM}(\boldsymbol{S})$ as

$$\begin{aligned}\mathcal{D}\circ\mathcal{D}f^*_{EM}(\boldsymbol{S}) + 2\gamma\boldsymbol{S}^{-1}\circ\mathcal{D}f^*_{EM}(\boldsymbol{S}) &= \alpha(\alpha-1)p_*\Big(\sum_{i=1}^k \mathrm{tr}\,\boldsymbol{S}_i\Big)^{\alpha-2} \\ &\quad + 2\alpha\gamma\Big(\sum_{i=1}^k \mathrm{tr}\,\boldsymbol{S}_i\Big)^{\alpha-1}\sum_{i=1}^k \mathrm{tr}\,\boldsymbol{S}_i^{-1} \\ &\leq \Big(\sum_{i=1}^k \mathrm{tr}\,\boldsymbol{S}_i\Big)^{\alpha-2} \times p_*\{\alpha^2 + \alpha(2\gamma p_* - 1)\}.\end{aligned}$$

Therefore we have

Corollary 6.1 *The estimator $\boldsymbol{\delta}^{IEM} = \boldsymbol{\delta}^{US} - 4\mathcal{D}\log f^*_{EM}(\boldsymbol{S})$ dominates the usual estimator $\boldsymbol{\delta}^{US}$ relative to the loss* (2.2) *if $-(2\gamma p_* - 1) < \alpha < 0$.*

We next check the condition (6.2) for $f^*_{AM}(\boldsymbol{S}) = |\boldsymbol{T}|^\varepsilon = |\boldsymbol{S}_1 + \cdots + \boldsymbol{S}_k|^\varepsilon$ when $p = p_1 = \cdots = p_k$. Assume that ε is a negative constant. To evaluate the condition (6.2) for $f^*_{AM}(\boldsymbol{S})$ we use the following fact: Define the arithmetic and the harmonic mean matrices of positive definite matrices $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_k$ as $(\boldsymbol{S}_1 + \cdots + \boldsymbol{S}_k)/k$ and $k(\boldsymbol{S}_1^{-1} + \cdots + \boldsymbol{S}_k^{-1})^{-1}$, respectively. We then get the matrix inequality for the arithmetic and the harmonic mean matrices of positive definite matrices.

Lemma 6.1 *For positive definite matrices $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_k$, we have the followings.*

(1) $k(\boldsymbol{S}_1^{-1} + \cdots + \boldsymbol{S}_k^{-1})^{-1} \leq (\boldsymbol{S}_1 + \cdots + \boldsymbol{S}_k)/k$,

(2) $(\boldsymbol{S}_1^{-1} + \cdots + \boldsymbol{S}_k^{-1})/k \geq k(\boldsymbol{S}_1 + \cdots + \boldsymbol{S}_k)^{-1}$.

The equalities hold in (1) *and* (2) *if and only if $\boldsymbol{S}_1 = \cdots = \boldsymbol{S}_k$.*

Proof. The matrix inequality (1) is proved by Sagae and Tanabe (1994) and the matrix inequality (2) is verified by combining (1) and the fact that $\boldsymbol{A}^{-1} \geq \boldsymbol{B}^{-1}$ if $\boldsymbol{A} \leq \boldsymbol{B}$ for two positive definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. ■

From the above lemma, it is seen that $\sum_{i=1}^k \mathrm{tr}\,\boldsymbol{S}_i^{-1}\boldsymbol{T}^{-1} \geq k^2 \mathrm{tr}\,\boldsymbol{T}^{-2}$ for $\boldsymbol{T} = \boldsymbol{S}_1 + \cdots +$

$\boldsymbol{S}_k$. We also use the inequality $\{\mathrm{tr}\,\boldsymbol{T}^{-1}\}^2 \le p\mathrm{tr}\,\boldsymbol{T}^{-2}$ to give that

$$\begin{aligned}
\mathcal{D}\circ\mathcal{D}f^*_{AM}(\boldsymbol{S}) + 2\gamma\boldsymbol{S}^{-1}\circ\mathcal{D}f^*_{AM}(\boldsymbol{S}) &= k\varepsilon^2|\boldsymbol{T}|^\varepsilon \mathrm{tr}\,\boldsymbol{T}^{-2} - \frac{k\varepsilon}{2}|\boldsymbol{T}|^\varepsilon(\{\mathrm{tr}\,\boldsymbol{T}^{-1}\}^2 + \mathrm{tr}\,\boldsymbol{T}^{-2}) \\
&\quad + 2\varepsilon\gamma|\boldsymbol{T}|^\varepsilon \sum_{i=1}^k \mathrm{tr}\,\boldsymbol{S}_i^{-1}\boldsymbol{T}^{-1} \\
&\le k\varepsilon^2|\boldsymbol{T}|^\varepsilon \mathrm{tr}\,\boldsymbol{T}^{-2} - (k\varepsilon/2)(p+1)|\boldsymbol{T}|^\varepsilon \mathrm{tr}\,\boldsymbol{T}^{-2} \\
&\quad + 2k^2\varepsilon\gamma|\boldsymbol{T}|^\varepsilon \mathrm{tr}\,\boldsymbol{T}^{-2} \\
&= k|\boldsymbol{T}|^\varepsilon \mathrm{tr}\,\boldsymbol{T}^{-2} \times \{\varepsilon^2 + \varepsilon(4k\gamma - p - 1)/2\}.
\end{aligned}$$

Hence we get

Corollary 6.2 *The estimator* $\boldsymbol{\delta}^{IAM} = \boldsymbol{\delta}^{US} - 4\mathcal{D}\log f^*_{AM}(\boldsymbol{S})$ *dominates the usual estimator* $\boldsymbol{\delta}^{US}$ *relative to the loss* (2.2) *if* $-(4k\gamma - p - 1)/2 < \varepsilon < 0$.

We finally examine $f^*_{JS}(\boldsymbol{S}) = \|\boldsymbol{S}\|_M^{-\beta}$. Using Theorem 6.1, we can see that $\boldsymbol{\delta}^{IJS} = \boldsymbol{\delta}^{US} - 4\mathcal{D}\log f^*_{JS}(\boldsymbol{S})$ dominates $\boldsymbol{\delta}^{US}$ under the condition that $0 < \beta < p_0 - 2$. Calculating the left hand-side of (6.2) for $f^*_{JS}(\boldsymbol{S})$ directly, on the other hand, we can show that

$$\begin{aligned}
\mathcal{D}\circ\mathcal{D}f^*_{JS}(\boldsymbol{S}) + 2\gamma\boldsymbol{S}^{-1}\circ\mathcal{D}f^*_{JS}(\boldsymbol{S}) &= \|\boldsymbol{S}\|_M^{-\beta-2}\{\beta(\beta+2) - \beta p_0\} - 2\beta\gamma p_*\|\boldsymbol{S}\|_M^{-\beta-2} \\
&= \|\boldsymbol{S}\|_M^{-\beta-2}\{\beta^2 - \beta(p_0 + 2\gamma p_* - 2)\},
\end{aligned}$$

where $p_0 = \sum_{i=1}^k p_i(p_i+1)/2$. Thus we obtain the better condition for the dominance of $\boldsymbol{\delta}^{IJS} = \boldsymbol{\delta}^{US} - 4\mathcal{D}\log f^*_{JS}(\boldsymbol{S})$ over $\boldsymbol{\delta}^{US}$.

Corollary 6.3 *If* $0 < \beta < p_0 + 2\gamma p_* - 2$, *then* $\boldsymbol{\delta}^{IJS} = \boldsymbol{\delta}^{US} - 4\mathcal{D}\log f^*_{JS}(\boldsymbol{S})$ *dominates* $\boldsymbol{\delta}^{US}$ *relative to the loss* (2.2).

In the special case of $p_1 = \cdots = p_k = 1$, the best constant γ of the usual estimator $\boldsymbol{\delta}^{US}$ is given by $\gamma = 1/2$. The condition that $\boldsymbol{\delta}^{IJS}$ improves upon $\boldsymbol{\delta}^{US}$ is $\beta^2 - 2\beta(k-1) < 0$ and $k \ge 2$, namely, it is possible to improve upon the best usual estimator

$$\boldsymbol{\delta}^{BU} = ((n_1 - 4)/s_1, \ldots, (n_k - 4)/s_k)$$

for even $k = 2$, but then impossible to improve upon the unbiased estimator $\boldsymbol{\delta}^{UB}$. See also Berger (1980).

7 Numerical studies

In this section we compare the risk functions of alternative estimators of $\boldsymbol{\Sigma}^{-1}$ under the loss function (2.2).

For risk comparison in case of $p = p_1 = \cdots = p_k$, we examined the six estimators $\boldsymbol{\delta}^{UB}$, $\boldsymbol{\delta}^{US}$, $\boldsymbol{\delta}^{IJS}$, $\boldsymbol{\delta}^{IEM}$, $\boldsymbol{\delta}^{IAM}$ and $\boldsymbol{\delta}^{IGB}$, of which i-th components are given by, respectively,

(1) $\boldsymbol{\delta}_i^{UB} = (n_i - p - 1)\boldsymbol{S}_i^{-1}$,

(2) $\boldsymbol{\delta}_i^{US} = (n_i - p - 1 - 4\gamma)\boldsymbol{S}_i^{-1}$,

(3) $\boldsymbol{\delta}_i^{IJS} = \boldsymbol{\delta}_i^{US} + \{2(p_0 + 2\gamma p_* - 2)/\|\boldsymbol{S}\|_M^2\}\boldsymbol{S}_i$,

(4) $\boldsymbol{\delta}_i^{IEM} = \boldsymbol{\delta}_i^{US} + \{2(2\gamma p_* - 1)/\sum_{i=1}^k \operatorname{tr} \boldsymbol{S}_i\}\boldsymbol{I}_p$,

(5) $\boldsymbol{\delta}_i^{IAM} = \boldsymbol{\delta}_i^{US} + (4k\gamma - p - 1)(\boldsymbol{S}_1 + \cdots + \boldsymbol{S}_k)^{-1}$,

(6) $\boldsymbol{\delta}_i^{IGB} = \boldsymbol{\delta}_i^{GB} + \{2(p_0 - 2)/\|\boldsymbol{S}\|_M^2\}\boldsymbol{S}_i$, where the prior of $\boldsymbol{\delta}_i^{GB}$ is given by (4.2) with $a = 1$ and $b = 1/2$.

Here $p_0 = kp(p+1)/2$ and $p_* = kp$. It is noted from Corollaries 6.1, 6.2 and 6.3 that $\boldsymbol{\delta}^{IJS}$, $\boldsymbol{\delta}^{IEM}$ and $\boldsymbol{\delta}^{IAM}$ dominate the usual estimator $\boldsymbol{\delta}^{US}$ and, also, from Theorem 6.1 that $\boldsymbol{\delta}^{IGB}$ dominates the unbiased estimator $\boldsymbol{\delta}^{UB}$ relative to the loss (2.2).

We note that $\boldsymbol{\delta}_i^{GB}$, $i = 1, \ldots, k$, is expressed by

$$\boldsymbol{\delta}_i^{GB} = \frac{E^{\boldsymbol{W}|\boldsymbol{S}}[\boldsymbol{W}_i \cdot \pi_B(\boldsymbol{W})]}{E^{\boldsymbol{W}|\boldsymbol{S}}[\pi_B(\boldsymbol{W})]}, \tag{7.1}$$

where $E^{\boldsymbol{W}|\boldsymbol{S}}$ denotes conditional expectation with respect to $\boldsymbol{W} = (\boldsymbol{W}_1, \ldots, \boldsymbol{W}_k)$ given $\boldsymbol{S}$. Here the conditional distribution of $\boldsymbol{W}_i$ given $\boldsymbol{S}_i$ is $\mathcal{W}_p(\boldsymbol{S}_i^{-1}, n_i - p - 1)$ for $i = 1, \ldots, k$ and the $\boldsymbol{W}_i$'s are independent. Hence the estimates of $\boldsymbol{\delta}^{GB}$ were derived from the Monte Carlo approximations for the two expectations in the denominator and the numerator of (7.1).

The estimates of risk values were computed by 10,000 independent replications. We chose $k = 3$, $p = 2$ and $\gamma = 1/2$ and took three sets of sample size $(n_1, n_2, n_3) = (10, 10, 10)$, $(30, 10, 50)$ and $(50, 70, 30)$. For the precision matrices $\boldsymbol{\Sigma}_i^{-1}$'s, we considered the following case: $\boldsymbol{\Sigma}_1^{-1} = \boldsymbol{I}_p$, $\boldsymbol{\Sigma}_2^{-1} = (1 + c)\boldsymbol{I}_p$, and $\boldsymbol{\Sigma}_3^{-1} = (1 + c)^{-1}\boldsymbol{I}_p$ for $c \geq 0$.

The simulation results are given in Figures 2, 3 and 4. The curves in Figures are those of the relative risks for each alternative estimator and the unbiased estimator, that is, the ratio of risks of an alternative estimator $\boldsymbol{\delta}$ and $\boldsymbol{\delta}^{UB}$,

$$\mathrm{RR} = R(\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1})/R(\boldsymbol{\delta}^{UB}, \boldsymbol{\Sigma}^{-1}).$$

Note that the RR is a function of c and that an estimator $\boldsymbol{\delta}$ is better than $\boldsymbol{\delta}^{UB}$ if $\mathrm{RR} < 1$. 'UB', 'US', 'IJS', 'IEM', 'IAM' and 'IGB' denote $\boldsymbol{\delta}^{UB}$, $\boldsymbol{\delta}^{US}$, $\boldsymbol{\delta}^{IJS}$, $\boldsymbol{\delta}^{IEM}$, $\boldsymbol{\delta}^{IAM}$ and $\boldsymbol{\delta}^{IGB}$, respectively.

The simulation results given in Figures 2, 3 and 4 show the following important observations.

1. In the case that the precision matrices $\boldsymbol{\Sigma}_i^{-1}$'s are the identity matrices, $\boldsymbol{\delta}^{IJS}$ is the best for all sets of sample size. For small sample size $(n_1, n_2, n_3) = (10, 10, 10)$, the RR of $\boldsymbol{\delta}^{IJS}$ is about 0.35 and $\boldsymbol{\delta}^{IJS}$ has substantial reduction in risk than $\boldsymbol{\delta}^{UB}$ and $\boldsymbol{\delta}^{US}$.

2. When the $\boldsymbol{\Sigma}_i^{-1}$'s are much different, $\boldsymbol{\delta}^{IEM}$ is excellent in the six estimators. Particularly in the case of $(n_1, n_2, n_3) = (30, 10, 50)$, it is favorable over wide range of c.

3. The risk behavior of $\boldsymbol{\delta}^{IEM}$ and $\boldsymbol{\delta}^{IAM}$ are very similar. However the RR of $\boldsymbol{\delta}^{IEM}$ is slightly better than that of $\boldsymbol{\delta}^{IAM}$.

4. The risk of $\boldsymbol{\delta}^{IGB}$ is near that of $\boldsymbol{\delta}^{IJS}$, and is better than that of $\boldsymbol{\delta}^{US}$ except that the $\boldsymbol{\Sigma}_i^{-1}$'s disperse for $(n_1, n_2, n_3) = (30, 10, 50)$.

5. The risk reduction of estimators are large when the sample size n_i's are small. In such case, the risk variation over c, too, are large.

6. In the case that the n_i's disperse, the maximum reduction in risks is not given by $c = 0$, namely, the $\boldsymbol{\Sigma}_i^{-1}$'s being the identity matrices, probably since any risk of alternative estimators is dependent on the $\boldsymbol{\Sigma}_i^{-1}$'s.

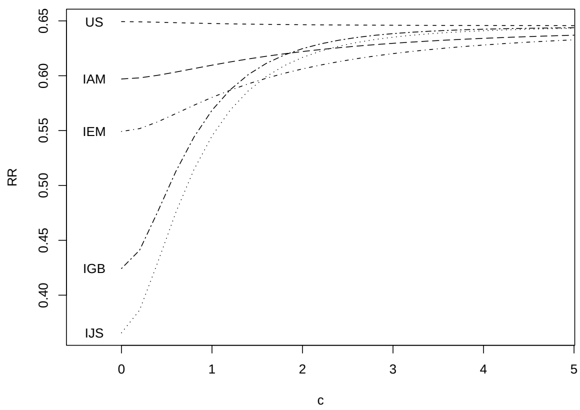

Figure 2: The relative risks in case of $(n_1, n_2, n_3) = (10, 10, 10)$.

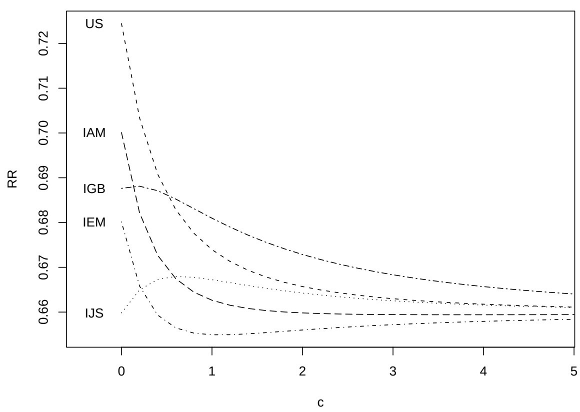

Figure 3: The relative risks in case of $(n_1, n_2, n_3) = (30, 10, 50)$.

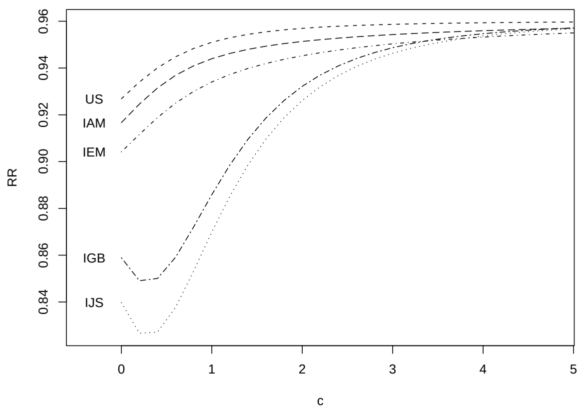

Figure 4: The relative risks in case of $(n_1, n_2, n_3) = (50, 70, 30)$.

References

[1] Dey, D.K. (1987). Improved estimation of a multinormal precision matrix, *Statist. Probab. Lett.*, **6**, 125–128.

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

[3] Haff, L.R. (1977). Minimax estimators for a multinormal precision matrix, *J. Multivariate Anal.*, **7**, 374–385.

[4] Haff, L.R. (1979). An identity for the Wishart distribution with applications, *J. Multivariate Anal.*, **9**, 531–544.

[5] Haff, L.R. (1981). Further identities for the Wishart distribution with applications in regression, *Canad. J. Statist.*, **9**, 215–224.

[6] Haff, L.R. (1982). Identities for the inverse Wishart distribution with computational results in linear and quadratic discrimination. *Sankhyā, ser.* **B 44**, 245–258.

[7] Haff, L.R. (1986). On linear log-odds and estimation of discriminant coefficients, *Commun. Statist.-Theor. Meth.*, **15**, 2131–2144.

[8] 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.

[9] Krishnamoorthy, K. and Gupta, A.K. (1989). Improved minimax estimation of a normal precision matrix, *Canad. J. Statist.*, **17**, 91–102.

[10] Kubokawa, T. (2005). A revisit to estimation of the precision matrix of the Wishart distribution, *J. Statist. Reserach*, **39**, 97–120.

[11] Sagae, M. and Tanabe, K. (1994). Upper and lower bounds for the arithmetic-geometric-harmonic means of positive definite matrices, *Linear and Multilinear Algebra*, **37**, 279-282.

[12] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution, *Ann. Statist.*, **9**, 1135–1151.

[13] Tsukuma, H. and Konno, Y. (2006). On improved estimation of normal precision matrix and discriminant coefficients, *Journal of Multivariate Analysis*, **97**, 1477–1500.