Estimation of the Precision Matrix of a Singular Wishart Distribution and its Application in High Dimensional Data

Tatsuya Kubokawa*and Muni S. Srivastava†
University of Tokyo and University of Toronto

July 28, 2005

Abstract

In this article, Stein-Haff identity is established for a singular Wishart distribution with a positive definite mean matrix but with the dimension larger than the degrees of freedom. This identity is then used to obtain estimators of the precision matrix improving on the estimator based on the Moore-Penrose inverse of the Wishart matrix under the Efron-Morris loss function and its variants. Ridge-type empirical Bayes estimators of the precision matrix are also given and their dominance properties over the usual one are shown using this identity. Finally, these precision estimators are used in a quadratic discriminant rule, and it is shown through simulation that the use of the ridge-type empirical Bayes estimators provides higher correct classification rates.

Key words and phrases: Covariance matrix, discriminant analysis, dominance property, Efron-Morris loss function, empirical Bayes procedure, multivariate classification, precision matrix, singular Wishart, Stein-Haff identity.

1 Introduction

The estimation of the precision matrix, namely the inverse of the covariance matrix $\mathbf{\Sigma}$, of a multivariate normal distribution has been an important issue in practical situations as well as from theoretical aspects, and when the dimension p is smaller than the number of observations n, Efron and Morris (1976) considered this problem. But, when $p > n$, the Wishart matrix is singular, and thus many estimators can be constructed by using a generalized inverse of the sample covariance matrix. However, Srivastava (2004) proposed the unique Moore-Penrose inverse of the sample covariance matrix as it uses the sufficient

*Faculty of Economics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: tatsuya@e.u-tokyo.ac.jpFaculty

†Department of Statistics, University of Toronto, 100 St George Street, Toronto, Ontario, CANADA M5S 3G3, E-Mail: srivasta@utstat.toronto.edu

statistic for $\boldsymbol{\Sigma}$. In this paper, we obtain several estimators theoretically improving on the Moore-Penrose inverse estimator of the precision matrix, some of which are shown to be very useful in discriminant analysis.

To specify the problem considered here, let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ be independently and identically distributed (i.i.d.) as multivariate normal with mean vector $\boldsymbol{\mu}$ and a $p \times p$ positive definite matrix $\boldsymbol{\Sigma}$ denoted as $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} > 0$. Let

$$\overline{\boldsymbol{x}} = N^{-1} \sum_{i=1}^{N} \boldsymbol{x}_i, \quad n = N - 1$$

and

$$\boldsymbol{W} = \sum_{i=1}^{N} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^t.$$

Then

$$\boldsymbol{W} = \boldsymbol{Y}\boldsymbol{Y}^t$$

where $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are i.i.d. $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{W}$ has a Wishart distribution with mean $n\boldsymbol{\Sigma}$ and degrees of freedom n, denoted as $\mathcal{W}_p(\boldsymbol{\Sigma}, n)$. When $n < p$, it is called a singular Wishart distribution, whose distribution has been recently studied by Srivastava (2003). In many inference procedures, an estimate of the precision matrix $\boldsymbol{\Sigma}^{-1}$ is required. Srivastava (2004) used $n\boldsymbol{W}^+$, where $\boldsymbol{W}^+$ is the Moore-Penrose inverse of $\boldsymbol{W}$. We shall consider a generalized version of this estimator for estimating the precision matrix. It is given by

$$\boldsymbol{\delta}_a = a\boldsymbol{W}^+$$

for a constant a. The main aim of this paper is to develop estimators of $\boldsymbol{\Sigma}^{-1}$ improving on the usual one $\boldsymbol{\delta}_a$ in term of risk in a decision-theoretic framework. To evaluate the risk of $\boldsymbol{\delta}_a$, however, we cannot employ the Stein loss function $L_S(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \mathrm{tr}\, \boldsymbol{\delta}\boldsymbol{\Sigma} - \log |\boldsymbol{\delta}\boldsymbol{\Sigma}| - p$ for estimator $\boldsymbol{\delta}$, because of the singularity of $\boldsymbol{W}^+$. Alternative loss functions are of the forms

$$L_k(\boldsymbol{\delta}, \boldsymbol{\Sigma}) = \mathrm{tr}\, (\boldsymbol{\delta} - \boldsymbol{\Sigma}^{-1})^2 \boldsymbol{W}^k \quad \text{for} \quad k = 0, 1, 2,$$

where the L_1-loss was used by Efron and Morris (1976), and the L_1- and L_0-losses were used by Haff (1979b). However, all the estimators that we obtain in Section 2 under the losses L_0 and L_1 dominating the estimator $n\boldsymbol{W}^+$, require not only that $p > n$ but that $n = O(p^{\varepsilon})$, $0 \leq \varepsilon < 1$. In practical applications this is a severe restriction. On the other hand, no such restriction on n and p are required under the loss function L_2. Thus, for obtaining a ridge type empirical Bayes estimator of the precision matrix $\boldsymbol{\Sigma}^{-1}$, we consider only L_2 loss function.

To develop analytical dominance properties of estimators, we need to derive the so-called Stein-Haff identity in the singular Wishart distribution. The Stein-Haff identity was derived by Stein (1977) and Haff (1979a) for the full rank Wishart distribution. A similar identity for the elliptically contoured model has been given by Kubokawa and Srivastava (1999). It has been well known that the Stein-Haff identity is a very powerful tool to develop significant dominance results. In the Appendix, we derive the Stein-Haff identity in the singular Wishart distribution, which is equally powerful. With the help of this identity, we obtain in Section 2 several estimators dominating $\boldsymbol{\delta}_a$ under the three loss functions L_0, L_1 and L_2. In Section 3, the empirical Bayes approach to the estimation of the precision matrix is given to provide ridge-type stable procedures dominating $\boldsymbol{\delta}_a$ under the loss function L_2.

It may be of great interest to investigate how much useful the improved estimators of the precision matrix are in practical multivariate inference procedures. While its application in tests and confidence intervals for mean vectors are currently under investigation, we in this article consider an equally important problem of classifying an observation vector into one of two groups with unequal covariance matrices. Through simulations we show that our emperical Bayes procedures using nonsingular ridge type estimators for the precision matrices provide significantly higher correct classification rates for the quadratic classification rules; these rates are very close to the rates obtained when all the parameters are known.

2 Estimation of the precision matrix

For estimating the precision matrix in the case of $p > n$, in this paper, we consider orthogonally equivariant estimators of the general form

$$\begin{aligned}\boldsymbol{\delta}(\boldsymbol{\Phi}) &= \boldsymbol{H}_1\boldsymbol{\Phi}(\boldsymbol{\ell})\boldsymbol{H}_1^t, \\ \boldsymbol{\Phi}(\boldsymbol{\ell}) &= \mathrm{diag}\,(\phi_1(\boldsymbol{\ell}),\dots,\phi_n(\boldsymbol{\ell}))\,.\end{aligned} \tag{2.1}$$

Instead of the function $\boldsymbol{\Phi}(\boldsymbol{\ell})$, we often use the function $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\ell}) = \mathrm{diag}\,(\psi_1(\boldsymbol{\ell}),\dots,\psi_n(\boldsymbol{\ell}))$ for

$$\begin{aligned}\boldsymbol{\Psi}(\boldsymbol{\ell}) =& \boldsymbol{L}\boldsymbol{\Phi}(\boldsymbol{\ell}), \\ \psi_i(\boldsymbol{\ell}) =& \ell_i\phi_i(\boldsymbol{\ell}), \quad i = 1,\dots,n.\end{aligned}$$

To evaluate the estimators, we use the following three loss functions

$$L_0(\boldsymbol{\delta},\boldsymbol{\Sigma}) = \mathrm{tr}\,(\boldsymbol{\delta}-\boldsymbol{\Sigma}^{-1})^2, \tag{2.2}$$

$$L_1(\boldsymbol{\delta},\boldsymbol{\Sigma}) = \mathrm{tr}\,(\boldsymbol{\delta}-\boldsymbol{\Sigma}^{-1})^2\boldsymbol{W}, \tag{2.3}$$

$$L_2(\boldsymbol{\delta},\boldsymbol{\Sigma}) = \mathrm{tr}\,(\boldsymbol{\delta}-\boldsymbol{\Sigma}^{-1})^2\boldsymbol{W}^2, \tag{2.4}$$

which are here called the L_0-loss, the L_1-loss and the L_2-loss functions. The risk function of estimator $\boldsymbol{\delta}$ relative to the L_k-loss is written by $R_k(\boldsymbol{\Sigma},\boldsymbol{\delta}) = E[L_k(\boldsymbol{\delta},\boldsymbol{\Sigma})]$ for $k = 0$, 1 and 2. Dominance results in terms of the risks are given below for the L_1, L_0 and L_2 loss functions, but all the proofs are given in the Appendix. It is specially noted that the Stein-Haff identity (A.1) in the singular Wishart distribution is quite useful for establishing the dominance properties and the derivation of the indentity is also given in the Appendix.

2.1 Dominance results relative to the L_1-loss

We first handle the L_1-loss, for it is the most tractable of the three. The risk function of $\boldsymbol{\delta}(\boldsymbol{\Phi})$ under the loss $L_1(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ is expressed as

$$R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}(\boldsymbol{\Phi})) = E\left[\mathrm{tr}\,\{\boldsymbol{\delta}(\boldsymbol{\Phi})\}^2\boldsymbol{W} - 2\mathrm{tr}\,\boldsymbol{\delta}(\boldsymbol{\Phi})\boldsymbol{W}\boldsymbol{\Sigma}^{-1}\right] + n\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}. \tag{2.5}$$

Then the Stein-Haff identity given by Lemma A.1 is applied to rewrite $E[\mathrm{tr}\,\boldsymbol{\delta}(\boldsymbol{\Phi})\boldsymbol{W}\boldsymbol{\Sigma}^{-1}]$ as

$$\begin{aligned} E\left[\mathrm{tr}\,\boldsymbol{\delta}(\boldsymbol{\Phi})\boldsymbol{W}\boldsymbol{\Sigma}^{-1}\right] =& E\left[\mathrm{tr}\,\boldsymbol{H}_1\boldsymbol{\Phi}\boldsymbol{L}\boldsymbol{H}_1^t\boldsymbol{\Sigma}^{-1}\right] \\ =& \sum_{i=1}^n E\left[(p-n-1)\phi_i + 2\frac{\partial}{\partial \ell_i}(\ell_i\phi_i) + 2\sum_{j>i}\frac{\ell_i\phi_i - \ell_j\phi_j}{\ell_i - \ell_j}\right]. \end{aligned} \tag{2.6}$$

Combining (2.5) and (2.6), we get the following expression of the risk function.

Proposition 2.1 *Assume that $p > n+1$. The risk function of the orthogonally equivariant estimator $\boldsymbol{\delta}(\boldsymbol{\Phi})$ relative to the L_1-loss* (2.3) *is expressed by*

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

In this article, we consider estimators of the kind $a\boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{H}_1^t$, $a > 0$. From Proposition 2.1, this estimator has the risk $\{a^2 - 2(p-n-1)a\}E[\mathrm{tr}\,\boldsymbol{L}^{-1}] + n\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}$, which is minimized at $a = p-n-1$. Hence, the estimator with the best multiple is

$$\boldsymbol{\delta}_1 = a_1\boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{H}_1^t, \quad a_1 = p-n-1$$

with the risk $R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}_1) = -a_1^2E[\mathrm{tr}\,\boldsymbol{L}^{-1}] + n\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}$. Although it is not possible to get an unbiased estimator of the risk $R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}(\boldsymbol{\Phi}))$ in the case of $p > n$, we can provide an unbiased estimator of the risk difference $R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}(\boldsymbol{\Phi})) - R_1(\boldsymbol{\Sigma}, \boldsymbol{\delta}_1)$, which gives a sufficient condition for improving on the estimator $\boldsymbol{\delta}_1$.

Proposition 2.2 *The estimator $\boldsymbol{\delta}(\boldsymbol{\Phi})$ dominates $\boldsymbol{\delta}_1$ relative to the L_1-loss* (2.3) *if $\psi_i(\boldsymbol{\ell})$'s satisfy the inequality*

$$\sum_{i=1}^n\left\{\frac{\psi_i^2}{\ell_i} - 2a_1\frac{\psi_i}{\ell_i} - 4\sum_{j>i}\frac{\psi_i-\psi_j}{\ell_i-\ell_j} - 4\frac{\partial\psi_i}{\partial\ell_i}\right\} \le -\sum_{i=1}^n\frac{a_1^2}{\ell_i},$$

for $p > n+1$, and $\psi_i = \ell_i\phi_i$.

The following proposition is very useful for developing improved estimators.

Proposition 2.3 *Assume that* $\boldsymbol{\Psi}(\boldsymbol{\ell}) = \mathrm{diag}\,(\psi_1(\boldsymbol{\ell}), \ldots, \psi_n(\boldsymbol{\ell}))$ *satisfies the following conditions for* $p > n+1$*:*

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

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

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

Then the estimator $\boldsymbol{\delta}(\boldsymbol{L}^{-1}\boldsymbol{\Psi}) = \boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{\Psi}(\boldsymbol{\ell})\boldsymbol{H}_1^t$ *dominates the estimator* $\boldsymbol{\delta}_1$ *relative to the* L_1*-loss* (2.3).

Proposition 2.3 directly provides an example of the Stein type estimator given by

$$\begin{aligned} \boldsymbol{\delta}^S =& \boldsymbol{H}_1\boldsymbol{D}\boldsymbol{L}^{-1}\boldsymbol{H}_1^t, \\ \boldsymbol{D} =& \mathrm{diag}\,(d_1, \ldots, d_n), \\ d_i =& n + p - 2i - 1 \quad \text{for} \quad i = 1, \ldots, n. \end{aligned} \tag{2.7}$$

This corresponds to the case of $\phi_i = d_i$ or $\psi_i = \ell_i d_i$, and the dominance property follows from Proposition 2.3.

Proposition 2.4 *The Stein type estimator* $\boldsymbol{\delta}^S$ *dominates* $\boldsymbol{\delta}_1$ *under the* L_1*-loss.*

The Stein type estimator $\boldsymbol{\delta}^S$ can be further improved on by the estimator

$$\boldsymbol{\delta}^{IS}(g) = \boldsymbol{\delta}^S + \frac{g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{W}}\boldsymbol{I}_p, \tag{2.8}$$

where $g(\boldsymbol{\ell})$ is an absolutely continuous function. This dominance property follows from Proposition 2.5.

Proposition 2.5 *Assume that* $g(\boldsymbol{\ell})$ *satisfies the conditions:*

(a) $\partial g(\boldsymbol{\ell})/\partial\ell_i \geq 0$ *for* $i = 1, \ldots, n$.

(b) $0 < g(\boldsymbol{\ell}) \leq 4(n-1)$.

Then the estimator $\boldsymbol{\delta}^{IS}(g)$ *dominates the Stein type estimator* $\boldsymbol{\delta}^S$ *under the* L_1*-loss.*

Putting $g(\boldsymbol{\ell}) = 2(n-1)$ in (2.8) gives the improved estimator

$$\boldsymbol{\delta}^{IS} = \boldsymbol{H}_1\boldsymbol{D}\boldsymbol{L}^{-1}\boldsymbol{H}_1^t + \frac{2(n-1)}{\mathrm{tr}\,\boldsymbol{W}}\boldsymbol{I}_p, \tag{2.9}$$

which we shall call *the improved Stein type estimator*. It is noted that $\boldsymbol{\delta}^{IS}$ has a form similar to the Efron-Morris type estimator given by

$$\boldsymbol{\delta}^{EM} = (p - n - 1)\boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{H}_1^t + \frac{(n-1)(n+2)}{\mathrm{tr}\,\boldsymbol{W}}\boldsymbol{I}_p. \tag{2.10}$$

Using the same arguments as in the proof of Proposition 2.5 shows that $\boldsymbol{\delta}^{EM}$ dominates $\boldsymbol{\delta}_1$ relative to the L_1-loss, but it is not known if it dominates $\boldsymbol{\delta}^S$ or $\boldsymbol{\delta}^{IS}$.

Proposition 2.3 allows us to produce a new type of improved estimator, given by

$$\boldsymbol{\delta}^R = \boldsymbol{H}_1\mathrm{diag}\,(\phi_1^R(\boldsymbol{\ell}), \ldots, \phi_n^R(\boldsymbol{\ell}))\boldsymbol{H}_1^t, \tag{2.11}$$

where for $i = 1, \ldots, n$,

$$\phi_i^R(\boldsymbol{\ell}) = \frac{d_i}{\ell_i + \hat{\lambda}_i} \quad \text{and} \quad d_i = n + p - 2i - 1.$$

Here, $\hat{\lambda}_n = 0$ and for $i = 1, \ldots, n-1$, $\hat{\lambda}_i$ is a function of $\ell_{i+1}, \ldots, \ell_n$ defined sequentially by

$$\hat{\lambda}_i = (d_i \hat{\lambda}_{i+1} + 2\ell_i)/d_{i+1}. \tag{2.12}$$

It is interesting to note that the estimator $\boldsymbol{\delta}^R$ is a ridge type because of nonnegativeness of $\hat{\lambda}_i$'s.

Proposition 2.6 *The ridge type estimator $\boldsymbol{\delta}^R$ dominates $\boldsymbol{\delta}_1$ under the L_1-loss.*

2.2 Dominance results relative to the L_0-loss

The risk function of $\boldsymbol{\delta}(\boldsymbol{\Phi})$ under the loss $L_0(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ is expressed as

$$R_0(\boldsymbol{\Sigma}, \boldsymbol{\delta}(\boldsymbol{\Phi})) = E\left[\mathrm{tr}\,\{\boldsymbol{\delta}(\boldsymbol{\Phi})\}^2 - 2\mathrm{tr}\,\boldsymbol{\delta}(\boldsymbol{\Phi})\boldsymbol{\Sigma}^{-1}\right] + \mathrm{tr}\,\boldsymbol{\Sigma}^{-2}.$$

Using the Stein-Haff identity given by Lemma A.1 for the term $E[\mathrm{tr}\,\boldsymbol{\delta}(\boldsymbol{\Phi})\boldsymbol{\Sigma}^{-1}]$, we get the following expression of the risk function.

Proposition 2.7 *Assume that $p > n+3$. The risk function of the orthogonally equivariant estimator $\boldsymbol{\delta}(\boldsymbol{\Phi})$ relative to the L_0-loss* (2.2) *is expressed by*

$$\begin{aligned} R_0(\boldsymbol{\Sigma},&\boldsymbol{\delta}(\boldsymbol{\Phi})) - \mathrm{tr}\,\boldsymbol{\Sigma}^{-2} \\ &= \sum_{i=1}^n E\left[\phi_i^2 - 2(p-n-1)\frac{\phi_i}{\ell_i} - 4\sum_{j>i}\frac{\phi_i - \phi_j}{\ell_i - \ell_j} - 4\frac{\partial \phi_i}{\partial \ell_i}\right] \\ &= \sum_{i=1}^n E\left[\frac{\psi_i^2 - 2(p-n-3)\psi_i}{\ell_i^2} - 4\sum_{j>i}\frac{\ell_j\psi_i - \ell_i\psi_j}{\ell_i\ell_j(\ell_i - \ell_j)} - \frac{4}{\ell_i}\frac{\partial \psi_i}{\partial \ell_i}\right], \end{aligned}$$

where $\boldsymbol{\Phi}(\boldsymbol{\ell}) = \boldsymbol{L}^{-1}\boldsymbol{\Psi}(\boldsymbol{\ell}) = \mathrm{diag}\,(\psi_1(\boldsymbol{\ell})/\ell_1, \ldots, \psi_n(\boldsymbol{\ell})/\ell_n)$ for $\psi_i = \ell_i\phi_i$.

From Proposition 2.7, the estimator of the form $a\boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{H}_1^t$ has the risk that

$$\begin{aligned} R_0(\boldsymbol{\Sigma},&a\boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{H}_1^t) - \mathrm{tr}\,\boldsymbol{\Sigma}^{-2} \\ &= \{a^2 - 2(p-n-3)a\}E[\mathrm{tr}\,\boldsymbol{L}^{-2}] + 4a\sum_{i=1}^n\sum_{j>i}E[1/(\ell_i\ell_j)] \\ &= \{a^2 - 2(p-n-2)a\}E[\mathrm{tr}\,\boldsymbol{L}^{-2}] + 2aE[(\mathrm{tr}\,\boldsymbol{L}^{-1})^2], \end{aligned} \tag{2.13}$$

since $2\sum_{i=1}^n\sum_{j>i}1/(\ell_i\ell_j) = (\mathrm{tr}\,\boldsymbol{L}^{-1})^2 - \mathrm{tr}\,\boldsymbol{L}^{-2}$. This expression shows that the best constant a does not exist, but we suggest a reasonable choice of a given by

$$\boldsymbol{\delta}_0 = a_0\boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{H}_1^t, \quad a_0 = p - n - 3.$$

As seen from (2.13), $R_0(\boldsymbol{\Sigma}, \boldsymbol{\delta}_0) \leq R_0(\boldsymbol{\Sigma}, a\boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{H}_1^t)$ for any $a > a_0$ and any $\boldsymbol{\Sigma}$, which implies that $\boldsymbol{\delta}_0$ dominates $\boldsymbol{\delta}_1$ as pointed out by Haff (1979b). An unbiased estimator of the risk difference $R_0(\boldsymbol{\Sigma}, \boldsymbol{\delta}(\boldsymbol{\Phi})) - R_0(\boldsymbol{\Sigma}, \boldsymbol{\delta}_0)$ can be provided and a sufficient condition for improving on the estimator $\boldsymbol{\delta}_1$ is given in the following.

Proposition 2.8 *The estimator $\boldsymbol{\delta}(\boldsymbol{\Phi})$ dominates $\boldsymbol{\delta}_0$ relative to the L_0-loss* (2.2) *if $\psi_i(\boldsymbol{\ell})$'s satisfy the inequality*

$$\sum_{i=1}^{n} \left\{ \frac{\psi_i^2 - 2a_0\psi_i}{\ell_i^2} - 4\sum_{j>i} \frac{\ell_j\psi_i - \ell_i\psi_j}{\ell_i\ell_j(\ell_i - \ell_j)} - \frac{4}{\ell_i}\frac{\partial\psi_i}{\partial\ell_i} \right\} \leq \sum_{i=1}^{n} \left\{ -\frac{a_0^2}{\ell_i^2} + \sum_{j>i} \frac{4a_0}{\ell_i\ell_j} \right\},$$

for $p > n + 3$.

Proposition 2.8 provides the condition for the estimator $\boldsymbol{\delta}(\boldsymbol{\Phi})$ to dominate $\boldsymbol{\delta}_0$ in the case of $p > n$. Some improved estimators proposed by Haff (1979b) and Dey (1987) for $n > p$ can hold dominance properties in the case $p > n$ by interchanging n and p. Of these, Dey (1987) proposed the use of estimators of the form

$$\boldsymbol{\delta}^{DE}(g) = \boldsymbol{\delta}_0 + \frac{g(\boldsymbol{\ell})}{\operatorname{tr} \boldsymbol{W}^2}\boldsymbol{W}, \tag{2.14}$$

where $g(\boldsymbol{\ell})$ is an absolutely continuous function. The following proposition provides conditions for $\boldsymbol{\delta}^{DE}(g)$ to dominate $\boldsymbol{\delta}_0$.

Proposition 2.9 *Assume that $g(\boldsymbol{\ell})$ satisfies the conditions:*
(a) $\partial g(\boldsymbol{\ell})/\partial\ell_i \geq 0$ *for $i = 1, \ldots, n$.*
(b) $0 < g(\boldsymbol{\ell}) \leq 2(n-1)(n+4)$.
Then the estimator $\boldsymbol{\delta}^{DE}(g)$ dominates the Stein type estimator $\boldsymbol{\delta}_0$ under the L_0-loss.

Putting $g(\boldsymbol{\ell}) = (n-1)(n+4)$ in (2.15) gives the improved estimator

$$\boldsymbol{\delta}^{DE} = \boldsymbol{\delta}_0 + \frac{(n-1)(n+4)}{\operatorname{tr} \boldsymbol{W}^2}\boldsymbol{W}, \tag{2.15}$$

which we shall call *the Dey type estimator.*

Finally, we shall derive a Stein type estimator dominating $\boldsymbol{\delta}_0$ like Propositions 2.3 and 2.4. Let $r = (n-1)/2$ if n is odd and $r = n/2$ if n is even. Define constants $d_i^\dagger$ by

$$\begin{aligned} d_i^\dagger &= \max\{a_0 + 2(n+1-2i),\ a_0\} \\ &= \max\{p + n - 4i - 1,\ p - n - 3\} \\ &= \begin{cases} p + n - 4i, & \text{if } i = 1, \ldots, r, \\ p - n - 3, & \text{if } i = r+1, \ldots, n. \end{cases} \end{aligned}$$

and let $\boldsymbol{D}^\dagger = \operatorname{diag}(d_1^\dagger, \ldots, d_n^\dagger)$. The resulting Stein type estimator is of the form

$$\boldsymbol{\delta}^{S\dagger} = \boldsymbol{H}_1\boldsymbol{D}^\dagger\boldsymbol{L}^{-1}\boldsymbol{H}_1^t.$$

The dominance property of $\boldsymbol{\delta}^{S\dagger}$ over $\boldsymbol{\delta}_0$ follows from the following proposition.

Proposition 2.10 *Assume that* $\mathbf{\Psi}(\boldsymbol{\ell}) = \mathrm{diag}\,(\psi_1(\boldsymbol{\ell}), \ldots, \psi_n(\boldsymbol{\ell}))$ *satisfies the following conditions for* $p > n+3$*:*

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

(b) $\psi_1(\boldsymbol{\ell}) \geq \cdots \geq \psi_n(\boldsymbol{\ell}) = p - n - 3 \equiv a_0$.

(c) $d_i^\dagger = \max\{p + n - 4i - 1,\ p - n - 3\} \geq \psi_i$ *for* $i = 1, \ldots, n$.

Then the estimator $\boldsymbol{\delta}(\boldsymbol{L}^{-1}\mathbf{\Psi}) = \boldsymbol{H}_1\boldsymbol{L}^{-1}\mathbf{\Psi}(\boldsymbol{\ell})\boldsymbol{H}_1^t$*, given by* (2.1)*, dominates the estimator* $\boldsymbol{\delta}_0$ *relative to the* L_0*-loss* (2.3)*.*

Proposition 2.10 provides not only the Stein type estimator $\boldsymbol{\delta}^{S\dagger}$, but also a ridge type estimator for improving on $\boldsymbol{\delta}_0$. Define $\hat{\lambda}_i^\dagger$ sequentially by

$$\hat{\lambda}_i^\dagger = \begin{cases} (d_i^\dagger\hat{\lambda}_{i+1}^\dagger + 4\ell_i)/d_{i+1}^\dagger, & \text{for } i = 1, \ldots, r, \\ 0, & \text{for } i = r+1, \ldots, n. \end{cases}$$

The ridge type estimator is given by

$$\boldsymbol{\delta}^{R\dagger} = \boldsymbol{H}_1 \mathrm{diag}\,(\phi_1^{R\dagger}(\boldsymbol{\ell}), \ldots, \phi_n^{R\dagger}(\boldsymbol{\ell}))\boldsymbol{H}_1^t, \tag{2.16}$$

where for $i = 1, \ldots, n$,

$$\phi_i^{R\dagger}(\boldsymbol{\ell}) = \frac{d_i^\dagger}{\ell_i + \hat{\lambda}_i^\dagger}.$$

Then the same arguments as in the proof of Proposition 2.6 can be used to show that $\boldsymbol{\delta}^{R\dagger}$ dominates $\boldsymbol{\delta}_0$.

Proposition 2.11 *The ridge type estimator* $\boldsymbol{\delta}^{R\dagger}$ *dominates* $\boldsymbol{\delta}_0$ *under the* L_0*-loss.*

2.3 Dominance results relative to the L_2-loss

The risk function of $\boldsymbol{\delta}(\mathbf{\Phi})$ under the loss $L_2(\boldsymbol{\delta}, \mathbf{\Sigma})$ is expressed as

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

Using the Stein-Haff identity given by Lemma A.1, we can derive an expression of the risk function.

Proposition 2.12 *Assume that* $p > n+1$*. The risk function of the orthogonally equivariant estimator* $\boldsymbol{\delta}(\mathbf{\Phi})$ *relative to the* L_2*-loss* (2.4) *is expressed by*

$$\begin{aligned} R_2(\mathbf{\Sigma},&\boldsymbol{\delta}(\mathbf{\Phi})) - E[\mathrm{tr}\,\boldsymbol{W}^2\mathbf{\Sigma}^{-2}] \\ &= \sum_{i=1}^n E\left[\psi_i^2 - 2(p+n+1-2i)\psi_i - 4\sum_{j>i}\frac{\ell_j(\psi_i - \psi_j)}{\ell_i - \ell_j} - 4\ell_i\frac{\partial\psi_i}{\partial\ell_i}\right], \end{aligned}$$

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

From Proposition 2.12, the estimator of the form $a\boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{H}_1^t$ has the risk that $n\{a^2 - 2pa\} + E[\mathrm{tr}\,\boldsymbol{W}^2\boldsymbol{\Sigma}^{-2}]$, which is minimized at $a = p$. Hence, the estimator with the best multiple is

$$\boldsymbol{\delta}_2 = a_2\boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{H}_1^t, \quad a_2 = p$$

with the risk $R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}_2) = -np^2 + +E[\mathrm{tr}\,\boldsymbol{W}^2\boldsymbol{\Sigma}^{-2}]$. Although it is not possible to get an unbiased estimator of the risk $R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}(\boldsymbol{\Phi}))$ in the case of $p > n$, we can provide an unbiased estimator of the risk difference $R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}(\boldsymbol{\Phi})) - R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}_1)$, which gives a sufficient condition for improving on the estimator $\boldsymbol{\delta}_2$.

Proposition 2.13 *The estimator $\boldsymbol{\delta}(\boldsymbol{\Phi})$ dominates $\boldsymbol{\delta}_2$ relative to the L_2-loss (2.4) if $\psi_i(\boldsymbol{\ell})$'s satisfy the inequality*

$$\sum_{i=1}^{n} \left\{ \psi_i^2 - 2(p+n+1-2i)\psi_i + p^2 - 4\sum_{j>i} \frac{\ell_j(\psi_i - \psi_j)}{\ell_i - \ell_j} - 4\ell_i \frac{\partial \psi_i}{\partial \ell_i} \right\} \le 0$$

for $p > n+1$.

One candidate for improving on $\boldsymbol{\delta}_2$ may be the Efron-Morris type estimator

$$\boldsymbol{\delta}^{EM*}(g) = \boldsymbol{\delta}_2 + \frac{g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{W}}\boldsymbol{I}_p, \tag{2.17}$$

where $g(\boldsymbol{\ell})$ is an absolutely continuous function. The following proposition provides the conditions for $\boldsymbol{\delta}^{EM*}(g)$ to dominate $\boldsymbol{\delta}_2$.

Proposition 2.14 *Assume that $g(\boldsymbol{\ell})$ satisfies the conditions:*
(a) $\partial g(\boldsymbol{\ell})/\partial \ell_i \ge 0$ *for* $i = 1, \dots, n$.
(b) $0 < g(\boldsymbol{\ell}) \le 2(n-1)$.
Then the estimator $\boldsymbol{\delta}^{EM*}(g)$ dominates the estimator $\boldsymbol{\delta}_2$ under the L_2-loss.

Putting $g(\boldsymbol{\ell}) = n - 1$ in (2.18) gives the improved estimator

$$\boldsymbol{\delta}^{EM*} = \boldsymbol{\delta}_2 + \frac{n-1}{\mathrm{tr}\,\boldsymbol{W}}\boldsymbol{I}_p. \tag{2.18}$$

From Proposition 2.13, we can also get another condition for the estimator $\boldsymbol{\delta}(\boldsymbol{\Phi})$ to dominate $\boldsymbol{\delta}_2$ in the case of $p > n+1$.

Proposition 2.15 *Assume that $\boldsymbol{\Psi}(\boldsymbol{\ell}) = \mathrm{diag}\,(\psi_1(\boldsymbol{\ell}), \dots, \psi_n(\boldsymbol{\ell}))$ satisfies the following conditions for $p > n+1$:*
(a) $\partial \psi_i(\boldsymbol{\ell})/\partial \ell_i \ge 0$ *for* $i = 1, \dots, n$.
(b) $\psi_1(\boldsymbol{\ell}) \ge \cdots \ge \psi_n(\boldsymbol{\ell})$.
(c) $\sum_{i=1}^n \{\psi_i^2 - 2(p+n+1-2i)\psi_i + p^2\} \le 0$.
Then the estimator $\boldsymbol{\delta}(\boldsymbol{L}^{-1}\boldsymbol{\Psi}) = \boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{\Psi}(\boldsymbol{\ell})\boldsymbol{H}_1^t$ dominates the estimator $\boldsymbol{\delta}_2$ relative to the L_2-loss (2.4).

Proposition 2.15 provides an example of the Stein type estimator given by

$$\begin{aligned}\boldsymbol{\delta}^{S*} =& \boldsymbol{H}_1 \boldsymbol{D}^* \boldsymbol{L}^{-1} \boldsymbol{H}_1^t,\\ \boldsymbol{D}^* =& \mathrm{diag}\,(d_1^*, \ldots, d_n^*),\\ d_i^* =& p + n - 2i + 1 \quad \text{for} \quad i = 1, \ldots, n.\end{aligned}$$

Proposition 2.16 *The Stein type estimator $\boldsymbol{\delta}^{S*}$ dominates $\boldsymbol{\delta}_2$ under the L_2-loss.*

In the above section, we have obtained estimators under the three loss functions L_0, L_1 and L_2. These estimators dominate the estimators of the kind $a\boldsymbol{W}^+$. However, all the estimators obtained under L_0 and L_1 loss functions involve a factor of $p - n$. This in turn implies that $n = O(p^\varepsilon)$, $0 \le \varepsilon < 1$. Such a restriction, however, is not needed under the loss function L_2. Thus, from now on, we will consider estimators only under the loss function L_2. But even under the L_2 loss function, the estimator of the kind $p\boldsymbol{W}^+$ is not stable due to smallness of smaller eigenvalues. Thus, it would be desirable to obtain ridge-type estimators of the precision matrix under the loss function L_2. This is obtained in the next section using empirical Bayes methods.

3 Empirical Bayes estimator of the precision matrix

The estimators of the precision matrix $\boldsymbol{\Sigma}^{-1}$ given in the previous section have the shortcoming of their singularity in the case of $p > n$. An approach to deriving nonsingular estimators of $\boldsymbol{\Sigma}^{-1}$ is to consider ridge type estimators of the form $a(\boldsymbol{W} + \lambda \boldsymbol{I}_p)^{-1}$ for positive constants a and λ. The important issue in the use of the ridge type estimators is how to choose the ridge parameter λ. We here employ an empirical Bayes method for giving estimators of λ and show that the resulting ridge type empirical Bayes estimators dominate the usual estimator $p\boldsymbol{W}^+$ relative to the L_2-loss function.

3.1 Empirical Bayes procedures

In our Bayesian setup, we assume that $\boldsymbol{\Sigma}^{-1}$ has a Wishart distribution with mean $r\lambda^{-1}\boldsymbol{I}_p$, $\lambda > 0$, $r > p$, and degrees of freedom r. That is, $\boldsymbol{\Sigma}^{-1} \sim \mathcal{W}_p(\lambda^{-1}\boldsymbol{I}_p, r)$, $\lambda > 0$, $r \ge p$, with the density

$$\pi(\boldsymbol{\Sigma}^{-1}) = c(p, r)|\lambda^{-1}\boldsymbol{I}_p|^{-r/2}|\boldsymbol{\Sigma}^{-1}|^{(r-p-1)/2}\mathrm{etr}(-\frac{1}{2}\lambda\boldsymbol{\Sigma}^{-1})$$

where etr $(\boldsymbol{A})$ stands for the exponential of the trace of the matrix $\boldsymbol{A}$ and

$$c(p, r) = \left[2^{(pr)/2}\Gamma_p(r/2)\right]^{-1}, \quad \Gamma_p(r/2) = \pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma(\frac{r - i + 1}{2}).$$

Since $\boldsymbol{W}$ is distributed as the singular Wishart distribution $\mathcal{W}_p(\boldsymbol{\Sigma}, n)$ for $n < p$, there exist a random variable $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ such that $\boldsymbol{W} = \boldsymbol{Y}\boldsymbol{Y}^t$ and $\boldsymbol{y}_i$'s are i.i.d. $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} > 0$. Then, the joint p.d.f. of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ given $\boldsymbol{\Sigma}^{-1}$ is given by

$$(2\pi)^{-(np)/2}|\boldsymbol{\Sigma}^{-1}|^{n/2}\mathrm{etr}\,\{-2^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\boldsymbol{Y}^t\}.$$

Hence, the joint p.d.f. of $\boldsymbol{Y}$ and $\boldsymbol{\Sigma}^{-1}$ is given by

$$c(p,r)(2\pi)^{-(pn)/2}|\lambda \boldsymbol{I}_p|^{r/2}|\boldsymbol{\Sigma}^{-1}|^{(n+r-p-1)/2}\mathrm{etr}\,\{-2^{-1}\boldsymbol{\Sigma}^{-1}(\lambda \boldsymbol{I}_p+\boldsymbol{Y}\boldsymbol{Y}^t)\}.$$

From this joint density, it is seen that the posterior distribution of $\boldsymbol{\Sigma}^{-1}$ given $\boldsymbol{Y}$ is given by $\mathcal{W}_p((\lambda \boldsymbol{I}_p+\boldsymbol{Y}\boldsymbol{Y}^t)^{-1}, n+r)$. Then, the Bayes estimator of $\boldsymbol{\Sigma}^{-1}$ is written by

$$\boldsymbol{\delta}^B(\lambda) = E[\boldsymbol{\Sigma}^{-1}|\boldsymbol{Y}] = (n+r)(\lambda \boldsymbol{I}_p+\boldsymbol{Y}\boldsymbol{Y}^t)^{-1}. \tag{3.1}$$

Since λ is unknown, it should be estimated from the marginal density of $\boldsymbol{Y}$ whose density is given by

$$\frac{c(p,r)}{c(p,n+r)}(2\pi)^{-pn/2}|\lambda \boldsymbol{I}_p|^{r/2}/|\lambda \boldsymbol{I}_p+\boldsymbol{Y}\boldsymbol{Y}^t|^{-(n+r)/2}.$$

Making the transformation $\boldsymbol{V}=\boldsymbol{Y}^t\boldsymbol{Y}$, we obtain its marginal density from Lemma 3.2.3 of Srivastava and Khatri (1979) as

$$\begin{aligned}
&\frac{c(p,r)}{c(p,n+r)}\frac{2^{-pn/2}}{\Gamma_n(p/2)}\lambda^{-np/2}|\boldsymbol{V}|^{(p-n-1)/2}\ |\boldsymbol{I}_n+\lambda^{-1}\boldsymbol{V}|^{-(n+r)/2}\\
&=\frac{c(p,r)}{c(p,n+r)}\frac{2^{-pn/2}}{\Gamma_n(p/2)}|\lambda^{-1}\boldsymbol{V}|^{(p-n-1)/2}\ |\boldsymbol{I}_n+\lambda^{-1}\boldsymbol{V}|^{-(n+r)/2}\ \lambda^{-n(n+1)/2}
\end{aligned}$$

with respect to the nonsingular matrix $\boldsymbol{V}$. Note the expectations $E[|\lambda^{-1}\boldsymbol{V}|^{1/n}]$, $E[\mathrm{tr}\,\lambda^{-1}\boldsymbol{V}]$ and $E[\mathrm{tr}\,\lambda\boldsymbol{V}^{-1}]$ are constants independent of λ. These suggest the use of the following moment estimators as possible candidates of estimators of λ:

$$\begin{aligned}
\hat\lambda_G =& c|\boldsymbol{V}|^{1/n} = c(\prod_{i=1}^n \ell_i)^{1/n},\\
\hat\lambda_A =& c\mathrm{tr}\,\boldsymbol{V} = (cn)\frac{1}{n}\sum_{i=1}^n \ell_i,\\
\hat\lambda_H =& c/\mathrm{tr}\,\boldsymbol{V}^{-1} = (c/n)\frac{n}{\sum_{i=1}^n \ell_i^{-1}},
\end{aligned} \tag{3.2}$$

for positive constant c and the eigenvalues $\boldsymbol{\ell}=(\ell_1,\dots,\ell_n)$ of $\boldsymbol{V}$. It is noted that the estimators $\hat\lambda_G$, $\hat\lambda_A$ and $\hat\lambda_H$ are based on the geometric, the arithmetic and the harmonic means of $\ell_1,\dots,\ell_n$. Another type of estimator is provided by the solution of the equation

$$\sum_{i=1}^n \frac{\hat\lambda_M}{\hat\lambda_M+\ell_i} = c, \tag{3.3}$$

for a constant c satisfying $0<c<n$. This is analogous to the maximum likelihood estimator in the marginal density.

An empirical Bayes estimator can be derived by substituting an estimator of the hyperparameter into a Bayes estimator. When λ is estimated by an estimator $\hat\lambda=\hat\lambda(\boldsymbol{\ell})$, the empirical Bayes estimator of $\boldsymbol{\Sigma}^{-1}$ is given by

$$\boldsymbol{\delta}_a^{EB}(\hat\lambda) = a\left(\boldsymbol{W}+\hat\lambda \boldsymbol{I}_p\right)^{-1}, \tag{3.4}$$

where a is a positive constant suitably chosen. In the Bayes estimator (3.1), a is given by $a = n + r$. For $r = p - n$, a is $a_2 = p$, which is the best multiple under the L_2-loss function. In the next subsection, we examine the dominance property of the estimator in (3.4) under the loss function L_2.

3.2 Dominance property under L_2-loss

Now we shall investigate whether the empirical Bayes estimator $\boldsymbol{\delta}_a^{EB}(\hat{\lambda})$ given in (3.4) dominate the estimator of the form $\boldsymbol{\delta}_a = a\boldsymbol{W}^+$ for $\boldsymbol{W}^+ = \boldsymbol{H}_1\boldsymbol{L}^{-1}\boldsymbol{H}_1^t$ relative to the L_2-loss (2.4). Using the Stein-Haff identity given by Lemma A.1, we derive in the following proposition an unbiased estimator of the risk difference of the two estimators $\boldsymbol{\delta}_a^{EB}(\hat{\lambda})$ and $\boldsymbol{\delta}_a$, the proof of which is given in the Appendix.

Proposition 3.1 *An unbiased estimator of the risk difference*

$$\Delta_2(a, \hat{\lambda}) = R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}_a^{EB}(\hat{\lambda})) - R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}_a)$$

relative to the L_2-loss is given by

$$\begin{aligned}\widehat{\Delta}_2(a, \hat{\lambda}) =& a\sum_{i=1}^{n} \frac{\hat{\lambda}}{\ell_i + \hat{\lambda}}\Big\{(a+2)\frac{\hat{\lambda}}{\ell_i + \hat{\lambda}} + 2\sum_{j=1}^{n} \frac{\hat{\lambda}}{\ell_j + \hat{\lambda}} \\ &- 2(a - p + n + 1) + 4\frac{\ell_i^2}{\hat{\lambda}(\ell_i + \hat{\lambda})}\frac{\partial \hat{\lambda}}{\partial \ell_i}\Big\}. \end{aligned} \tag{3.5}$$

As demonstrated in Section 2.3, it is noted that the best multiple a of estimators $a\boldsymbol{W}^+$ is given by $a = p$ relative to the L_2-loss. From Proposition 3.1, it is seen that the ridge type empirical Bayes estimator

$$\boldsymbol{\delta}^{EB}(\hat{\lambda}) = p(\boldsymbol{W} + \hat{\lambda}\boldsymbol{I}_p)^{-1}$$

dominates the estimator $\boldsymbol{\delta}_2 = p\boldsymbol{W}^+$ if the ridge function $\hat{\lambda}$ satisfies the inequality

$$\sum_{i=1}^{n} \frac{\hat{\lambda}}{\ell_i + \hat{\lambda}}\Big\{(p+2)\frac{\hat{\lambda}}{\ell_i + \hat{\lambda}} + 2\sum_{j=1}^{n} \frac{\hat{\lambda}}{\ell_j + \hat{\lambda}} - 2(n+1) + 4\frac{\ell_i^2}{\hat{\lambda}(\ell_i + \hat{\lambda})}\frac{\partial \hat{\lambda}}{\partial \ell_i}\Big\} \le 0. \tag{3.6}$$

Using the condition (3.6), we first obtain a condition on c for the function $\hat{\lambda}_M$ to satisfy the inequality (3.6), where $\hat{\lambda}_M$ is the solution of the equation

$$\sum_{i=1}^{n} \frac{\hat{\lambda}_M}{\hat{\lambda}_M + \ell_i} = c. \tag{3.7}$$

Then from the implicit function theorem, we can see that the partial derivative $\partial\hat{\lambda}_M/\partial\ell_i$ is given by

$$\frac{\partial \hat{\lambda}_M}{\partial \ell_i} = \frac{\hat{\lambda}_M/(\ell_i + \hat{\lambda}_M)^2}{\sum_{j=1}^{n} \ell_j/(\ell_j + \hat{\lambda}_M)^2},$$

which is used to get that

$$\begin{aligned}\sum_i \frac{\ell_i^2(\partial\hat{\lambda}_M/\partial\ell_i)}{(\ell_i+\hat{\lambda}_M)^2} =&\hat{\lambda}_M \frac{\sum_i \ell_i^2/(\ell_i+\hat{\lambda}_M)^4}{\sum_i \ell_i/(\ell_i+\hat{\lambda}_M)^2} \\ \le&\hat{\lambda}_M \frac{\{\sum_i \ell_i/(\ell_i+\hat{\lambda}_M)^2\}^2}{\sum_i \ell_i/(\ell_i+\hat{\lambda}_M)^2} \\ =&\hat{\lambda}_M \sum_i \frac{\ell_i}{(\ell_i+\hat{\lambda}_M)^2} = \sum_i \frac{\hat{\lambda}_M}{\ell_i+\hat{\lambda}_M} - \sum_i \frac{\hat{\lambda}_M^2}{(\ell_i+\hat{\lambda}_M)^2}.\end{aligned}$$

From the inequality (3.6) and the equation $c = \sum_{i=1}^n \hat{\lambda}_M/(\ell_i+\hat{\lambda}_M)$, we get a sufficient condition given by

$$(p-2)\sum_i\sum_i \frac{\hat{\lambda}_M^2}{(\ell_i+\hat{\lambda}_M)^2} - 2(n-1)c + 2c^2 \le 0,$$

which can be satisfied for $0 < c \le 2(n-1)/p$ since $\sum_i \hat{\lambda}_M^2/(\ell_i+\hat{\lambda}_M)^2 \le c^2$. We thus get the following dominance result.

Proposition 3.2 *Assume that the constant c satisfies the inequality $0 < c \le 2(n-1)/p$. Let $\hat{\lambda}_M$ be the unique solution of the equation (3.7). Then, the ridge type empirical Bayes estimator*

$$\boldsymbol{\delta}^{EB}(\hat{\lambda}_M) = p(\boldsymbol{W} + \hat{\lambda}_M \boldsymbol{I}_p)^{-1} \tag{3.8}$$

dominates $\boldsymbol{\delta}_2 = p\boldsymbol{W}^+$ under the L_2-loss.

We next show that the function $\hat{\lambda}_H = c/\mathrm{tr}\,\boldsymbol{V}^{-1} = c/\sum_{i=1}^n \ell_i^{-1}$ satisfies the inequality (3.6). It is noted that

$$\begin{aligned}\frac{\ell_i^2}{\hat{\lambda}_H}\frac{\partial\hat{\lambda}_H}{\partial\ell_i} &= \frac{\hat{\lambda}_H}{c}, \\ \hat{\lambda}_H/\ell_i &= c/(1+\sum_{j\ne i}\ell_i/\ell_j) \le c.\end{aligned} \tag{3.9}$$

Then from the condition (3.6), we can get a sufficient condition given by

$$(p+2)\frac{c}{1+c} + 2n\frac{c}{1+c} - 2(n+1) + 4\frac{1}{1+c} \le 0,$$

which can be satisfied for $0 < c \le 2(n-1)/p$.

Proposition 3.3 *Assume that the constant c satisfies the inequality $0 < c \le 2(n-1)/p$. Then, the ridge type empirical Bayes estimator $\boldsymbol{\delta}^{EB}(\hat{\lambda}_H)$ dominates $\boldsymbol{\delta}_2 = p\boldsymbol{W}^+$ under the L_2-loss.*

From Propositions 3.2 and 3.3, it is seen that $\hat{\lambda}_M$ and $\hat{\lambda}_H$ are two superior estimators of λ in the sense that the resulting ridge type empirical Bayes estimators $\boldsymbol{\delta}^{EB}(\hat{\lambda}_M)$ and $\boldsymbol{\delta}^{EB}(\hat{\lambda}_H)$ have smaller risks than $\boldsymbol{\delta}_2$ relative to the L_2-loss. For the other estimators $\hat{\lambda}_G$ and $\hat{\lambda}_A$ given by (3.2), however, we could not show similar dominance properties of the resulting empirical Bayes estimators under the L_2-loss. This may be due to the unboundedness of the functions $\hat{\lambda}_G/\ell_i$ and $\hat{\lambda}_A/\ell_i$.

4 Application to multivariate classification

Now we investigate how much useful the improved estimators of the precision matrix are in the multivariate discriminant analysis. It should be noted that the use of the improved precision estimators does not theoretically guarantee the improvement in reducing the classification errors. Although the idea of using the improved precision estimators in the discriminant rule is quite intuitive, it is worth inspecting through the simulation studies. The related problems have been studied by Friedman (1989), Loh (1997) and Zhao, Honda and Konishi (1996) and others. Of these, Friedman (1989) handled the case of the dimension p being much larger than the degrees of freedom n, and proposed regularized discriminant rules where the ridge parameters are determined by the cross-validation method. We here try to answer the query about whether the correct classification rates can be improved or not by using the improved precision estimators derived in the previous sections.

We treat the problem of classifying observations into two classes of the distributions: $\pi_i : \mathcal{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ for unknown $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$, $i = 1, 2$. For $i = 1, 2$, let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{n_i}$ be a training sample from π_i, and suppose that $\boldsymbol{\mu}_i$ is estimated by the sample mean $\overline{\boldsymbol{x}}_i$ and that the precision matrix $\boldsymbol{\Sigma}_i^{-1}$ is estimated by $\boldsymbol{\delta}_i$ based on $\boldsymbol{W}_i = \sum_{j=1}^{n_i}(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_i)^t$. A new observation $\boldsymbol{x}$ is classified into π_1 if

$$(\boldsymbol{x} - \overline{\boldsymbol{x}}_1)^t \boldsymbol{\delta}_1 (\boldsymbol{x} - \overline{\boldsymbol{x}}_1) < (\boldsymbol{x} - \overline{\boldsymbol{x}}_2)^t \boldsymbol{\delta}_2 (\boldsymbol{x} - \overline{\boldsymbol{x}}_2), \tag{4.1}$$

and into π_2 otherwise. The simulation experiment is planned as follows: 100 training samples are generated and each training sample constructs the above quadratic discriminant rule, which classifies 200 new observations into the two classes of the distributions, where 100 observations are generated from π_1 and the other 100 from π_2. Thus, the correct classification rates are computed based on 100×200 total classifications.

As the estimators of the precision matrix $\boldsymbol{\Sigma}^{-1}$ based on $\boldsymbol{W}$, we use the following estimators: the Srivastava type estimators $n\boldsymbol{W}^+$ and $p\boldsymbol{W}^+$, the improved Stein estimator $\boldsymbol{\delta}^{IS}$ given by (2.9), the Efron-Morris estimator $\boldsymbol{\delta}^{EM}$ given by (2.10), the ridge-type empirical Bayes estimators $\boldsymbol{\delta}^{EB}(\hat{\lambda}_H)$, $\boldsymbol{\delta}^{EB}(\hat{\lambda}_M)$ given by (3.8) for $c = (n-1)/2$, and $\boldsymbol{\delta}^{EB}(\mathrm{tr}\, \boldsymbol{W}/p)$ and $\boldsymbol{\delta}^{EB}(\mathrm{tr}\, \boldsymbol{W}/n)$, which are abbreviated by SRn, SRp, IS, EM, RH, RM, RAp and RAn. All these estimators are used in the estimators $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ of the classification rule (4.1), for example, the estimator $n\boldsymbol{W}^+$, SRn, gives the rule $(\boldsymbol{x} - \overline{\boldsymbol{x}}_1)^t n_1 \boldsymbol{W}_1^+ (\boldsymbol{x} - \overline{\boldsymbol{x}}_1) < (\boldsymbol{x} - \overline{\boldsymbol{x}}_2)^t n_2 \boldsymbol{W}_2^+ (\boldsymbol{x} - \overline{\boldsymbol{x}}_2)$. In the simulation experiments, it is supposed that $\boldsymbol{\mu}_1 = \mu \times (1, \dots, 1)^t$ for $\mu = 0.8$, 1.0 and 1.5, $\boldsymbol{\mu}_2 = \boldsymbol{0}$, $\boldsymbol{\Sigma}_2 = \boldsymbol{I}_p$ and

$$\boldsymbol{\Sigma}_1 = \mathrm{diag}\,(\sigma_1, \dots, \sigma_p)(\rho_{ij})\mathrm{diag}\,(\sigma_1, \dots, \sigma_p),$$

where $\sigma_i = i/p$ for $i = 1, \dots, p$, and (ρ_{ij}) is a $p \times p$ matrix with $\rho_{ij} = (k/6)^{|i-j|/2}$ for $k = 1$, 2, 3, 4 and 5.

Table 1 reports the correct classification rates when the estimators SRn, SRp, IS, EM, RH, RM, RAp and RAn are used for $p = 100$, $(n_1, n_2) = (5, 5)$, $(5, 50)$ and $(50, 5)$, and $\mu = 0.8$, 1.0 and 1.5, where TR means the correct classifications rates based on the true rule

$$(\boldsymbol{x} - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_1) < (\boldsymbol{x} - \boldsymbol{\mu}_2)^t \boldsymbol{\Sigma}_2^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_2),$$

Table 1: Correct Classification Rates by Simulation for $p = 100$, $\mu = 0.8$, 1.0 and 1.5, and $(n_1, n_2) = (5, 5)$, $(5, 50)$ and $(50, 5)$.

	k	TR	SRn	SRp	IS	EM	RH	RM	RAp	RAn
	1	71.9	60.0	60.0	61.8	62.6	72.7	73.0	77.7	75.9
$\mu = 0.8$	2	72.1	57.2	57.2	58.4	59.3	73.1	73.4	80.2	77.6
$n_1 = 5$	3	72.3	54.9	54.9	55.7	56.7	73.7	74.0	82.9	79.3
$n_2 = 5$	4	71.2	52.6	52.6	53.1	54.0	73.9	74.1	85.4	81.0
	5	70.1	51.2	51.2	51.6	52.3	74.3	74.5	89.5	83.8
	1	97.2	64.6	64.6	67.2	70.2	90.6	90.8	93.5	92.7
$\mu = 1.0$	2	95.8	61.0	61.0	62.8	65.4	89.1	89.3	93.4	92.1
$n_1 = 5$	3	93.8	57.7	57.7	58.8	61.2	87.2	87.5	93.1	91.0
$n_2 = 5$	4	90.4	54.5	54.5	55.2	57.1	85.2	85.4	93.3	90.7
	5	85.1	52.2	52.2	52.7	54.1	82.9	83.1	94.4	90.4
	1	100.0	75.6	75.6	80.4	87.6	100.0	100.0	100.0	100.0
$\mu = 1.5$	2	100.0	70.5	70.5	74.8	81.9	99.7	99.7	99.9	99.8
$n_1 = 5$	3	100.0	65.5	65.5	68.8	75.3	99.3	99.3	99.8	99.6
$n_2 = 5$	4	100.0	60.0	60.0	62.5	67.8	98.0	98.0	99.5	99.0
	5	99.8	55.7	55.7	57.2	61.0	95.7	95.7	99.0	97.9
	1	97.4	50.0	53.9	60.4	58.1	66.4	66.3	50.0	70.0
$\mu = 1.0$	2	95.8	50.0	58.2	66.8	64.1	67.0	67.0	50.0	71.5
$n_1 = 5$	3	93.8	50.0	65.7	76.6	73.7	68.5	68.5	50.0	73.6
$n_2 = 50$	4	90.7	50.0	75.5	84.6	83.2	69.5	69.6	50.0	76.2
	5	85.3	50.0	83.2	85.1	87.3	71.0	70.8	50.0	79.9
	1	97.4	50.0	50.0	50.0	50.0	99.1	99.3	50.0	100.0
$\mu = 1.0$	2	95.8	50.0	50.0	50.0	50.0	98.9	99.1	50.0	100.0
$n_1 = 50$	3	93.8	50.0	50.0	50.0	50.0	98.2	98.0	50.0	100.0
$n_2 = 5$	4	90.7	50.0	50.0	50.0	50.0	97.2	97.7	50.0	100.0
	5	85.3	50.0	50.0	50.0	50.0	94.8	94.8	50.0	100.0

Table 2: Correct Classification Rates by Simulation for $p = 50$, $\mu = 1.5$ and $(n_1, n_2) = (30, 30)$, $(10, 10)$, $(5, 5)$, $(10, 30)$ and $(30, 10)$.

(n_1, n_2)	k	TR	SRn	SRp	IS	EM	RH	RM	RAp	RAn
	1	100.0	83.3	83.3	86.7	95.7	98.1	98.2	100.0	100.0
$n_1 = 30$	2	100.0	81.8	81.8	84.5	94.5	97.8	98.1	100.0	100.0
$n_2 = 30$	3	99.9	80.5	80.5	82.7	93.3	97.3	97.8	100.0	100.0
	4	99.8	79.8	79.8	80.8	91.8	96.4	96.4	100.0	100.0
	5	98.9	78.4	78.4	79.0	89.8	95.0	94.8	100.0	100.0
	1	100.0	79.1	79.1	82.3	92.1	99.4	99.5	99.9	99.9
$n_1 = 10$	2	100.0	74.3	74.3	77.4	88.5	98.9	99.0	99.9	99.7
$n_2 = 10$	3	100.0	68.8	68.8	71.5	83.4	98.1	98.3	99.8	99.5
	4	99.8	64.6	64.6	66.7	78.6	96.9	97.2	99.7	99.2
	5	98.9	61.1	61.1	62.8	74.2	95.4	95.4	99.8	99.0
	1	100.0	76.5	76.5	81.8	88.7	99.1	99.1	99.5	99.4
$n_1 = 5$	2	100.0	71.7	71.7	76.2	83.5	98.5	98.6	99.4	99.1
$n_2 = 5$	3	100.0	67.1	67.1	71.0	78.1	97.4	97.5	99.1	98.7
	4	99.8	62.1	62.1	65.3	71.9	96.2	96.3	98.8	97.9
	5	98.7	57.9	57.9	60.3	65.6	93.0	93.3	98.3	96.8
	1	100.0	66.9	95.1	99.3	100.0	98.4	98.4	59.5	99.4
$n_1 = 10$	2	100.0	74.6	97.8	99.5	99.9	97.7	97.7	65.4	98.9
$n_2 = 30$	3	100.0	81.3	98.6	98.7	99.4	96.9	97.0	72.3	98.8
	4	99.8	88.1	98.2	97.0	98.5	95.7	95.9	81.2	98.4
	5	98.8	93.5	96.2	93.2	95.8	93.6	93.3	90.8	98.3
	1	100.0	50.0	50.2	50.9	55.1	99.2	99.3	99.8	100.0
$n_1 = 30$	2	100.0	50.0	50.2	50.7	55.1	98.9	99.2	99.7	100.0
$n_2 = 10$	3	100.0	50.0	50.2	50.6	54.9	98.5	98.9	98.9	100.0
	4	99.8	50.0	50.3	50.7	55.5	98.0	98.1	97.3	100.0
	5	98.8	50.0	50.3	50.9	56.6	96.7	96.8	93.8	100.0

which uses the true parameters instead of their estimators. Table 2 reports the correct classification rates when the same estimators are used for $p = 50$, $\mu = 1.5$ and $(n_1, n_2) = (30, 30)$, $(10, 10)$, $(5, 5)$, $(10, 30)$ and $(30, 10)$.

Overview through the simulation results in Tables 1 and 2 reveals that the ridge-type empirical Bayes estimators RH, RM and RAn provide significant improvements in the correct classification rates for all the cases, and their rates are close to those of the true classification rule TR. Although the other ridge-type estimator RAp, which uses $\mathrm{tr}\, \boldsymbol{W}/p$ as an estimator of λ, provides good behaviors in the cases of equal sample sizes $n_1 = n_2$, but not good for sample sizes n_1 and n_2 extremely different. Use of the improved estimators IS and EM gains relatively small improvements in the classification. The difference between SRn and SRp appears in the fourth and the fifth part of the Table 1 and in the fourth part of Table 2. Overall it appears that SRn is better than SRp in terms of the correct classification rates for all the cases.

We conclude this section with giving comments for the interesting query about whether the use of the improved precision estimator leads to the improvement in the correct classification rates. It is noted that the estimator IS, EM, RH and RM dominate the estimators of the form $a\boldsymbol{W}^+$, especially RH and RM beat SRp as estimators of the precision matrix. Certainly, IS, EM, RH and RM outperform SRn and SRp for most cases, but the improvements of IS and EM are much smaller than those of RH and RM. On the other hand, the use of the ridge-type estimator RAn yields the substantial improvements although the dominance property of RAn over SRp cannot be guaranteed. Taking these observations into account, we can guess that the ridge-form in the precision estimators significantly affects the improvement in the correct classification rates rather than the dominance property of the precision estimators. In the use of the ridge-type estimators, the estimator of the ridge-parameter λ sensitively affects the behaviors of the discriminant rules, and the estimators $\hat{\lambda}_H$, $\hat{\lambda}_M$ and $\mathrm{tr}\, \boldsymbol{W}/n$, which correspond to RH, RM and RAn, are good choices in the quadratic discriminant rule. Among these three, RAn is the simplest with performance comparable to RM.

A Appendix

We here give the proofs of the propositions in Sections 2 and 3. For this purpose, we first develop the so-called Stein-Haff identity for the singular Wishart distribution.

A.1 Stein-Haff identity for the singular Wishart distribution

Let $\boldsymbol{H}_1$ be a $p \times n$ matrix such that $\boldsymbol{H}_1^t \boldsymbol{H}_1 = \boldsymbol{I}_n$, that is, $\boldsymbol{H}_1 \in \mathcal{H}_{n,p}$, the Stiefel manifold, and

$$\boldsymbol{W} = \boldsymbol{H}_1 \boldsymbol{L} \boldsymbol{H}_1^t,$$

where $\boldsymbol{L} = \mathrm{diag}\,(\ell_1, \ell_2, \ldots, \ell_n)$, an $n \times n$ diagonal matrix, where $\ell_1 > \ell_2 > \cdots > \ell_n$ are the n non-zero eigenvalues of the $p \times p$ matrix $\boldsymbol{W}$ of rank n. Let $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_n)$. Then, from Srivastava (2003, p.1549), the joint probability density function (p.d.f.) of $\boldsymbol{\ell}$ and $\boldsymbol{H}_1$

is given by

$$f(\boldsymbol{\ell}, \boldsymbol{H}_1, \boldsymbol{\Sigma}) = c_n(\boldsymbol{\Sigma}) b(\boldsymbol{\ell}) g_{m,p}(\boldsymbol{H}_1^t) \exp\left\{-2^{-1} \mathrm{tr}\, \boldsymbol{H}_1 \boldsymbol{L} \boldsymbol{H}_1^t \boldsymbol{\Sigma}^{-1}\right\},$$

where

$$\begin{aligned} b(\boldsymbol{\ell}) =& \prod_{i=1}^{n} \left\{ \ell_i^{(p-n-1)/2} \prod_{i<j} (\ell_i - \ell_j) \right\}, \\ c_n(\boldsymbol{\Sigma}) =& \frac{2^{-n} c(n,n)}{2^n (2\pi)^{pn/2} |\boldsymbol{\Sigma}|^{n/2}}, \end{aligned}$$

for

$$c(n,n) = 2^n \pi^{n^2/2} \Gamma_n(n/2) \quad \text{and} \quad \Gamma_r(m/2) = \pi^{r(r-1)/4} \prod_{i=1}^{r} \Gamma((m-i+1)/2).$$

Hence, the p.d.f. of $\boldsymbol{\ell}$ is given by

$$f_1(\boldsymbol{\ell}, \boldsymbol{\Sigma}) = c_n(\boldsymbol{\Sigma}) b(\boldsymbol{\ell}) \int_{\mathcal{H}_{n,p}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} a_{ii}(\boldsymbol{H}_1) \ell_i \right\} g_{n,p}(\boldsymbol{H}_1^t) \mathrm{d}\boldsymbol{H}_1,$$

where $\boldsymbol{A} = (a_{ij}) = \boldsymbol{H}_1^t \boldsymbol{\Sigma}^{-1} \boldsymbol{H}_1$ is an $n \times n$ matrix.

Next, we state a lemma stating the Stein-Haff identity for the singular Wishart matrix $\boldsymbol{W}$; a similar identity for $n > p$ has been obtained by Sheena (1995).

Lemma A.1 *Let $\boldsymbol{W}$ have a singular Wishart distribution $\mathcal{W}_p(\boldsymbol{\Sigma}, n)$, $\boldsymbol{\Sigma} > 0$, $n < p$, $\boldsymbol{W} = \boldsymbol{H}_1 \boldsymbol{L} \boldsymbol{H}_1^t$ and $\boldsymbol{\Phi}(\boldsymbol{\ell}) = \mathrm{diag}\,(\phi_1(\boldsymbol{\ell}), \ldots, \phi_n(\boldsymbol{\ell}))$, where $\boldsymbol{H}_1^t \boldsymbol{H}_1 = \boldsymbol{I}_n$, and $\boldsymbol{L}$ is the diagonal matrix with ordered non-zero eigenvalues of the matrix $\boldsymbol{W}$. Then the following identity holds:*

$$E\left[\mathrm{tr}\, \boldsymbol{H}_1 \boldsymbol{\Phi}(\boldsymbol{\ell}) \boldsymbol{H}_1^t \boldsymbol{\Sigma}^{-1}\right] = \sum_{i=1}^{n} E\left[(p-n-1) \frac{\phi_i}{\ell_i} + 2 \frac{\partial}{\partial \ell_i} \phi_i + 2 \sum_{j>i} \frac{\phi_i - \phi_j}{\ell_i - \ell_j} \right]. \tag{A.1}$$

Proof. We follow Sheena (1995) in proving this identity. Let $\ell_0 = \infty$, $\ell_{n+1} = 0$ and $\mathrm{d}\boldsymbol{L}_{(i)} = \prod_{j \neq i} \mathrm{d}\ell_j$, where the product term does not include the term $\mathrm{d}\ell_{(i)}$. Let $\mathcal{L}_{(i)}$ be the set defined by

$$\mathcal{L}_{(i)} = \{(\ell_1, \ldots, \ell_{i-1}, \ell_{i+1}, \ldots, \ell_n) | \ell_1 > \cdots > \ell_{i-1} > \ell_{i+1} > \cdots > \ell_n > 0\}.$$

Let $I = E[\mathrm{tr}\, \boldsymbol{H}_1 \boldsymbol{\Phi} \boldsymbol{H}_1^t \boldsymbol{\Sigma}^{-1}] = \sum_{i=1}^{n} E[\phi_i a_{ii}]$. Then, I is expressed as

$$\begin{aligned} I =& \sum_{i=1}^{n} \int_{\mathcal{L}_{(i)}} \int_{\ell_{i+1}}^{\ell_{i-1}} \phi_i b(\boldsymbol{\ell}) \left[\int_{\mathcal{H}_{n,p}} a_{ii} \exp\{-\frac{1}{2} \sum_{k=1}^{n} a_{kk} \ell_k\} g_{n,p}(\boldsymbol{H}_1^t) \mathrm{d}\boldsymbol{H}_1 \right] \mathrm{d}\ell_i \mathrm{d}\boldsymbol{L}_{(i)} \\ =& -2 \sum_{i=1}^{n} \int_{\mathcal{L}_{(i)}} \int_{\ell_{i+1}}^{\ell_{i-1}} \phi_i b(\boldsymbol{\ell}) \frac{\partial}{\partial \ell_i} \left[\int_{\mathcal{H}_{n,p}} \exp\{-\frac{1}{2} \sum_{k=1}^{n} a_{kk} \ell_k\} g_{n,p}(\boldsymbol{H}_1^t) \mathrm{d}\boldsymbol{H}_1 \right] \mathrm{d}\ell_i \mathrm{d}\boldsymbol{L}_{(i)}. \end{aligned}$$

Using integration by parts, we rewrite I as

$$I = 2\sum_{i=1}^{n}\int_{\mathcal{L}_{(i)}}\int_{\ell_{i+1}}^{\ell_{i-1}}\frac{\partial}{\partial \ell_i}\{\phi_i b(\boldsymbol{L})\}\int_{\mathcal{H}_{n,p}}\exp\{-\frac{1}{2}\sum_{k=1}^{n}a_{kk}\ell_k\}g_{n,p}(\boldsymbol{H}_1^t)\mathrm{d}\boldsymbol{H}_1\mathrm{d}\ell_i\mathrm{d}\boldsymbol{L}_{(i)},$$

which is equal to

$$2\sum_{i=1}^{n}E\left[\frac{1}{b(\boldsymbol{\ell})}\frac{\partial}{\partial \ell_i}\{\phi_i b(\boldsymbol{\ell})\}\right] = 2\sum_{i=1}^{n}E\left[\frac{\partial \phi_i}{\partial \ell_i} + \phi_i\frac{\partial}{\partial \ell_i}\log b(\boldsymbol{\ell})\right].$$

Since $\log b(\boldsymbol{\ell}) = \sum_{k=1}^{n}\{2^{-1}(p-n-1)\log \ell_k + \sum_{k<j}\log(\ell_k-\ell_j)\}$, it is noted that

$$\frac{\partial \log b(\boldsymbol{\ell})}{\partial \ell_i} = \frac{p-n-1}{2\ell_i} + \sum_{j\neq i}\frac{1}{\ell_i-\ell_j},$$

which implies that

$$I = \sum_{i=1}^{n}E\left[(p-n-1)\frac{\phi_i}{\ell_i} + 2\frac{\partial \phi_i}{\partial \ell_i} + 2\sum_{j\neq i}\frac{\phi_i}{\ell_i-\ell_j}\right]$$

This proves Lemma A.1 since $\sum_{i=1}^{n}\sum_{j\neq i}\phi_i/(\ell_i-\ell_j) = \sum_{i=1}^{n}\sum_{j>i}(\phi_i-\phi_j)/(\ell_i-\ell_j)$. ■

A.2 Proofs of the propositions

Proof of Proposition 2.3. It is noted that

$$\begin{aligned}\sum_{j=i+1}^{n}\frac{\psi_i-\psi_j}{\ell_i-\ell_j} =& \frac{1}{\ell_i}\sum_{j=i+1}^{n}(\psi_i-\psi_j) + \frac{1}{\ell_i}\sum_{j=i+1}^{n}\frac{\ell_j(\psi_i-\psi_j)}{\ell_i-\ell_j}\\ =& \frac{1}{\ell_i}\left\{(n-i)\psi_i - \sum_{j=i+1}^{n}\psi_j\right\} + \frac{1}{\ell_i}\sum_{j=i+1}^{n}\frac{\ell_j(\psi_i-\psi_j)}{\ell_i-\ell_j}.\end{aligned} \tag{A.2}$$

Then, the l.h.s. of the inequality in Proposition 2.2 is expressed by

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

which, from the conditions (a) and (b), can be seen to be less than or equal to

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

From the conditions (b) and (c), it is noted that $n+p-2i-1 \geq \psi_i \geq \psi_{i+1}$, so that $\psi_i^2 - 2(n+p-2i-1)\psi_i \leq \psi_{i+1}^2 - 2(n+p-2i-1)\psi_{i+1}$. Repeating this argument, we see that

$$
\begin{aligned}
&\psi_i^2 - 2(n+p-2i-1)\psi_i + 4\sum_{j>i}\psi_j \\
&\le \psi_{i+1}^2 - 2(n+p-2i-1)\psi_{i+1} + 4\psi_{i+1} + 4\sum_{j>i+1}\psi_j \\
&= \psi_{i+1}^2 - 2\{n+p-2(i+1)-1\}\psi_{i+1} + 4\sum_{j>i+1}\psi_j \\
&\le \cdots \le \psi_n^2 - 2(p-n-1)\psi_n = -(p-n-1)^2,
\end{aligned} \tag{A.3}
$$

which implies that

$$
\sum_{i=1}^n \frac{1}{\ell_i}\left\{\psi_i^2 - 2(n+p-2i-1)\psi_i + 4\sum_{j>i}\psi_j\right\} \le -\sum_{i=1}^n \frac{(p-n-1)^2}{\ell_i}.
$$

Hence, Proposition 2.3 is proved. ■

Proof of Proposition 2.5. The risk difference of the two estimators $\boldsymbol{\delta}^S$ and $\boldsymbol{\delta}^{IS}(g)$ is written as

$$
\begin{aligned}
\Delta =& R(\boldsymbol{\Sigma}, \boldsymbol{\delta}^{IS}(g)) - R(\boldsymbol{\Sigma}, \boldsymbol{\delta}_b^S) \\
=& \mathrm{tr}\, E\left[\frac{g(\boldsymbol{\ell})^2}{(\mathrm{tr}\,\boldsymbol{W})^2}\boldsymbol{W} + 2\frac{g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{W}}(\boldsymbol{\delta}^S - \boldsymbol{\Sigma}^{-1})\boldsymbol{W}\right] \\
=& E\left[\frac{g(\boldsymbol{\ell})^2}{\mathrm{tr}\,\boldsymbol{W}} + 2\frac{g(\boldsymbol{\ell})\mathrm{tr}\,\boldsymbol{D}}{\mathrm{tr}\,\boldsymbol{W}} - 2\frac{g(\boldsymbol{\ell})\mathrm{tr}\,\boldsymbol{W}\boldsymbol{\Sigma}^{-1}}{\mathrm{tr}\,\boldsymbol{W}}\right],
\end{aligned} \tag{A.4}
$$

since $\mathrm{tr}\,\boldsymbol{\delta}^S\boldsymbol{W} = \mathrm{tr}\,\boldsymbol{D}$. The Stein-Haff identity given in Lemma A.1 is used to evaluate $E[g(\boldsymbol{\ell})\mathrm{tr}\,\boldsymbol{W}\boldsymbol{\Sigma}^{-1}/\mathrm{tr}\,\boldsymbol{W}]$ as

$$
\begin{aligned}
&E\left[\mathrm{tr}\,\boldsymbol{H}_1\frac{g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{L}}\boldsymbol{L}\boldsymbol{H}_1^t\boldsymbol{\Sigma}^{-1}\right] \\
&= \sum_{i=1}^n E\left[(p-n-1)\frac{g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{L}} + 2\frac{g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{L}} - 2\frac{g(\boldsymbol{\ell})\ell_i}{(\mathrm{tr}\,\boldsymbol{L})^2} + 2\frac{\ell_i}{\mathrm{tr}\,\boldsymbol{L}}\frac{\partial g(\boldsymbol{\ell})}{\partial \ell_i} + 2\sum_{j>i}\frac{g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{L}}\right] \\
&= \sum_{i=1}^n E\left[(p+n-2i-1)\frac{g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{L}} + 2\frac{g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{L}} - 2\frac{g(\boldsymbol{\ell})\ell_i}{(\mathrm{tr}\,\boldsymbol{L})^2} + 2\frac{\ell_i}{\mathrm{tr}\,\boldsymbol{L}}\frac{\partial g(\boldsymbol{\ell})}{\partial \ell_i}\right] \\
&= E\left[\mathrm{tr}\,\boldsymbol{D}\frac{g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{L}} + 2(n-1)\frac{g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{L}} + 2\sum_{i=1}^n\frac{\ell_i}{\mathrm{tr}\,\boldsymbol{L}}\frac{\partial g(\boldsymbol{\ell})}{\partial \ell_i}\right].
\end{aligned} \tag{A.5}
$$

Combining (A.4) and (A.5) gives the expression

$$
\Delta = E\left[\frac{g(\boldsymbol{\ell})^2 - 4(n-1)g(\boldsymbol{\ell})}{\mathrm{tr}\,\boldsymbol{W}} - 4\sum_{i=1}^n\frac{\ell_i}{\mathrm{tr}\,\boldsymbol{L}}\frac{\partial g(\boldsymbol{\ell})}{\partial \ell_i}\right],
$$

which leads to the conditions given in Proposition 2.5 for the domination of $\boldsymbol{\delta}^{IS}(g)$ over $\boldsymbol{\delta}^S$. ■

Proof of Proposition 2.6. It is sufficient to check the conditions (a), (b) and (c) of Proposition 2.3 for

$$\psi_i^R = \ell_i \phi_i^R = \frac{d_i \ell_i}{\ell_i + (d_i \hat{\lambda}_{i+1} + 2\ell_i)/d_{i+1}}.$$

Since $\hat{\lambda}_{i+1}$ does not depend on ℓ_i, it is easy to see that ψ_i^R is increasing in ℓ_i. For the condition (b), we have that $\psi_n^R = d_n$ since $\hat{\lambda}_n = 0$. Also it is seen that the inequality $\psi_i \geq \psi_{i+1}$ is equivalent to the inequality

$$\hat{\lambda}_i \leq \frac{\ell_i}{d_{i+1}} \left(d_i - d_{i+1} + \frac{1}{\ell_{i+1}} d_i \hat{\lambda}_{i+1} \right) = \frac{1}{d_{i+1}} \left(2\ell_i + \frac{\ell_i}{\ell_{i+1}} d_i \hat{\lambda}_{i+1} \right).$$

Hence, the condition (b) follows from the definition of $\hat{\lambda}_i$ and the fact that $\ell_i/\ell_{i+1} > 1$. Finally, it is easily verified that $\phi_i^R \leq d_i$ for $i = 1, \ldots, n$. ■

Proof of Proposition 2.9. Since the estimator $\boldsymbol{\delta}^{DE}(g)$ belongs to the class of the estimators (2.1) as $\psi = a_0 + \ell_i^2 g(\boldsymbol{\ell})/\mathrm{tr}\, \boldsymbol{L}^2$, we can evaluate the condition in Proposition 2.8 as

$$\sum_{i=1}^n \left\{ \frac{\ell_i^2 g^2}{(\mathrm{tr}\, \boldsymbol{L}^2)^2} - 4(n-i)\frac{g}{\mathrm{tr}\, \boldsymbol{L}^2} - 8\frac{g}{\mathrm{tr}\, \boldsymbol{L}^2} + 8\frac{\ell_i^2 g}{(\mathrm{tr}\, \boldsymbol{L}^2)^2} - 4\frac{\ell_i(\partial g/\partial \ell_i)}{\mathrm{tr}\, \boldsymbol{L}^2} \right\} \leq 0,$$

or

$$\frac{1}{\mathrm{tr}\, \boldsymbol{L}^2} \left\{ g^2 - 2(n-1)(n+4)g - 4\sum_{i=1}^n \frac{\ell_i \partial g}{\partial \ell_i} \right\} \leq 0.$$

Hence, we get the conditions of Proposition 2.9 for the dominance of $\boldsymbol{\delta}^{DE}(g)$ over $\boldsymbol{\delta}_0$. ■

Proof of Proposition 2.10. It is noted that

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

where the equation (A.2) is used to get the third equation. Hence, the condition given in Proposition 2.8 is expressed by

$$\begin{aligned}
\sum_{i=1}^n \Big\{ & \frac{\psi_i^2 - 2a_0\psi_i - 4(n-i)\psi_i + 4\sum_{j>i}\psi_j + a_0^2}{\ell_i^2} + 4\sum_{j>i} \frac{\psi_j - a_0}{\ell_i \ell_j} \\
& - \frac{4}{\ell_i^2} \sum_{j>i} \frac{\ell_j(\psi_i - \psi_j)}{\ell_i - \ell_j} - \frac{4}{\ell_i} \frac{\partial \psi_i}{\partial \ell_i} \Big\} \leq 0.
\end{aligned} \tag{A.6}$$

From the condition (b), $\psi_i - a_0$ is nonnegative, and we observe that

$$\sum_i \sum_{j>i} \frac{\psi_j - a_0}{\ell_i \ell_j} \le \sum_i \sum_{j>i} \frac{\psi_j - a_0}{\ell_j^2} = \sum_i (i-1) \frac{\psi_i - a_0}{\ell_i^2}.$$

Using the conditions (a) and (b), we see that the inequality (A.6) holds if $h_i(\boldsymbol{\ell}) \le 0$, where

$$h_i(\boldsymbol{\ell}) = \psi_i^2 - 2(a_0 + 2n - 2i)\psi_i + 4\sum_{j>i} \psi_j + a_0^2 + 4(i-1)(\psi_i - a_0),$$

which can be rewritten by

$$h_i(\boldsymbol{\ell}) = (\psi_i - a_0)^2 - 4(n - 2i + 1)(\psi_i - a_0) + 4\sum_{j>i}(\psi_j - a_0). \tag{A.7}$$

To prove the inequality $h_i(\boldsymbol{\ell}) \le 0$, note that

$$2\max\{n - 2i + 1, 0\} \ge \psi_i - a_0 \ge \psi_{i+1} - a_0 \ge 0. \tag{A.8}$$

Then, (A.8) implies that $\psi_i - a_0 = 0$ for $i = r+1, \ldots, n$, so that it is easy to see that $h_i(\boldsymbol{\ell}) \le 0$ for $i = r+1, \ldots, n$. For $i = 1, \ldots, r-1$, from the inequalities in (A.8) and the same arguments as in the proof of Proposition 2.3, it follows that

$$\begin{aligned}
h_i(\boldsymbol{\ell}) \le& (\psi_{i+1} - a_0)^2 - 4(n - 2i + 1)(\psi_{i+1} - a_0) + 4(\psi_{i+1} - a_0) + 4\sum_{j>i+1}(\psi_j - a_0)\\
=& (\psi_{i+1} - a_0)^2 - 4(n - 2(i+1) + 1)(\psi_{i+1} - a_0) - 4(\psi_{i+1} - a_0) + 4\sum_{j>i+1}(\psi_j - a_0)\\
\le& (\psi_{i+1} - a_0)^2 - 4(n - 2(i+1) + 1)(\psi_{i+1} - a_0) + 4\sum_{j>i+1}(\psi_j - a_0)\\
\le& \cdots \le (\psi_{r-1} - a_0)^2 - 4(n - 2(r-1) + 1)(\psi_{r-1} - a_0) + 4(\psi_r - a_0)\\
\le& (\psi_r - a_0)^2 - 4(n - 2r + 1)(\psi_r - a_0) - 4(\psi_r - a_0),
\end{aligned}$$

which is not positive. Therefore we get Proposition 2.10. ■

Proof of Proposition 2.14. Since the estimator $\boldsymbol{\delta}^{EM*}(g)$ belongs to the class of the estimators (2.1) as $\psi = p + \ell_i g(\boldsymbol{\ell})/\mathrm{tr}\,\boldsymbol{L}$, we can evaluate the condition in Proposition 2.13 as

$$\begin{aligned}
\sum_{i=1}^{n} \Big\{ & \frac{\ell_i^2 g^2}{(\mathrm{tr}\,\boldsymbol{L})^2} + 2p\frac{\ell_i g}{\mathrm{tr}\,\boldsymbol{L}} - 2(p + n - 2i + 1)\frac{\ell_i g}{\mathrm{tr}\,\boldsymbol{L}}\\
& - 4\frac{\sum_{j>i}\ell_j g}{\mathrm{tr}\,\boldsymbol{L}} - 4\frac{\ell_i g}{\mathrm{tr}\,\boldsymbol{L}} + 4\frac{\ell_i^2 g}{(\mathrm{tr}\,\boldsymbol{L})^2} - 4\frac{\ell_i}{\mathrm{tr}\,\boldsymbol{L}}\frac{\partial g}{\partial \ell_i}\Big\} \le 0,
\end{aligned}$$

or

$$g^2 \frac{\mathrm{tr}\,\boldsymbol{L}^2}{(\mathrm{tr}\,\boldsymbol{L})^2} - 2(n+1)g + 4g\frac{\mathrm{tr}\,\boldsymbol{L}^2}{(\mathrm{tr}\,\boldsymbol{L})^2} - 4\sum_{i=1}^{n} \frac{\ell_i}{\mathrm{tr}\,\boldsymbol{L}}\frac{\partial g}{\partial \ell_i} \le 0.$$

Since $\mathrm{tr}\,\boldsymbol{L}^2 \leq (\mathrm{tr}\,\boldsymbol{L})^2$, we get the conditions given by Proposition 2.14 for $\boldsymbol{\delta}^{EM*}(g)$ to dominate $\boldsymbol{\delta}_2$ relative to the L_2-loss. ■

Proof of Proposition 3.1. We first consider the empirical Bayes estimator $\boldsymbol{\delta}_a^{EB}(\hat{\lambda}) = a(\boldsymbol{W} + \hat{\lambda}\boldsymbol{I}_p)^{-1}$ and the estimator

$$\boldsymbol{\delta}_a^{+}(\hat{\lambda}) = a\boldsymbol{H}_1(\boldsymbol{L} + \hat{\lambda}\boldsymbol{I}_n)^{-1}\boldsymbol{H}_1^t.$$

It is noted that both are $p \times p$ matricial estimators of $\boldsymbol{\Sigma}^{-1}$, $\boldsymbol{\delta}_a^{+}(\hat{\lambda})$ is singular while $\boldsymbol{\delta}_a^{EB}(\hat{\lambda})$ has the full rank p and is nonsingular. However, we will show that both estimators have the same risk under the L_2-loss function (2.4). Let

$$\boldsymbol{H} = (\boldsymbol{H}_1, \boldsymbol{H}_2)$$

where $\boldsymbol{H}_2$ is a $p \times (p-n)$ matrix belonging to $\mathcal{H}_{p-n,p}$, $\boldsymbol{H}_2^t\boldsymbol{H}_2 = \boldsymbol{I}_{p-n}$. Thus, $\boldsymbol{H}$ is an orthogonal matrix, and it is noted that

$$\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_1^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1 & \boldsymbol{H}_1^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_2 \\ \boldsymbol{H}_2^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1 & \boldsymbol{H}_2^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_2 \end{pmatrix}$$

and $(\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H})_{11} = \boldsymbol{H}_1^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_1$. Then we observe that

$$\begin{aligned}
\mathrm{tr}\,\{\boldsymbol{\delta}_a^{EB}(\hat{\lambda})\}^2\boldsymbol{W}^2 =& a^2\mathrm{tr}\,\boldsymbol{H}\begin{pmatrix} \boldsymbol{L} + \hat{\lambda}\boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\lambda}\boldsymbol{I}_{p-n} \end{pmatrix}^{-2}\boldsymbol{H}\boldsymbol{H}^t\begin{pmatrix} \boldsymbol{L}^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{H}^t \\
=& a^2\mathrm{tr}\,(\boldsymbol{L} + \hat{\lambda}\boldsymbol{I}_n)^{-2}\boldsymbol{L}^2 \\
=& \mathrm{tr}\,\{\boldsymbol{\delta}_a^{+}(\hat{\lambda})\}^2\boldsymbol{W}^2,
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{tr}\,\boldsymbol{W}^2\boldsymbol{\delta}_a^{EB}(\hat{\lambda})\boldsymbol{\Sigma}^{-1} =& a\mathrm{tr}\,\boldsymbol{H}\begin{pmatrix} \boldsymbol{L}^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\boldsymbol{H}\boldsymbol{H}^t\begin{pmatrix} \boldsymbol{L} + \hat{\lambda}\boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\lambda}\boldsymbol{I}_{p-n} \end{pmatrix}^{-1}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1} \\
=& a\mathrm{tr}\,\boldsymbol{L}^2(\boldsymbol{L} + \hat{\lambda}\boldsymbol{I}_n)^{-1}(\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H})_{11} \\
=& \mathrm{tr}\,\boldsymbol{W}^2\boldsymbol{\delta}_a^{+}(\hat{\lambda})\boldsymbol{\Sigma}^{-1}.
\end{aligned}$$

Thus, the two estimators $\boldsymbol{\delta}_a^{EB}(\hat{\lambda})$ and $\boldsymbol{\delta}_a^{+}(\hat{\lambda})$ have the same risk under the L_2-loss.

We next apply Proposition 2.12 to get the unbiased estimator of the risk difference of the estimators $\boldsymbol{\delta}_a^{+}(\hat{\lambda})$ and $\boldsymbol{\delta}_a$ where $\psi_i(\boldsymbol{\ell}) = a\ell_i/(\ell_i + \hat{\lambda}(\boldsymbol{\ell}))$ in the estimator (2.1). Then, we have

$$\begin{aligned}
\Delta_2(a,\hat{\lambda}) =& R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}_a^{EB}(\hat{\lambda})) - R_2(\boldsymbol{\Sigma}, \boldsymbol{\delta}_a) \\
=& \sum_{i=1}^n E\Big\{\frac{a^2\ell_i^2}{(\ell_i + \hat{\lambda})^2} - 2(p+n-2i+1)\frac{a\ell_i}{\ell_i + \hat{\lambda}} - 4\frac{a\ell_i}{\ell_i + \hat{\lambda}} \\
& - 4a\sum_{j>i}\frac{\ell_j\hat{\lambda}}{(\ell_i + \hat{\lambda})(\ell_j + \hat{\lambda})} - (a^2 - 2pa) + 4a\frac{\ell_i^2(1 + \partial\hat{\lambda}/\partial\ell_i)}{(\ell_i + \hat{\lambda})^2}\Big\},
\end{aligned}$$

which, since $\ell_i/(\ell_i+\hat\lambda) = 1-\hat\lambda/(\ell_i+\hat\lambda)$, can be rewritten as

$$a\sum_{i=1}^n E\Big\{(a+4)\frac{\hat\lambda^2}{(\ell_i+\hat\lambda)^2} - 2(a-p+n+1)\frac{\hat\lambda}{\ell_i+\hat\lambda} + 4\frac{\hat\lambda}{\ell_i+\hat\lambda}\sum_{j>i}\frac{\hat\lambda}{\ell_j+\hat\lambda} + 4\frac{\ell_i^2(\partial\hat\lambda/\partial\ell_i)}{(\ell_i+\hat\lambda)^2}\Big\}.$$

Using the equation

$$\Big\{\sum_i \frac{\hat\lambda}{\ell_i+\hat\lambda}\Big\}^2 = \sum_i \frac{\hat\lambda^2}{(\ell_i+\hat\lambda)^2} + 2\sum_i\sum_{j>i}\frac{\hat\lambda}{\ell_i+\hat\lambda}\frac{\hat\lambda}{\ell_j+\hat\lambda},$$

we can get the expression (3.5) of the risk difference. ■

Acknowledgments. The research of the first author was supported in part by grants from the Ministry of Education, Japan, Nos. 13680371, 15200021 and 15200022 and in part by a grant from the 21st Century COE Program at Faculty of Economics, University of Tokyo.

REFERENCES

Dey, D.K. (1987). Improved estimation of a multinormal precision matrix. *Statist. Probab. Letters*, **6**, 125-128.

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

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

Friedman, J.H. (1989). Regularized discriminant analysis. *J. Amer. Statist. Assoc.*, **84**, 165-175.

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

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

Kubokawa, T. (2005). A revisit to estimation of the precision matrix of the Wishart distribution. *J. Statist. Res.*, **39**, 91-114.

Kubokawa, T. and Srivastava, M.S. (1999). Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution. *Ann. Statist.*, **27**, 600-609.

Loh, W.L. (1997). Linear discrimination with adaptive ridge classification rules. *J. Multivariate Anal.*, **62**, 169-180.

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

Srivastava, M.S. (2003). Singular Wishart and multivariate beta distributions. *Ann. Statist.*, **31**, 1537-1560.

Srivastava, M.S. (2004). Mutivariate theory for analyzing high dimensional data. Tech. Report, University of Toronto.

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

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

Zhao, Y., Honda, M. and Konishi, S. (1996). Effect of a shrinkage estimator on the linear discriminant function. *American J. Math. Management Sci.*, **16**, 109-130.