CIRJE-F-378

Methods for Improvement in Estimation of a Normal Mean Matrix

Hisayuki Tsukuma
University of Tokyo

Tatsuya Kubokawa
University of Tokyo

September 2005

Methods for Improvement in Estimation of a Normal Mean Matrix

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

September 28, 2005

Abstract

This paper is concerned with the problem of estimating a matrix of means in multivariate normal distributions with an unknown covariance matrix under the quadratic loss function. It is first shown that the modified Efron-Morris estimator is characterized as certain empirical Bayes estimator. This estimator modifies the crude Efron-Morris estimator by adding a scalar shrinkage term. It is next shown that the idea of this modification provides the general method for improvement of estimators, which results in the further improvement of several minimax estimators including the Stein, Dey and Haff estimators. As a new method for improvement, a random combination of the modified Stein and the James-Stein estimators is also proposed and is shown to be minimax. Through Monte Carlo studies for the risk behaviors, it is numerically shown that the proposed, combined estimator inherits the nice risk properties of both individual estimators and thus it has a very favorable risk behavior in a small sample case.

Key words and phrases: Decision theory, empirical Bayes estimator, James-Stein estimator, MANOVA model, minimaxity, multivariate linear regression model, shrinkage estimation, simultaneous estimation.

1 Introduction

The estimation of a mean matrix of a multivariate normal distribution with a known covariance matrix has received theoretical interest in the literature since the seminal works of Efron and Morris (1972, 76) who extended the breakthrough of James and Stein (1961) to the multivariate setup. Especially, Efron and Morris (1976) showed not only that a matricial shrinkage estimator can be characterized as an empirical Bayes estimator, but also that the matricial shrinkage estimator can be further improved on by the modification of adding a scalar shrinkage term. Another important finding in their paper is that the estimation of the mean matrix is connected to that of a covariance

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

or precision matrix, which implies that the methods used for estimating the covariance matrix produce the corresponding minimax estimators of the mean matrix. For the recent development from the aspect of admissibility, one can refer to Berger, Strawderman and Tang (2005). Although the results for the known covariance matrix are of theoretical interest, their extensions to the case of an unknown covariance matrix are important from the practical aspect, because the mean matrix corresponds to regression coefficients in a multivariate linear regression model and small area means in a multivariate mixed linear model. Using the technique of the unbiased estimate of risk, Bilodeau and Kariya (1989) and Konno (1990, 1991, 1992) extended the above minimaxity results to the case of the unknown covariance matrix. In this paper, we point out that the estimators given in the previous studies have a room for the improvement, and construct new types of minimax estimators with favorable risk behaviors.

To explain the subjects addressed in the paper, we begin with describing the model and the estimation problem. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)^t$ be an $m \times p$ random matrix, where the row vectors are mutually independent and the i-th row vector $\boldsymbol{x}_i$ has a multivariate normal distribution with mean vector $\boldsymbol{\theta}_i$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. Also, let $\boldsymbol{S}$ be a $p \times p$ random matrix having the Wishart distribution with degrees of freedom n and mean $n\boldsymbol{\Sigma}$. These are abbreviated to

$$\boldsymbol{X} \sim \mathcal{N}_{m\times p}(\boldsymbol{\Theta}, \boldsymbol{I}_m \otimes \boldsymbol{\Sigma}) \quad \text{and} \quad \boldsymbol{S} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma}), \tag{1.1}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m)^t$. It is assumed that $\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}$ are unknown and that $\boldsymbol{X}$ and $\boldsymbol{S}$ are mutually independent. This is a canonical form of a multivariate linear regression model. Our aim is to construct an estimator of the mean matrix $\boldsymbol{\Theta}$ on the basis of $\boldsymbol{X}$ and $\boldsymbol{S}$ relative to the quadratic loss function

$$L(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}}) = \mathrm{tr}\,(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})^t. \tag{1.2}$$

Every estimator is evaluated by the risk function $R(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}}) = E[L(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}})]$.

The maximum likelihood estimator of $\boldsymbol{\Theta}$ is $\widehat{\boldsymbol{\Theta}}^{ML} = \boldsymbol{X}$, which is a minimax estimator with the constant risk mp. One of estimators improving on $\boldsymbol{X}$ is the Efron-Morris estimator

$$\widehat{\boldsymbol{\Theta}}^{EM} = \begin{cases} \boldsymbol{X}\left\{\boldsymbol{I}_p - \alpha(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{S}\right\} & \text{if } m \geq p+2, \\ \left\{\boldsymbol{I}_m - \alpha(\boldsymbol{X}^t\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\right\}\boldsymbol{X} & \text{if } p \geq m+2, \end{cases} \tag{1.3}$$

where $\alpha = \{|m-p|-1\}/\{n + (2m-p)\wedge p + 1\}$ with $a \wedge b = \min(a, b)$. Konno (1991, 92) showed the minimaxity of $\widehat{\boldsymbol{\Theta}}^{EM}$ and obtained the further dominance result that $\widehat{\boldsymbol{\Theta}}^{EM}$ can be improved on by the modified Efron-Morris estimator

$$\widehat{\boldsymbol{\Theta}}^{MEM} = \widehat{\boldsymbol{\Theta}}^{EM} - \frac{\beta}{\mathrm{tr}\,\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1}}\boldsymbol{X} \tag{1.4}$$

for a nonnegative constant β. This procedure modifies the matricial shrinkage estimator by adding the scalar shrinkage term $-(\beta/\mathrm{tr}\,\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1})\boldsymbol{X}$, and this modification yields the further improvement. The following queries are here raised:

(a) The modified Efron-Morris estimator $\widehat{\boldsymbol{\Theta}}^{MEM}$ consists of two kinds of shrinkage terms: matricial shrinkage and scalar shrinkage. Can this modified Efron-Morris estimator be characterized as an empirical Bayes procedure ? If so, adding the scalar shrinkage term may be considered as a natural modification.

(b) Can the modification rule of adding the scalar shrinkage term be established as the general method for improving estimators ?

The first objective of this paper is to address the problems of resolving these queries. Section 2 handles the query (a) in a Bayesian framework. As prior distributions, it is assumed that $\mathbf{\Theta}$ has a multivariate normal distribution and that $\mathbf{\Sigma}^{-1}$ has a multivariate F-distribution in a setup similar to Kiefer and Schwartz (1965). It is shown that the modified Efron-Morris estimator can be derived as an empirical Bayes estimator under the setup.

Section 3 is concerned with the query (b). To explain the derived results, let $\boldsymbol{F} = \mathrm{diag}\,(f_1, \dots, f_{m\wedge p})$ be a diagonal matrix based on the eigenvalues $f_1 \geq \cdots \geq f_{m\wedge p} \geq 0$ such that for an $m \times m$ orthogonal matrix $\boldsymbol{R}$ and a $p \times p$ nonsingular matrix $\boldsymbol{Q}$,

$$\begin{cases} \boldsymbol{Q}^t\boldsymbol{S}\boldsymbol{Q} = \boldsymbol{I}_p \quad \text{and} \quad \boldsymbol{Q}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{Q} = \boldsymbol{F} & \text{if } m \geq p, \\ \boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t = \boldsymbol{R}\boldsymbol{F}\boldsymbol{R}^t & \text{if } m < p. \end{cases}$$

Then, we consider the general class of the shrinkage estimator

$$\widehat{\mathbf{\Theta}}(\mathbf{\Psi}) = \begin{cases} \boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{Q}\mathbf{\Psi}(\boldsymbol{F})\boldsymbol{Q}^{-1}) & \text{if } m \geq p, \\ (\boldsymbol{I}_m - \boldsymbol{R}\mathbf{\Psi}(\boldsymbol{F})\boldsymbol{R}^t)\boldsymbol{X} & \text{if } m < p, \end{cases} \tag{1.5}$$

which is equivariant under a transformation group, where $\mathbf{\Psi}(\boldsymbol{F})$ is a $(m \wedge p) \times (m \wedge p)$ diagonal matrix whose elements are functions of $\boldsymbol{F}$. Using the same idea as in the modified Efron-Morris estimator $\widehat{\mathbf{\Theta}}^{MEM}$, we consider to modify $\widehat{\mathbf{\Theta}}(\mathbf{\Psi})$ as

$$\widehat{\mathbf{\Theta}}^{M} = \widehat{\mathbf{\Theta}}(\mathbf{\Psi}) - \frac{\beta}{\mathrm{tr}\,\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1}}\boldsymbol{X}.$$

In Section 3, we obtain the general conditions on β and $\mathbf{\Psi}$ for the estimator $\widehat{\mathbf{\Theta}}^{M}$ to dominate $\widehat{\mathbf{\Theta}}(\mathbf{\Psi})$. This provides a unified method for improving estimators. Two simple applications are the minimaxity of the James-Stein estimator $\widehat{\mathbf{\Theta}}^{JS} = \boldsymbol{X} - (\beta/\mathrm{tr}\,\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1})\boldsymbol{X}$ and the domination of $\widehat{\mathbf{\Theta}}^{MEM}$ over $\widehat{\mathbf{\Theta}}^{EM}$. Another interesting example is to modify the Stein estimator $\widehat{\mathbf{\Theta}}^{ST}$ given by Konno (1991, 92) and it is shown that the modified Stein estimator $\widehat{\mathbf{\Theta}}^{MST} = \widehat{\mathbf{\Theta}}^{ST} - (\beta/\mathrm{tr}\,\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1})\boldsymbol{X}$ dominates $\widehat{\mathbf{\Theta}}^{ST}$ under a condition on β. The method can be also applied to get the improvements on the estimators motivated from Haff (1980) and Dey (1987).

Section 4 handles the method of combining the James-Stein estimator $\widehat{\mathbf{\Theta}}^{JS}$ and the modified Stein estimator $\widehat{\mathbf{\Theta}}^{MST}$. The former estimator is known to give the significant improvement near $\mathbf{\Theta} = \mathbf{0}$, while the latter is better than the former when $\mathbf{\Theta}$ is far away from zero. We want to choose the weighting function $\varepsilon = \varepsilon(\boldsymbol{F})$ such that the combined estimator $\widehat{\mathbf{\Theta}}^{CM} = \varepsilon\widehat{\mathbf{\Theta}}^{JS} + (1-\varepsilon)\widehat{\mathbf{\Theta}}^{MST}$ inherits these nice risk properties of the individual estimators. For this aim, it is reasonable to take a test statistic for testing the null hypothesis $H_0 : \mathbf{\Theta} = \mathbf{0}$ against $H_1 : \mathbf{\Theta} \neq \mathbf{0}$. Since the likelihood ratio statistic is of the form $\exp\{-n\mathrm{tr}\,\boldsymbol{F}/2\}$, a good choice of ε may be $\varepsilon = \exp\{-\gamma\mathrm{tr}\,\boldsymbol{F}\}$ for a positive constant γ. Although it is very hard to establish the minimaxity of the combined estimator $\widehat{\boldsymbol{\Theta}}^{CM}$, in Section 4, we succeed in deriving a condition for the minimaxity.

Monte Carlo simulation studies for comparing the estimators derived in this paper are provided in Section 5 in the case of $m > p$. The competitors include the modified shrinkage estimators given in Section 3, the combined estimator $\widehat{\boldsymbol{\Theta}}^{CM}$ given in Section 4 and an empirical Bayes estimator recommended by Shieh (1993). The Monte Carlo studies report that the combined estimator $\widehat{\boldsymbol{\Theta}}^{CM}$ has an excellent risk behavior such that $\widehat{\boldsymbol{\Theta}}^{CM}$ inherits the nice risk properties of both the estimators $\widehat{\boldsymbol{\Theta}}^{JS}$ and $\widehat{\boldsymbol{\Theta}}^{MST}$.

Finally, it is noted that the proofs in this paper shall be done in the case of $m \geq p$ since the proofs for $m < p$ can be given by replacing (n, m, p) with $(n + m - p, p, m)$ in the proof for $m \geq p$.

2 Empirical Bayes methods

We consider an empirical Bayes estimation of the normal mean matrix in the model (1.1) and show that the resulting empirical Bayes estimators correspond to the Efron-Morris and its modified estimators.

2.1 Case of $m \geq p$

We first treat the case of $m \geq p$. Assume that the prior distribution of $\boldsymbol{\Theta}$ is distributed as

$$\boldsymbol{\Theta} \sim \mathcal{N}_{m\times p}(\boldsymbol{0}, \boldsymbol{I}_m \otimes \boldsymbol{A}),$$

where $\boldsymbol{A}$ is an unknown $p \times p$ matrix. Also assume that $\boldsymbol{\Sigma}$ has a prior distribution, which will be specified later. Then, given $\boldsymbol{\Sigma}$, the posterior distribution of $\boldsymbol{\Theta}$ and the marginal distribution of $\boldsymbol{X}$ are, respectively, given by

$$\begin{aligned} \boldsymbol{\Theta}|\boldsymbol{X}, \boldsymbol{\Sigma} &\sim \mathcal{N}_{m\times p}(\boldsymbol{X}(\boldsymbol{I}_p - \Xi), \boldsymbol{I}_m \otimes (\boldsymbol{\Sigma}^{-1} + \boldsymbol{A}^{-1})^{-1}), \\ \boldsymbol{X}|\boldsymbol{\Sigma} &\sim \mathcal{N}_{m\times p}(\boldsymbol{0}, \boldsymbol{I}_m \otimes (\boldsymbol{\Sigma} + \boldsymbol{A})), \end{aligned}$$

where $\Xi = (\boldsymbol{\Sigma} + \boldsymbol{A})^{-1}\boldsymbol{\Sigma}$. The Bayes estimator is the posterior mean $\widehat{\boldsymbol{\Theta}}^B = \boldsymbol{X}(\boldsymbol{I}_p - \Xi)$. Since the ratio of covariance matrices Ξ is unknown, it may be estimated from the marginal distributions of $\boldsymbol{S}$ and $\boldsymbol{W} = \boldsymbol{X}^t\boldsymbol{X}$, respectively, given by

$$\boldsymbol{S}|\boldsymbol{\Sigma} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma}) \quad \text{and} \quad \boldsymbol{W}|\boldsymbol{\Sigma} \sim \mathcal{W}_p(m, \boldsymbol{\Sigma}_2), \tag{2.1}$$

for $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma} + \boldsymbol{A}$. It is noted that the parameter space is restricted by $\boldsymbol{\Sigma}_2 > \boldsymbol{\Sigma}$ or $\Xi < \boldsymbol{I}_p$. When Ξ is estimated by a function of $\boldsymbol{S}$ and $\boldsymbol{W}$, denoted by $\widehat{\Xi}$, substituting $\widehat{\Xi}$ into $\widehat{\boldsymbol{\Theta}}^B$ results in an empirical Bayes estimator of the form $\widehat{\boldsymbol{\Theta}}^{EB} = \boldsymbol{X}(\boldsymbol{I}_p - \widehat{\Xi})$.

The expected risk of the empirical Bayes estimator $\widehat{\boldsymbol{\Theta}}^{EB} = \boldsymbol{X}(\boldsymbol{I}_p - \widehat{\Xi})$ is written as

$$\begin{aligned} E_{\boldsymbol{\Theta},\boldsymbol{\Sigma}}[R(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}}^{EB})] &= E_{\boldsymbol{\Theta},\boldsymbol{\Sigma}}[E_{\boldsymbol{X},\boldsymbol{S}|\boldsymbol{\Theta},\boldsymbol{\Sigma}}[\mathrm{tr}\,(\widehat{\boldsymbol{\Theta}}^{EB} - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Theta}}^{EB} - \boldsymbol{\Theta})^t]] \\ &= E_{\boldsymbol{X},\boldsymbol{S},\boldsymbol{\Sigma}}[E_{\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{S},\boldsymbol{\Sigma}}[\mathrm{tr}\,(\widehat{\boldsymbol{\Theta}}^{EB} - \widehat{\boldsymbol{\Theta}}^B)\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Theta}}^{EB} - \widehat{\boldsymbol{\Theta}}^B)^t]] \\ &\quad + E_{\boldsymbol{X},\boldsymbol{S},\boldsymbol{\Sigma}}[E_{\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{S},\boldsymbol{\Sigma}}[\mathrm{tr}\,(\widehat{\boldsymbol{\Theta}}^B - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Theta}}^B - \boldsymbol{\Theta})^t]], \end{aligned}$$

each term of which can be evaluated as

$$E_{\boldsymbol{X},\boldsymbol{S},\boldsymbol{\Sigma}}[E_{\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{S},\boldsymbol{\Sigma}}[\mathrm{tr}\,(\widehat{\boldsymbol{\Theta}}^B-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Theta}}^B-\boldsymbol{\Theta})^t]]=E_{\boldsymbol{\Sigma}}[m\mathrm{tr}\,(\boldsymbol{I}_p-\Xi)]$$

and

$$\begin{aligned}&E_{\boldsymbol{X},\boldsymbol{S},\boldsymbol{\Sigma}}[E_{\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{S},\boldsymbol{\Sigma}}[\mathrm{tr}\,(\widehat{\boldsymbol{\Theta}}^{EB}-\widehat{\boldsymbol{\Theta}}^B)\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Theta}}^{EB}-\widehat{\boldsymbol{\Theta}}^B)^t]]\\&=E_{\boldsymbol{X},\boldsymbol{S},\boldsymbol{\Sigma}}[\mathrm{tr}\,\boldsymbol{X}^t\boldsymbol{X}(\widehat{\Xi}-\Xi)\boldsymbol{\Sigma}^{-1}(\widehat{\Xi}-\Xi)^t].\end{aligned}$$

Since $\widehat{\Xi}$ is a function of $\boldsymbol{S}$ and $\boldsymbol{W}=\boldsymbol{X}^t\boldsymbol{X}$, we observe that

$$E_{\boldsymbol{X},\boldsymbol{S},\boldsymbol{\Sigma}}[\mathrm{tr}\,\boldsymbol{X}^t\boldsymbol{X}(\widehat{\Xi}-\Xi)\boldsymbol{\Sigma}^{-1}(\widehat{\Xi}-\Xi)^t]=E_{\boldsymbol{\Sigma}}[E_{\boldsymbol{W},\boldsymbol{S}|\boldsymbol{\Sigma}}[\mathrm{tr}\,\boldsymbol{W}(\widehat{\Xi}-\Xi)\boldsymbol{\Sigma}^{-1}(\widehat{\Xi}-\Xi)^t]].$$

Thus, the expected risk is expressed by

$$E_{\boldsymbol{\Theta},\boldsymbol{\Sigma}}[R(\boldsymbol{\Theta},\widehat{\boldsymbol{\Theta}}^{EB})]=E_{\boldsymbol{\Sigma}}\left[m\mathrm{tr}\,(\boldsymbol{I}_p-\Xi)+E_{\boldsymbol{W},\boldsymbol{S}|\boldsymbol{\Sigma}}[\mathrm{tr}\,\boldsymbol{W}(\widehat{\Xi}-\Xi)\boldsymbol{\Sigma}^{-1}(\widehat{\Xi}-\Xi)^t]\right].$$

Since $(\boldsymbol{S},\boldsymbol{W})$ is a complete statistic for $(\boldsymbol{\Sigma},\Xi)$, the same argument as in Efron and Morris (1976) can be used to get the expression

$$R(\boldsymbol{\Theta},\widehat{\boldsymbol{\Theta}}^{EB})=m\mathrm{tr}\,(\boldsymbol{I}_p-\Xi)+E_{\boldsymbol{W},\boldsymbol{S}|\boldsymbol{\Sigma}}[\mathrm{tr}\,\boldsymbol{W}(\widehat{\Xi}-\Xi)\boldsymbol{\Sigma}^{-1}(\widehat{\Xi}-\Xi)^t] \tag{2.2}$$

and the expected risk

$$E_{\boldsymbol{\Theta},\boldsymbol{\Sigma}}[R(\boldsymbol{\Theta},\widehat{\boldsymbol{\Theta}}^{EB})]=E_{\boldsymbol{\Theta},\boldsymbol{\Sigma}}\left[m\mathrm{tr}\,(\boldsymbol{I}_p-\Xi)+E_{\boldsymbol{W},\boldsymbol{S}|\boldsymbol{\Sigma}}[\mathrm{tr}\,\boldsymbol{W}(\widehat{\Xi}-\Xi)\boldsymbol{\Sigma}^{-1}(\widehat{\Xi}-\Xi)^t]\right].$$

This implies that the problem of estimating the mean matrix $\boldsymbol{\Theta}$ by using an estimator $\widehat{\boldsymbol{\Theta}}^{EB}=\boldsymbol{X}(\boldsymbol{I}_p-\widehat{\Xi})$ is reduced to that of estimating Ξ relative to the loss function $\mathrm{tr}\,\boldsymbol{W}(\widehat{\Xi}-\Xi)\boldsymbol{\Sigma}^{-1}(\widehat{\Xi}-\Xi)^t$ for $\Xi=\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}$ under the model (2.1). This estimation problem is similar to that considered by Loh (1988, 1991).

It is reasonable to estimate Ξ by an estimator of the form $\alpha\boldsymbol{W}^{-1}\boldsymbol{S}$ for a positive constant α. Then the best α in terms of minimizing the risk $R(\boldsymbol{\Theta},\boldsymbol{X}(\boldsymbol{I}-\alpha(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{S}))$ is given by $\alpha=(m-p-1)/(n+p+1)$. Replacing Ξ with the estimator $\{(m-p-1)/(n+p+1)\}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{S}$, we obtain the empirical Bayes estimator

$$\widehat{\boldsymbol{\Theta}}^{EM}=\boldsymbol{X}\Big\{\boldsymbol{I}_p-\frac{m-p-1}{n+p+1}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{S}\Big\},$$

which is called the Efron-Morris estimator. Konno (1992) showed that the Efron-Morris estimator $\widehat{\boldsymbol{\Theta}}^{EM}$ is better than $\widehat{\boldsymbol{\Theta}}^{ML}=\boldsymbol{X}$ relative to the loss (1.2), that is, $\widehat{\boldsymbol{\Theta}}^{EM}$ is minimax.

It is interesting to show that the modified Efron-Morris estimator $\widehat{\boldsymbol{\Theta}}^{MEM}$ given by (1.4) can be derived as an empirical Bayes estimator. For the purpose, we begin with the model (2.1) and use the prior distribution similar to Kiefer and Schwartz (1965). Assume that

$$\boldsymbol{\Sigma}_2^{-1}=\lambda\boldsymbol{\Sigma}^{-1}+\boldsymbol{C},$$

where λ is a scalar and $\boldsymbol{C}$ is a $p \times p$ positive definite matrix. Then the joint density of $(\boldsymbol{W}, \boldsymbol{S})$ is proportional to

$$
\begin{aligned}
p(\boldsymbol{W}, \boldsymbol{S}|\boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{W}|^{(m-p-1)/2}|\boldsymbol{S}|^{(n-p-1)/2}|\lambda\boldsymbol{\Sigma}^{-1} + \boldsymbol{C}|^{m/2}|\boldsymbol{\Sigma}^{-1}|^{n/2} \\
\times \exp\left\{-\mathrm{tr}\,(\lambda\boldsymbol{W} + \boldsymbol{S})\boldsymbol{\Sigma}^{-1}/2 - \mathrm{tr}\,\boldsymbol{W}\boldsymbol{C}/2\right\}.
\end{aligned}
$$

Assume that the prior distribution of $\boldsymbol{\Sigma}^{-1}$ has a multivariate F-distribution whose density has the form

$$
p(\boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}^{-1}|^{d/2}|\lambda\boldsymbol{\Sigma}^{-1} + \boldsymbol{C}|^{-m/2}
$$

for $d > -2$ and $m > d + 2p$. Then the posterior distribution of $\boldsymbol{\Sigma}^{-1}$ can be expressed as

$$
p(\boldsymbol{\Sigma}^{-1}|\boldsymbol{W}, \boldsymbol{S}) \propto |\boldsymbol{\Sigma}^{-1}|^{(n+d)/2} \exp\left\{-\mathrm{tr}\,(\lambda\boldsymbol{W} + \boldsymbol{S})\boldsymbol{\Sigma}^{-1}/2\right\},
$$

that is, the Wishart distribution $\mathcal{W}_p(n + d + p + 1, (\lambda\boldsymbol{W} + \boldsymbol{S})^{-1})$. Since the posterior distribution $\boldsymbol{\Sigma}$ is the inverse Wishart distribution $\mathcal{W}_p^{-1}(n + d + 2p + 2, \lambda\boldsymbol{W} + \boldsymbol{S})$, the posterior mean of $\boldsymbol{\Xi} = \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}$ is

$$
E[\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}|\boldsymbol{W}, \boldsymbol{S}, \lambda, \boldsymbol{C}] = \lambda\boldsymbol{I}_p + \boldsymbol{C}E[\boldsymbol{\Sigma}|\boldsymbol{W}, \boldsymbol{S}] = \lambda\boldsymbol{I}_p + a_0\boldsymbol{C}(\lambda\boldsymbol{W} + \boldsymbol{S}),
$$

where $a_0 = 1/(n + d)$. We need to derive estimators of the hyperparameters λ and $\boldsymbol{C}$ from the marginal distribution of $(\boldsymbol{W}, \boldsymbol{S})$, given by

$$
\begin{aligned}
p(\boldsymbol{W}, \boldsymbol{S}|\lambda, \boldsymbol{C}) \propto \lambda^{p(d+p+1)/2}|\boldsymbol{C}|^{m/2-(d+p+1)/2} \\
\times |\boldsymbol{W}|^{(m-p-1)/2}|\boldsymbol{S}|^{(n-p-1)/2}|\lambda\boldsymbol{W} + \boldsymbol{S}|^{-(n+d+p+1)/2} \exp(-\mathrm{tr}\,\boldsymbol{W}\boldsymbol{C}/2).
\end{aligned}
$$

From the marginal distribution, the covariance matrix $\boldsymbol{C}$ may be estimated by $a_1\boldsymbol{W}^{-1}$ for a constant a_1. Using the first order approximation of the marginal likelihood function as used in Haff (1980, page 589), we may estimate λ by the form $\hat{\lambda} = a_2/\mathrm{tr}\,\boldsymbol{W}\boldsymbol{S}^{-1}$ for a constant a_2. Thus, $\boldsymbol{\Xi} = \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}$ can be estimated by

$$
\begin{aligned}
\widehat{\boldsymbol{\Xi}}^{EB} =& \hat{\lambda}\boldsymbol{I}_p + a_0\widehat{\boldsymbol{C}}(\hat{\lambda}\boldsymbol{W} + \boldsymbol{S}) \\
=& \alpha\boldsymbol{W}^{-1}\boldsymbol{S} + \frac{\beta}{\mathrm{tr}\,\boldsymbol{W}\boldsymbol{S}^{-1}}\boldsymbol{I}_p
\end{aligned}
$$

for positive constants α and β. The resulting empirical Bayes estimator of $\boldsymbol{\Theta}$ is

$$
\widehat{\boldsymbol{\Theta}} = \boldsymbol{X}\left\{\boldsymbol{I}_p - \alpha(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{S} - \frac{\beta}{\mathrm{tr}\,\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1}}\boldsymbol{I}_p\right\}.
$$

The best α in terms of minimizing the risk function is given by $\alpha = (m-p-1)/(n+p+1)$. Then, the empirical Bayes estimator of $\boldsymbol{\Theta}$ is

$$
\widehat{\boldsymbol{\Theta}}^{MEM}(\beta) = \widehat{\boldsymbol{\Theta}}^{EM} - \frac{\beta}{\mathrm{tr}\,\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1}}\boldsymbol{X},
$$

which is the modified Efron-Morris estimator.

2.2 Case of $m < p$

We next handle the case of $m < p$. Assume that the prior distribution of $\boldsymbol{\Theta}$ is

$$\boldsymbol{\Theta} \sim \mathcal{N}_{m\times p}(\boldsymbol{0}, \boldsymbol{B} \otimes \boldsymbol{\Sigma}),$$

where $\boldsymbol{B}$ is an $m \times m$ unknown positive definite matrix. Then, the posterior distribution of $\boldsymbol{\Theta}$ and the marginal distribution of $\boldsymbol{X}$ are, respectively, given as

$$\begin{array}{ll} \boldsymbol{\Theta}|\boldsymbol{X} & \sim \mathcal{N}_{m\times p}((\boldsymbol{I}_m - \boldsymbol{\Xi})\boldsymbol{X}, (\boldsymbol{I}_m - \boldsymbol{\Xi}) \otimes \boldsymbol{\Sigma}), \\ \boldsymbol{X} & \sim \mathcal{N}_{m\times p}(\boldsymbol{0}, \boldsymbol{\Xi}^{-1} \otimes \boldsymbol{\Sigma}), \end{array}$$

where $\boldsymbol{\Xi} = (\boldsymbol{I}_m + \boldsymbol{B})^{-1}$. The Bayes estimator is thus given by $\widehat{\boldsymbol{\Theta}}^B = (\boldsymbol{I}_m - \boldsymbol{\Xi})\boldsymbol{X}$. Since $\boldsymbol{\Xi}$ is unknown, we need to estimate it. For the purpose, we concentrate our attention on the distribution of $\boldsymbol{V} = (\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t)^{-1}$. It is noted that $\boldsymbol{X}$ and $\boldsymbol{S}$ are marginally distributed as $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Xi}^{-1} \otimes \boldsymbol{\Sigma})$ and $\boldsymbol{S} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$. Combining Theorems 4.2.1, 5.3.22, 5.3.6, and the equation (1.3.5) of Gupta and Nagar (1999), we can see that the density of $\boldsymbol{V} = (\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t)^{-1}$ is written by

$$p(\boldsymbol{V}|\boldsymbol{\Xi}) \propto |\boldsymbol{\Xi}|^{p/2}|\boldsymbol{\Xi} + \boldsymbol{V}|^{-(n+m)/2}|\boldsymbol{V}|^{(n-p-1)/2}.$$

From Theorem 5.3.20 of Gupta and Nagar (1999), it follows that $E[\boldsymbol{V}] = (n-p+m)\boldsymbol{\Xi}/(p-m-1)$. Hence, an empirical Bayes estimator of $\boldsymbol{\Theta}$ is

$$\widehat{\boldsymbol{\Theta}}^{EM}(\alpha) = \left\{\boldsymbol{I}_m - \alpha(\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t)^{-1}\right\}\boldsymbol{X},$$

where α is a constant. The best α is $\alpha_0 = (p - m - 1)/(n + 2m - p + 1)$, and we call $\widehat{\boldsymbol{\Theta}}^{EM} = \widehat{\boldsymbol{\Theta}}^{EM}(\alpha_0)$ the Efron-Morris estimator.

It is more interesting to characterize the modified Efron-Morris estimator through the empirical method for $m < p$. Let

$$\boldsymbol{\Xi} = \lambda\boldsymbol{I}_m + \boldsymbol{C},$$

where λ is a scalar and $\boldsymbol{C}$ is an $m \times m$ positive definite matrix. Then, the marginal density of $\boldsymbol{V}$ is

$$p(\boldsymbol{V}|\lambda, \boldsymbol{C}) \propto |\lambda\boldsymbol{I}_m + \boldsymbol{C}|^{p/2}|\lambda\boldsymbol{I}_m + \boldsymbol{C} + \boldsymbol{V}|^{-(n+m)/2}|\boldsymbol{V}|^{(n-p-1)/2}.$$

Assuming that the prior distribution of $\boldsymbol{C}$ has the density

$$p(\boldsymbol{C}|\lambda) \propto \lambda^{m(p-m-1-b_0)/2}|\boldsymbol{C}|^{b_0/2}|\lambda\boldsymbol{I}_m + \boldsymbol{C}|^{-p/2}$$

for a constant b_0, we have the posterior distribution of $\boldsymbol{C}$ as

$$p(\boldsymbol{C}|\lambda, \boldsymbol{V}) \propto |\boldsymbol{C}|^{b_0/2}|\lambda\boldsymbol{I}_m + \boldsymbol{C} + \boldsymbol{V}|^{-(n+m)/2},$$

and the posterior mean of $\boldsymbol{\Xi}$ as

$$E[\boldsymbol{\Xi}|\lambda, \boldsymbol{V}] = \lambda\boldsymbol{I}_m + E[\boldsymbol{C}|\boldsymbol{V}, \lambda] = (1 + b_1)\lambda\boldsymbol{I}_m + b_1\boldsymbol{V}$$

for $b_1 = (b_0 + m + 1)/(n - m - 2 - b_0)$. Since the marginal distribution of $\boldsymbol{V}$ is written by

$$p(\boldsymbol{V}|\lambda) \propto \lambda^{m(p-m-1-b_0)/2} |\boldsymbol{V}|^{(n-p-1)/2} |\lambda \boldsymbol{I}_m + \boldsymbol{V}|^{-(n-b_0-1)/2},$$

using the arguments as in Haff (1980) provides a reasonable estimator of λ, given by $\hat{\lambda} = b_2/\mathrm{tr}\, \boldsymbol{V}^{-1}$ for a constant b_2. Thus, $\boldsymbol{\Xi}$ can be estimated by

$$\begin{aligned}\widehat{\boldsymbol{\Xi}} =& (b_2(1+b_1)/\mathrm{tr}\, \boldsymbol{V}^{-1})\boldsymbol{I}_m + b_1 \boldsymbol{V} \\ =& \alpha(\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t)^{-1} + \frac{\beta}{\mathrm{tr}\, \boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t} \boldsymbol{I}_m\end{aligned}$$

for constants α and β. The resulting empirical Bayes estimator of $\boldsymbol{\Theta}$ is

$$\widehat{\boldsymbol{\Theta}} = \left\{ \boldsymbol{I}_m - \alpha(\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t)^{-1} - \frac{\beta}{\mathrm{tr}\, \boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t} \boldsymbol{I}_m \right\} \boldsymbol{X}.$$

Since the best α is given by $\alpha = (p - m - 1)/(n + 2m - p + 1)$, we have the empirical Bayes estimator

$$\widehat{\boldsymbol{\Theta}}^{MEM}(\beta) = \widehat{\boldsymbol{\Theta}}^{EM} - \frac{\beta}{\mathrm{tr}\, \boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t} \boldsymbol{X},$$

which is the modified Efron-Morris estimator.

3 A unified method for the improvement

3.1 Improvement by a scalar shrinkage

In the previous section, the modified Efron-Morris estimator has been characterized as an empirical Bayes estimator, which modifies the crude Efron-Morris estimator $\widehat{\boldsymbol{\Theta}}^{EM}$ by adding the scalar shrinkage term $-(\beta/\mathrm{tr}\, \boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1})\boldsymbol{X}$ for a positive constant β. As proved by Konno (1991, 1992), this modification yields the further improvement. In this section, we investigate whether the idea of this modification can be established as the general method for improving estimators.

Consider the general class of estimators of the form

$$\widehat{\boldsymbol{\Theta}}(\boldsymbol{\Psi}) = \begin{cases} \boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^{-1}) & \text{if } m \geq p, \\ (\boldsymbol{I}_m - \boldsymbol{R}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{R}^t)\boldsymbol{X} & \text{if } m < p, \end{cases} \tag{3.1}$$

where $\boldsymbol{\Psi} = \mathrm{diag}\,(\psi_1, \ldots, \psi_{m\wedge p})$ for $m \wedge p = \min(m, p)$. It is noted that this class of estimators is equivariant under the group of transformations $\boldsymbol{X} \to \boldsymbol{O}\boldsymbol{X}\boldsymbol{P}$ and $\boldsymbol{S} \to \boldsymbol{P}^t\boldsymbol{S}\boldsymbol{P}$ where $\boldsymbol{O}$ is an $m \times m$ orthogonal matrix and $\boldsymbol{P}$ is a $p \times p$ nonsingular matrix. It is also noted that the class (3.1) includes several shrinkage estimators proposed in Bilodeau and Kariya (1989) and Konno (1991, 1992), but the empirical Bayes estimators given by Ghosh and Shieh (1991, 92) and Shieh (1993) do not belong to the class. Employing the same idea as appeared in $\widehat{\boldsymbol{\Theta}}^{MEM}$, we shall modify $\widehat{\boldsymbol{\Theta}}(\boldsymbol{\Psi})$ as

$$\widehat{\boldsymbol{\Theta}}^M = \widehat{\boldsymbol{\Theta}}(\boldsymbol{\Psi}) - (\beta/\mathrm{tr}\, \boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1})\boldsymbol{X}, \tag{3.2}$$

which is rewritten by

$$\widehat{\boldsymbol{\Theta}}^M = \begin{cases} \boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{Q}\boldsymbol{\Psi}^M\boldsymbol{Q}^{-1}) & \text{if } m \geq p, \\ (\boldsymbol{I}_m - \boldsymbol{R}\boldsymbol{\Psi}^M\boldsymbol{R}^t)\boldsymbol{X} & \text{if } m < p, \end{cases}$$

where $\boldsymbol{\Psi}^M = \mathrm{diag}\,(\psi_1^M, \ldots, \psi_{m\wedge p}^M)$ with $\psi_i^M = \psi_i + \beta/\mathrm{tr}\,\boldsymbol{F}$. The following lemma which will be proved in the next subsection provides the conditions on $\boldsymbol{\Psi}$ and β for $\widehat{\boldsymbol{\Theta}}^M$ to dominate $\widehat{\boldsymbol{\Theta}}(\boldsymbol{\Psi})$. For the convenience, define $\mathcal{H}(\boldsymbol{F}, \boldsymbol{\Psi})$ by

$$\begin{aligned} \mathcal{H}(\boldsymbol{F}, \boldsymbol{\Psi}) =& \{n + (2m-p)\wedge p - 3\}\frac{\mathrm{tr}\,\boldsymbol{F}\boldsymbol{\Psi}}{\mathrm{tr}\,\boldsymbol{F}} + 2\frac{\mathrm{tr}\,\boldsymbol{F}^2\boldsymbol{\Psi}}{(\mathrm{tr}\,\boldsymbol{F})^2} \\ & - \frac{2}{\mathrm{tr}\,\boldsymbol{F}}\sum_{i=1}^{m\wedge p}\Big\{f_i^2\frac{\partial\psi_i}{\partial f_i} + \sum_{j>i}\frac{f_i^2\psi_i - f_j^2\psi_j}{f_i - f_j}\Big\}. \end{aligned} \tag{3.3}$$

Lemma 3.1 *Assume that $\boldsymbol{\Psi}$, β and a constant c satisfy the following conditions for $c < mp - 2$:*

(a) $\mathcal{H}(\boldsymbol{F}, \boldsymbol{\Psi}) \leq c/\mathrm{tr}\,\boldsymbol{F}$,

(b) $0 < \beta \leq 2(mp - 2 - c)/(n - p + 3)$.

Then, the modified shrinkage estimator $\widehat{\boldsymbol{\Theta}}^M$ improves on the crude one $\widehat{\boldsymbol{\Theta}}(\boldsymbol{\Psi})$ relative to the loss (1.2)

This lemma is very useful for deriving improved estimators. A simple application of the lemma is the improvement of $\widehat{\boldsymbol{\Theta}}^{ML} = \boldsymbol{X}$, which corresponds to the case of $\boldsymbol{\Psi} = \boldsymbol{0}$. Lemma 3.1 for $c = 0$ implies that $\boldsymbol{X}$ is dominated by the James-Stein estimator $\widehat{\boldsymbol{\Theta}}^{JS}(\beta) = (1 - \beta/\mathrm{tr}\,\boldsymbol{F})\boldsymbol{X}$ for $0 < \beta \leq 2(mp-2)/(n-p+3)$. Another simple example is the application to the Efron-Morris estimator $\widehat{\boldsymbol{\Theta}}^{EM}$ given by (1.3). Since $\boldsymbol{\Psi} = \alpha\boldsymbol{F}^{-1}$, $\mathcal{H}(\boldsymbol{F}, \boldsymbol{\Psi})$ is written as

$$\begin{aligned} \mathcal{H}(\boldsymbol{F}, \alpha\boldsymbol{F}^{-1}) =& \{n + (2m-p)\wedge p - 3\}\alpha\frac{m\wedge p}{\mathrm{tr}\,\boldsymbol{F}} + \frac{2\alpha}{\mathrm{tr}\,\boldsymbol{F}} - \frac{2}{\mathrm{tr}\,\boldsymbol{F}}\sum_{i=1}^{m\wedge p}\Big\{-\alpha + \sum_{j>i}\alpha\Big\} \\ =& \frac{\alpha\{(n + (2m-p)\wedge p - m\wedge p)(m\wedge p) + 2\}}{\mathrm{tr}\,\boldsymbol{F}}. \end{aligned}$$

Applying Lemma 3.1 for $c = \{n + (2m-p)\wedge p - m\wedge p\}(m\wedge p)\alpha + 2\alpha$, we can see that the Efron-Morris estimator $\widehat{\boldsymbol{\Theta}}^{EM}(\alpha)$ is dominated by the modified one $\widehat{\boldsymbol{\Theta}}^{MEM}(\beta) = \widehat{\boldsymbol{\Theta}}^{EM}(\alpha) - (\beta/\mathrm{tr}\,\boldsymbol{F})\boldsymbol{X}$ if β satisfies the condition $0 < \beta \leq 2[mp - 2 - \{n + (2m-p)\wedge p - m\wedge p\}(m\wedge p)\alpha - 2\alpha]/(n-p+3)$. Since the best α is given by $\alpha = (|m-p|-1)/\{n + (2m-p)\wedge p + 1\}$, this condition can be rewritten by

$$0 < \beta \leq 2\frac{(m\wedge p - 1)(m\wedge p + 2)(n+m)}{\{n + (2m-p)\wedge p + 1\}(n-p+3)},$$

which was derived by Konno (1992).

A nice application of Lemma 3.1 is obtained for the Stein estimator $\widehat{\boldsymbol{\Theta}}^{ST}$ given by

$$\widehat{\boldsymbol{\Theta}}^{ST} = \begin{cases} \boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{Q}\boldsymbol{D}\boldsymbol{F}^{-1}\boldsymbol{Q}^{-1}) & \text{if } m \geq p+2, \\ (\boldsymbol{I}_m - \boldsymbol{R}\boldsymbol{D}\boldsymbol{F}^{-1}\boldsymbol{R}^t)\boldsymbol{X} & \text{if } p \geq m+2, \end{cases} \tag{3.4}$$

where $\boldsymbol{D} = \mathrm{diag}\,(d_1, \ldots, d_{m\wedge p})$ with $d_i = (m+p-2i-1)/(n-p+2i+1)$. Consider the modified Stein estimator

$$\widehat{\boldsymbol{\Theta}}^{MST} = \widehat{\boldsymbol{\Theta}}^{ST} - \frac{\beta}{\mathrm{tr}\,\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1}}\boldsymbol{X} = \widehat{\boldsymbol{\Theta}}^{ST} - \frac{\beta}{\mathrm{tr}\,\boldsymbol{F}}\boldsymbol{X},$$

and we obtain the following dominance result:

Theorem 3.1 *The Stein estimator $\widehat{\boldsymbol{\Theta}}^{ST}$ is dominated by the modified Stein estimator $\widehat{\boldsymbol{\Theta}}^{MST}$ relative to the loss (1.2) if β satisfies the condition*

$$0 < \beta \le \frac{4(m\wedge p - 1 + \sum_{i=2}^{m\wedge p} d_i)}{n-p+3}.$$

Proof. Letting $\boldsymbol{\Psi} = \boldsymbol{D}\boldsymbol{F}^{-1}$, we can see that $\mathcal{H}(\boldsymbol{F}, \boldsymbol{\Psi})$ for $m \ge p$ is written as

$$\mathcal{H}(\boldsymbol{F}, \boldsymbol{\Psi}) = (n+p-1)\frac{\mathrm{tr}\,\boldsymbol{D}}{\mathrm{tr}\,\boldsymbol{F}} + 2\frac{\mathrm{tr}\,\boldsymbol{F}\boldsymbol{D}}{(\mathrm{tr}\,\boldsymbol{F})^2} - \frac{2}{\mathrm{tr}\,\boldsymbol{F}}\sum_{i=1}^{p}\sum_{j>i}\frac{f_i d_i - f_j d_j}{f_i - f_j}.$$

It is noted that

$$\begin{aligned}
\sum_{i=1}^{p}\sum_{j>i}\frac{f_i d_i - f_j d_j}{f_i - f_j} &= \sum_{i=1}^{p}\sum_{j>i}\left\{d_i + \frac{f_j(d_i - d_j)}{f_i - f_j}\right\} \\
&= \sum_{i=1}^{p}(p-i)d_i + \sum_{i=1}^{p}\sum_{j>i}\frac{f_j(d_i-d_j)}{f_i - f_j} \ge \sum_{i=1}^{p}(p-i)d_i, \\
\frac{\mathrm{tr}\,\boldsymbol{F}\boldsymbol{D}}{(\mathrm{tr}\,\boldsymbol{F})^2} &\le \frac{d_1}{\mathrm{tr}\,\boldsymbol{F}} = \frac{\mathrm{tr}\,\boldsymbol{D}}{\mathrm{tr}\,\boldsymbol{F}} - \frac{\sum_{i=2}^{p} d_i}{\mathrm{tr}\,\boldsymbol{F}},
\end{aligned}$$

which give that

$$\mathcal{H}(\boldsymbol{F}, \boldsymbol{\Psi}) \le \frac{1}{\mathrm{tr}\,\boldsymbol{F}}\sum_{i=1}^{p}(n-p+2i+1)d_i - 2\frac{\sum_{i=2}^{p} d_i}{\mathrm{tr}\,\boldsymbol{F}} = \frac{mp - 2p - 2\sum_{i=2}^{p} d_i}{\mathrm{tr}\,\boldsymbol{F}},$$

since $\sum_{i=1}^{p}(n-p+2i+1)d_i = \sum_{i=1}^{p}(m+p-2i-1) = p(m-2)$. Then Lemma 3.1 is applied to complete the proof. The result for $m < p$ can be similarly verified. ■

The risk expression (2.2) means that the estimation of the mean matrix is related to that of ratio of covariance matrices. This suggests that the estimators proposed for a covariance matrix or a ratio of covariance matrices can be employed for our problem. It is clear that the Efron-Morris and the Stein estimators $\widehat{\boldsymbol{\Theta}}^{EM}$ and $\widehat{\boldsymbol{\Theta}}^{ST}$ can be interpreted through the same idea. We here handle the other estimators induced from the estimators given by Dey (1987) and Haff (1980) for the covariance matrix. These estimators have the forms

$$\widehat{\boldsymbol{\Theta}}^{DY} = \boldsymbol{X}\left\{\boldsymbol{I}_p - \frac{\alpha}{\mathrm{tr}\,\boldsymbol{F}^2}\boldsymbol{S}^{-1}\boldsymbol{X}^t\boldsymbol{X}\right\} = \left\{\boldsymbol{I}_m - \frac{\alpha}{\mathrm{tr}\,\boldsymbol{F}^2}\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t\right\}\boldsymbol{X},$$

$$\widehat{\boldsymbol{\Theta}}^{HF} = \begin{cases}\boldsymbol{X}\{\boldsymbol{I}_p - \alpha\boldsymbol{Q}(\boldsymbol{F} + \delta\boldsymbol{I}_p/\mathrm{tr}\,\boldsymbol{F}^{-1})^{-1}\boldsymbol{Q}^{-1}\} & \text{if } m \ge p, \\ \{\boldsymbol{I}_m - \alpha\boldsymbol{R}(\boldsymbol{F} + \delta\boldsymbol{I}_m/\mathrm{tr}\,\boldsymbol{F}^{-1})^{-1}\boldsymbol{R}^t\}\boldsymbol{X} & \text{if } m < p,\end{cases}$$

where α and δ are positive constants. The estimators $\widehat{\boldsymbol{\Theta}}^{DY}$ and $\widehat{\boldsymbol{\Theta}}^{HF}$ are respectively called the Dey and Haff estimators in this paper. It is noted that the Dey estimator can be expressed by

$$\widehat{\boldsymbol{\Theta}}^{DY} = \boldsymbol{X}\Big\{\boldsymbol{I}_p - \frac{\alpha}{\mathrm{tr}\,\boldsymbol{F}^2}\boldsymbol{Q}\boldsymbol{F}\boldsymbol{Q}^{-1}\Big\} = \Big\{\boldsymbol{I}_m - \frac{\alpha}{\mathrm{tr}\,\boldsymbol{F}^2}\boldsymbol{R}\boldsymbol{F}\boldsymbol{R}^t\Big\}\boldsymbol{X},$$

which means that $\widehat{\boldsymbol{\Theta}}^{DY}$ is the same for both $m \ge p$ and $m < p$. The minimaxities of the Dey and Haff estimators relative to the loss (1.2) can be guaranteed by the following lemma, whose proof will be given in the next subsection.

Lemma 3.2
(1) *The Dey estimator* $\widehat{\boldsymbol{\Theta}}^{DY}$ *is minimax if* $0 < \alpha \le 2(m+p-3)/(n-p+3)$.
(2) *The Haff estimator* $\widehat{\boldsymbol{\Theta}}^{HF}$ *is minimax if* $m \ge p+2$ *and* $0 < \alpha \le 2(m-p-1)/(n+p+1)$ *or if* $p \ge m+2$ *and* $0 < \alpha \le 2(p-m-1)/(n+2m-p+1)$.

It is interesting to show that these estimators $\widehat{\boldsymbol{\Theta}}^{DY}$ and $\widehat{\boldsymbol{\Theta}}^{HF}$ can be further improved on by their modified estimators, respectively,

$$\begin{aligned}\widehat{\boldsymbol{\Theta}}^{MDY} &= \widehat{\boldsymbol{\Theta}}^{DY} - (\beta/\mathrm{tr}\,\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1})\boldsymbol{X} = \widehat{\boldsymbol{\Theta}}^{DY} - (\beta/\mathrm{tr}\,\boldsymbol{F})\boldsymbol{X},\\ \widehat{\boldsymbol{\Theta}}^{MHF} &= \widehat{\boldsymbol{\Theta}}^{HF} - (\beta/\mathrm{tr}\,\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1})\boldsymbol{X} = \widehat{\boldsymbol{\Theta}}^{HF} - (\beta/\mathrm{tr}\,\boldsymbol{F})\boldsymbol{X}.\end{aligned}$$

Theorem 3.2
(1) *If* $0 < \beta \le 2\{mp-2-\alpha(n-p+3)\}/(n-p+3)$ *then* $\widehat{\boldsymbol{\Theta}}^{MDY}$ *dominates* $\widehat{\boldsymbol{\Theta}}^{DY}$ *relative to the loss* (1.2).
(2) *If* $m \ge p+2$ *and* $0 < \beta \le 2\{mp-2-p(n+p+1)\alpha\}/(n-p+3)$ *or if* $p \ge m+2$ *and* $0 < \beta \le 2\{mp-2-m(n+2m-p+1)\alpha\}/(n-p+3)$, *then* $\widehat{\boldsymbol{\Theta}}^{MHF}$ *dominates* $\widehat{\boldsymbol{\Theta}}^{HF}$ *relative to the loss* (1.2).

Proof. The proof in the case of $m \ge p$ is stated here. For the proof of (1), note that $\partial\psi_i/\partial f_i = \alpha/\mathrm{tr}\,\boldsymbol{F}^2 - 2\alpha f_i^2/(\mathrm{tr}\,\boldsymbol{F}^2)^2$ for $\psi_i = \alpha f_i/\mathrm{tr}\,\boldsymbol{F}^2$. Then, $\mathcal{H}(\boldsymbol{F},\boldsymbol{\Psi})$ given by (3.3) is equal to

$$\mathcal{H}(\boldsymbol{F},\boldsymbol{\Psi}) = \frac{(n+p-5)\alpha}{\mathrm{tr}\,\boldsymbol{F}} + \frac{2\alpha\mathrm{tr}\,\boldsymbol{F}^3}{\mathrm{tr}\,\boldsymbol{F}^2(\mathrm{tr}\,\boldsymbol{F})^2} + \frac{4\alpha\mathrm{tr}\,\boldsymbol{F}^4}{\mathrm{tr}\,\boldsymbol{F}(\mathrm{tr}\,\boldsymbol{F}^2)^2} - \frac{2\alpha}{\mathrm{tr}\,\boldsymbol{F}\mathrm{tr}\,\boldsymbol{F}^2}\sum_{i=1}^p\sum_{j>i}\frac{f_i^3-f_j^3}{f_i-f_j}.$$

It is noted that $\mathrm{tr}\,\boldsymbol{F}^3 \le \mathrm{tr}\,\boldsymbol{F}\mathrm{tr}\,\boldsymbol{F}^2$, $\mathrm{tr}\,\boldsymbol{F}^4 \le (\mathrm{tr}\,\boldsymbol{F}^2)^2$ and

$$\sum_{i=1}^p\sum_{j>i}\frac{f_i^3-f_j^3}{f_i-f_j} = (p-1)\mathrm{tr}\,\boldsymbol{F}^2 + \sum_{i=1}^p\sum_{j>i} f_if_j \ge (p-1)\mathrm{tr}\,\boldsymbol{F}^2.$$

Using the inequalities given above, we get that $\mathcal{H}(\boldsymbol{F},\boldsymbol{\Psi}) \le (n-p+3)\alpha/\mathrm{tr}\,\boldsymbol{F}$, so that Lemma 3.1 can be applied to obtain the requested result.

For the proof of (2), ψ_i is written by $\psi_i = \alpha/\ell_i$ for $\ell_i = f_i + \delta/\mathrm{tr}\,\boldsymbol{F}^{-1}$. The partial derivative can be evaluated as

$$\frac{\partial\psi_i}{\partial f_i} = -\frac{\alpha}{\ell_i^2}\left\{1+\frac{\delta}{f_i^2(\mathrm{tr}\,\boldsymbol{F}^{-1})^2}\right\} \ge -\frac{\alpha}{\ell_i^2}\left\{1+\frac{\delta}{f_i\mathrm{tr}\,\boldsymbol{F}^{-1}}\right\} = -\frac{\alpha}{f_i\ell_i}.$$

It is noted that $f_i/\ell_i \leq 1$ and

$$\frac{f_i^2}{\ell_i} - \frac{f_j^2}{\ell_j} = \frac{1}{\ell_i \ell_j}\Big\{ f_i f_j (f_i - f_j) + \frac{\delta}{\mathrm{tr}\, \boldsymbol{F}^{-1}} (f_i^2 - f_j^2) \Big\} \geq 0$$

for $j > i$. Then, $\mathcal{H}(\boldsymbol{F}, \boldsymbol{\Psi})$ can be written by

$$\begin{aligned} \mathcal{H}(\boldsymbol{F}, \boldsymbol{\Psi}) = & \sum_{i=1}^{p} \Big\{ (n+p-3) \frac{\alpha f_i}{\ell_i \mathrm{tr}\, \boldsymbol{F}} + \frac{2\alpha f_i^2}{\ell_i (\mathrm{tr}\, \boldsymbol{F})^2} - \frac{2 f_i^2}{\mathrm{tr}\, \boldsymbol{F}} \frac{\partial}{\partial f_i} \frac{\alpha}{\ell_i} \Big\} \\ & - \frac{2\alpha}{\mathrm{tr}\, \boldsymbol{F}} \sum_{i=1}^{p} \sum_{j>i} \frac{f_i^2/\ell_i - f_j^2/\ell_j}{f_i - f_j}, \end{aligned}$$

which is less than or equal to $p(n+p+1)\alpha/\mathrm{tr}\, \boldsymbol{F}$. Hence from Lemma 3.1, we obtain the result (2) of Theorem 3.2. ■

Remark 3.1 As other modification rules of the estimator $\widehat{\boldsymbol{\Theta}}(\boldsymbol{\Psi})$ given by (3.1), we can consider the procedures $\widehat{\boldsymbol{\Theta}}(\boldsymbol{\Psi}) - (\beta/\mathrm{tr}\, \boldsymbol{\Psi}^{-1})\boldsymbol{X}$ and $\widehat{\boldsymbol{\Theta}}(\boldsymbol{\Psi}) - (\beta/\mathrm{tr}\, \boldsymbol{F}^2)\boldsymbol{X}\boldsymbol{S}^{-1}\boldsymbol{X}^t\boldsymbol{X}$ although the details are omitted.

3.2 Proofs of Lemmas

All the results in this paper can be proved based on the following lemma which provides the unbiased estimate of the risk function of the estimator (1.5) or (3.1). For the proof, see Konno (1992).

Lemma 3.3 *The unbiased risk estimate of the estimator* (1.5) *or* (3.1) *is given by*

$$\begin{aligned} \widehat{R}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}}) = mp + \sum_{i=1}^{m \wedge p} \Bigg\{ & (n + (2m-p) \wedge p - 3) f_i \psi_i^2 - 4 f_i^2 \psi_i \frac{\partial \psi_i}{\partial f_i} - 2 \sum_{j>i} \frac{f_i^2 \psi_i^2 - f_j^2 \psi_j^2}{f_i - f_j} \\ & -2(m \vee p - m \wedge p + 1)\psi_i - 4 f_i \frac{\partial \psi_i}{\partial f_i} - 4 \sum_{j>i} \frac{f_i \psi_i - f_j \psi_j}{f_i - f_j} \Bigg\} \end{aligned}$$

for $a \vee b = \max(a, b)$ *and* $a \wedge b = \min(a, b)$.

From Lemma 3.3, the unbiased risk estimate of the ML estimator $\widehat{\boldsymbol{\Theta}}^{ML} = \boldsymbol{X}$ is mp, which is the minimax risk. This means that an estimator whose unbiased risk estimate $\widehat{R}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}})$ is smaller than mp is minimax. Through the paper, we shall provide the proofs in the case of $m \geq p$ and omit the proofs for $m < p$ since they can be similarly done with replacing (n, m, p) with $(n + m - p, p, m)$.

Proof of Lemma 3.1. Using Lemma 3.3 for $m \geq p$, we can write the difference of $\widehat{R}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}}^M)$ and $\widehat{R}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}}(\boldsymbol{\Psi}))$ as

$$
\begin{aligned}
\widehat{\Delta} &= \widehat{R}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}}^M) - \widehat{R}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}}(\boldsymbol{\Psi})) \\
&= \sum_{i=1}^{p} \Big\{ (n+p-3) f_i \Big(\frac{2\psi_i \beta}{\mathrm{tr}\, \boldsymbol{F}} + \frac{\beta^2}{(\mathrm{tr}\, \boldsymbol{F})^2} \Big) - 4 f_i^2 \psi_i \frac{\partial}{\partial f_i} \frac{\beta}{\mathrm{tr}\, \boldsymbol{F}} - \frac{4 f_i^2 \beta}{\mathrm{tr}\, \boldsymbol{F}} \frac{\partial}{\partial f_i} \Big(\psi_i + \frac{\beta}{\mathrm{tr}\, \boldsymbol{F}} \Big) \\
&\quad - \frac{4\beta}{\mathrm{tr}\, \boldsymbol{F}} \sum_{j>i} \frac{f_i^2 \psi_i - f_j^2 \psi_j}{f_i - f_j} - \frac{2\beta^2}{(\mathrm{tr}\, \boldsymbol{F})^2} \sum_{j>i} \frac{f_i^2 - f_j^2}{f_i - f_j} - 2(m-p+1) \frac{\beta}{\mathrm{tr}\, \boldsymbol{F}} \\
&\quad - 4 f_i \frac{\partial}{\partial f_i} \frac{\beta}{\mathrm{tr}\, \boldsymbol{F}} - \frac{4\beta}{\mathrm{tr}\, \boldsymbol{F}} \sum_{j>i} \frac{f_i - f_j}{f_i - f_j} \Big\}.
\end{aligned}
$$

It is noted that $(\partial/\partial f_i)(\beta/\mathrm{tr}\, \boldsymbol{F}) = -\beta/(\mathrm{tr}\, \boldsymbol{F})^2$, $\sum_{i=1}^{p} \sum_{j>i} (f_i - f_j)/(f_i - f_j) = p(p-1)/2$ and

$$
\sum_{i=1}^{p} \sum_{j>i} \frac{f_i^2 - f_j^2}{f_i - f_j} = (p-1) \sum_{i=1}^{p} f_i = (p-1) \mathrm{tr}\, \boldsymbol{F},
$$

which imply that

$$
\begin{aligned}
\widehat{\Delta} &= (n-p-1) \frac{\beta^2}{\mathrm{tr}\, \boldsymbol{F}} + 2\beta(n+p-3) \frac{\mathrm{tr}\, \boldsymbol{F}\boldsymbol{\Psi}}{\mathrm{tr}\, \boldsymbol{F}} + 4\beta \frac{\mathrm{tr}\, \boldsymbol{F}^2 \boldsymbol{\Psi}}{(\mathrm{tr}\, \boldsymbol{F})^2} + 4\beta^2 \frac{\mathrm{tr}\, \boldsymbol{F}^2}{(\mathrm{tr}\, \boldsymbol{F})^3} \\
&\quad - \frac{4\beta}{\mathrm{tr}\, \boldsymbol{F}} \sum_{i=1}^{p} f_i^2 \frac{\partial \psi_i}{\partial f_i} - \frac{4\beta}{\mathrm{tr}\, \boldsymbol{F}} \sum_{i=1}^{p} \sum_{j>i} \frac{f_i^2 \psi_i - f_j^2 \psi_j}{f_i - f_j} - 2(mp-2) \frac{\beta}{\mathrm{tr}\, \boldsymbol{F}}.
\end{aligned}
$$

Since $\mathrm{tr}\, \boldsymbol{F}^2/(\mathrm{tr}\, \boldsymbol{F})^3 \leq 1/\mathrm{tr}\, \boldsymbol{F}$, the difference $\widehat{\Delta}$ is evaluated as

$$
\begin{aligned}
\widehat{\Delta} &\leq (n-p+3) \frac{\beta^2}{\mathrm{tr}\, \boldsymbol{F}} - 2(mp-2) \frac{\beta}{\mathrm{tr}\, \boldsymbol{F}} + 2\beta(n+p-3) \frac{\mathrm{tr}\, \boldsymbol{F}\boldsymbol{\Psi}}{\mathrm{tr}\, \boldsymbol{F}} \\
&\quad + 4\beta \frac{\mathrm{tr}\, \boldsymbol{F}^2 \boldsymbol{\Psi}}{(\mathrm{tr}\, \boldsymbol{F})^2} - \frac{4\beta}{\mathrm{tr}\, \boldsymbol{F}} \sum_{i=1}^{p} f_i^2 \frac{\partial \psi_i}{\partial f_i} - \frac{4\beta}{\mathrm{tr}\, \boldsymbol{F}} \sum_{i=1}^{p} \sum_{j>i} \frac{f_i^2 \psi_i - f_j^2 \psi_j}{f_i - f_j}.
\end{aligned}
$$

From the assumption (3.3), we get the inequality

$$
\widehat{\Delta} \leq (n-p+3) \frac{\beta^2}{\mathrm{tr}\, \boldsymbol{F}} - 2(mp-2-c) \frac{\beta}{\mathrm{tr}\, \boldsymbol{F}},
$$

which is not positive if $0 < \beta \leq 2(mp-2-c)/(n-p+3)$. Hence the proof is complete in the case of $m \geq p$. ■

Proof of Lemma 3.2. For the proof of (1), note the proof of Theorem 3.2 (1). Then from Lemma 3.3, the unbiased estimate of the risk difference is written by

$$
\begin{aligned}
\widehat{\Delta}_{DY} &= \widehat{R}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}}^{DY}) - mp \\
&= \frac{\alpha^2}{(\mathrm{tr}\, \boldsymbol{F}^2)^2} \Big\{ (n+p-7) \mathrm{tr}\, \boldsymbol{F}^3 + 8 \frac{\mathrm{tr}\, \boldsymbol{F}^5}{\mathrm{tr}\, \boldsymbol{F}^2} - 2 \sum_{i=1}^{p} \sum_{j>i} \frac{f_i^4 - f_j^4}{f_i - f_j} \Big\} \\
&\quad - \frac{2\alpha}{\mathrm{tr}\, \boldsymbol{F}^2} \Big\{ (m-p+3) \mathrm{tr}\, \boldsymbol{F} - 4 \frac{\mathrm{tr}\, \boldsymbol{F}^3}{\mathrm{tr}\, \boldsymbol{F}^2} + 2 \sum_{i=1}^{p} \sum_{j>i} \frac{f_i^2 - f_j^2}{f_i - f_j} \Big\}.
\end{aligned}
$$

It is observed that

$$
\begin{aligned}
\sum_{i=1}^{p}\sum_{j>i}\frac{f_i^4-f_j^4}{f_i-f_j} =&(p-1)\mathrm{tr}\,\boldsymbol{F}^3+\sum_{i=1}^{p}\sum_{j>i}(f_i^2f_j+f_if_j^2)\geq (p-1)\mathrm{tr}\,\boldsymbol{F}^3,\\
\sum_{i=1}^{p}\sum_{j>i}\frac{f_i^2-f_j^2}{f_i-f_j} =&(p-1)\mathrm{tr}\,\boldsymbol{F},
\end{aligned}\tag{3.5}
$$

which can be used to get that

$$
\widehat{\Delta}_{DY}\leq \frac{\alpha\mathrm{tr}\,\boldsymbol{F}}{\mathrm{tr}\,\boldsymbol{F}^2}\left\{(n-p-5)\alpha\frac{\mathrm{tr}\,\boldsymbol{F}^3}{\mathrm{tr}\,\boldsymbol{F}\mathrm{tr}\,\boldsymbol{F}^2}+\frac{8\alpha\mathrm{tr}\,\boldsymbol{F}^5}{(\mathrm{tr}\,\boldsymbol{F}^2)^2\mathrm{tr}\,\boldsymbol{F}}-2(m+p+1)+\frac{8\mathrm{tr}\,\boldsymbol{F}^3}{\mathrm{tr}\,\boldsymbol{F}\mathrm{tr}\,\boldsymbol{F}^2}\right\}.
$$

Since $\mathrm{tr}\,\boldsymbol{F}^3\leq\mathrm{tr}\,\boldsymbol{F}\mathrm{tr}\,\boldsymbol{F}^2$ and $\mathrm{tr}\,\boldsymbol{F}^5\leq\mathrm{tr}\,\boldsymbol{F}(\mathrm{tr}\,\boldsymbol{F}^2)^2$, $\widehat{\Delta}_{DY}$ can be evaluated as

$$
\widehat{\Delta}_{DY}\leq\frac{\alpha\mathrm{tr}\,\boldsymbol{F}}{\mathrm{tr}\,\boldsymbol{F}^2}\Big[(n-p+3)\alpha-2(m+p-3)\Big],
$$

which proves the result (1) of Lemma 3.2 for $m\geq p$.

For the proof of (2), recall the notation and techniques used in the proof of Theorem 3.2 (2). Then from Lemma 3.3, we can see that

$$
\begin{aligned}
\widehat{\Delta}_{HF}=&\widehat{R}(\boldsymbol{\Theta},\widehat{\boldsymbol{\Theta}}^{HF})-mp\\
&\sum_{i=1}^{p}\Big\{(n+p-3)\frac{f_i\alpha^2}{\ell_i^2}+\frac{4\alpha^2f_i}{\ell_i^3}\Big(f_i+\frac{\delta/f_i}{(\mathrm{tr}\,\boldsymbol{F}^{-1})^2}\Big)+\frac{4\alpha}{\ell_i^2}\Big(f_i+\frac{\delta/f_i}{(\mathrm{tr}\,\boldsymbol{F}^{-1})^2}\Big)\\
&-2(m-p+1)\frac{\alpha}{\ell_i}-2\alpha^2\sum_{i=1}^{p}\sum_{j>i}\frac{f_i^2/\ell_i^2-f_j^2/\ell_j^2}{f_i-f_j}-4\alpha\sum_{i=1}^{p}\sum_{j>i}\frac{f_i/\ell_i-f_j/\ell_j}{f_i-f_j}\Big\}.
\end{aligned}
$$

The following inequalities are useful for evaluating the risk difference:

$$
\begin{aligned}
&f_i+\frac{\delta/f_i}{(\mathrm{tr}\,\boldsymbol{F}^{-1})^2}\leq f_i+\frac{\delta}{\mathrm{tr}\,\boldsymbol{F}^{-1}}=\ell_i,\\
&\frac{f_i}{\ell_i}-\frac{f_j}{\ell_j}=\frac{\delta(f_i-f_j)}{\ell_i\ell_j\mathrm{tr}\,\boldsymbol{F}^{-1}}\geq 0,\\
&\frac{f_i^2}{\ell_i^2}-\frac{f_j^2}{\ell_j^2}=\Big(\frac{f_i}{\ell_i}+\frac{f_j}{\ell_j}\Big)\Big(\frac{f_i}{\ell_i}-\frac{f_j}{\ell_j}\Big)\geq 0
\end{aligned}
$$

for $j>i$. Then, we obtain that

$$
\begin{aligned}
\widehat{\Delta}_{HF}\leq&\sum_{i=1}^{p}\Big\{(n+p+1)\frac{f_i\alpha^2}{\ell_i^2}-2(m-p-1)\frac{\alpha}{\ell_i}\Big\}\\
\leq&\sum_{i=1}^{p}\frac{\alpha}{\ell_i}\big\{(n+p+1)\alpha-2(m-p-1)\big\},
\end{aligned}
$$

which proves the result (2) of Lemma 3.2 for $m\geq p$. ■

4 Improvement by a combined method

There are many minimax estimators and their risk behaviors have various characteristics. Of these, in this section, we look into the James-Stein estimator

$$\widehat{\boldsymbol{\Theta}}^{JS} = (1 - \beta_0/\mathrm{tr}\, \boldsymbol{X}^t \boldsymbol{X} \boldsymbol{S}^{-1})\boldsymbol{X} \quad \text{for} \quad \beta_0 = \frac{mp-2}{n-p+3}$$

and the modified Stein estimator

$$\widehat{\boldsymbol{\Theta}}^{MST} = \widehat{\boldsymbol{\Theta}}^{ST} - \frac{\beta_1}{\mathrm{tr}\, \boldsymbol{X}^t \boldsymbol{X} \boldsymbol{S}^{-1}}\boldsymbol{X} \quad \text{for} \quad \beta_1 = \frac{2(m \wedge p - 1 + \sum_{i=2}^{m\wedge p} d_i)}{n-p+3}.$$

The James-Stein estimator $\widehat{\boldsymbol{\Theta}}^{JS}$ is known to give the significant improvement near $\boldsymbol{\Theta} = \boldsymbol{0}$, while the modified Stein estimator $\widehat{\boldsymbol{\Theta}}^{MST}$ is much better than $\widehat{\boldsymbol{\Theta}}^{JS}$ when $\boldsymbol{\Theta}$ is far away from zero. In this section, we want to construct a combined estimator of $\widehat{\boldsymbol{\Theta}}^{JS}$ and $\widehat{\boldsymbol{\Theta}}^{MST}$ such that its risk behavior inherits the nice risk properties of both $\widehat{\boldsymbol{\Theta}}^{JS}$ and $\widehat{\boldsymbol{\Theta}}^{MST}$.

A simple combination with the form $(1-\varepsilon_0)\widehat{\boldsymbol{\Theta}}^{MST} + \varepsilon_0 \widehat{\boldsymbol{\Theta}}^{JS}$ for a constant $\varepsilon_0 \in [0,1]$ is minimax from the convexity of the loss function. However, such a simple combined estimator may be the second best. We thus consider a random combination of the estimators given by

$$\widehat{\boldsymbol{\Theta}}^{CM} = (1-\varepsilon)\widehat{\boldsymbol{\Theta}}^{MST} + \varepsilon \widehat{\boldsymbol{\Theta}}^{JS},$$

where $\varepsilon = \varepsilon(\boldsymbol{F})$ is a function of $\boldsymbol{F}$ satisfying $0 \le \varepsilon(\boldsymbol{F}) \le 1$. We want to choose the weighting function $\varepsilon = \varepsilon(\boldsymbol{F})$ such that the combined estimator $\widehat{\boldsymbol{\Theta}}^{CM}$ inherits the nice risk properties of both $\widehat{\boldsymbol{\Theta}}^{JS}$ and $\widehat{\boldsymbol{\Theta}}^{MST}$. For this aim, it is reasonable to take a test statistic for testing the null hypothesis $H_0 : \boldsymbol{\Theta} = \boldsymbol{0}$ against $H_1 : \boldsymbol{\Theta} \neq \boldsymbol{0}$. Since the likelihood ratio statistic is of the form $\exp\{-n\mathrm{tr}\, \boldsymbol{F}/2\}$, a good choice of ε may be $\varepsilon = \exp\{-\gamma \mathrm{tr}\, \boldsymbol{F}\}$ for a positive constant γ. It is noted that $\varepsilon(\boldsymbol{F})$ may be small, that is, $\widehat{\boldsymbol{\Theta}}^{CM}$ may be close to $\widehat{\boldsymbol{\Theta}}^{JS}$ if each element of the mean matrix $\boldsymbol{\Theta}$ is near zero.

We now provide the condition for the minimaxity of the combined estimator $\widehat{\boldsymbol{\Theta}}^{CM}$. For $m \ge p$, let

$$g_p(\varepsilon) = c_0 + c_1\varepsilon + c_2\varepsilon^2, \tag{4.1}$$

where

$$\begin{aligned}
c_0 &= -p^2(m-p-1)d_p - \beta_1^2(n-p+3) \\
&\qquad + (4/e)\Big\{(\beta_0-\beta_1)(1+d_1+\beta_1) - \sum_{i=1}^p d_i - \sum_{i=1}^p d_i^2 - \beta_1 d_p\Big\}, \\
c_1 &= -2(n-p+3)\beta_1(\beta_0-\beta_1) \\
&\qquad + (4/e)\Big\{(\beta_0-\beta_1)^2 - (\beta_0-\beta_1)(d_1+d_p) + \sum_{i=1}^p d_i^2\Big\}, \\
c_2 &= p^2(m-p-1)d_p + 2(mp-2)\beta_1 - \beta_1^2(n-p+3) - (mp-2)\beta_0.
\end{aligned}$$

Also, for $m < p$, define $g_m(\varepsilon)$ as $g_p(\varepsilon)$ with replacing (n,m,p) with $(n-p+m, p, m)$. Then we get the following result:

Theorem 4.1 *The combined estimator* $\widehat{\Theta}^{CM}$ *is minimax relative to the loss* (1.2) *if the function* $g_{m\wedge p}(\varepsilon)$ *satisfies the condition*

$$\sup_{0<\varepsilon<1} g_{m\wedge p}(\varepsilon) \le 0. \tag{4.2}$$

For a value of the constant γ, we recommend the use of $\gamma = (n-p-1)/mp$ from the numerical investigation given in the next section.

Remark 4.1 Since $g_{m\wedge p}(\varepsilon)$ is a quadratic function of $\varepsilon \in [0,1]$, for example, the condition (4.2) is satisfied if $g_{m\wedge p}(0) \le 0$, $g_{m\wedge p}(1) \le 0$ and $g'_{m\wedge p}(0) \le 0$. That is, the condition (4.2) holds if (a) $c_0 \le 0$, (b) $c_1 \le 0$ and (c) $c_0 + c_1 + c_2 \le 0$. Checking these conditions numerically, we can reveal that they may be satisfied when $m > p+1$ for $m \ge p$ or $p > m+1$ for $m < p$. In fact, the numerical values of the coefficients (c_0, c_1, c_2) for several cases of (p, m, n) are reported in Tables 1, 2 and 3, which show that the coefficients c_0, c_1 and c_2 are negative for all the cases investigated here. This means that the condition (4.2) holds for the cases.

For large n, it is easily checked that the condition (4.2) holds for $m \ge p+2$ in the case of $m \ge p$. In fact, assuming that $\lim_{n\to\infty} \varepsilon_1 = A$, a constant in $[0,1]$, we can see that

$$\begin{aligned}\lim_{n\to\infty} n \times g_p(\varepsilon_1) = &- p^2(m-p-1) - 4(p-1)^2 \\ &- 4p(p-1)(m-2)A - p^2(p-1)(2m-p-3)A^2.\end{aligned}$$

■

Remark 4.2 Another reasonable choice of the weighting function $\varepsilon = \varepsilon(\boldsymbol{F})$ is given by $\varepsilon^* = \gamma\{(m\wedge p)|\boldsymbol{F}|^{1/(m\wedge p)}/\mathrm{tr}\,\boldsymbol{F}\}^\delta$ for constants γ and δ. In the model (2.1) of the mariginal distribution of $\boldsymbol{S}$ and $\boldsymbol{W}$, we consider testing the null hypothesis $H_0 : \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}^{1/2} = \lambda \boldsymbol{I}_p$ against $H_1 : \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}^{1/2} \ne \lambda \boldsymbol{I}_p$. Under H_0, the parameter λ may be estimated by $\hat{\lambda} = \beta/\mathrm{tr}\,\boldsymbol{W}\boldsymbol{S}^{-1}$, which yields the James-Stein estimator $\widehat{\Theta}^{JS}$. For H_1, on the other hand, $\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}$ is estimated by $\alpha\boldsymbol{W}^{-1}\boldsymbol{S}$, which gives the Efron-Morris estimator $\widehat{\Theta}^{EM}$. Since the weighting function ε^* corresponds to the likelihood ratio test statistic for testing the sphericity hypothesis H_0, so that it may be quite reasonable to consider the combined estimator

$$\widehat{\Theta}^* = (1-\varepsilon^*)\widehat{\Theta}^{EM} + \varepsilon^*\widehat{\Theta}^{JS}.$$

Based on ε^*, various combined estimators including $(1-\varepsilon^*)\widehat{\Theta}^{MST} + \varepsilon^*\widehat{\Theta}^{JS}$ are provided, and we can show the minimaxity of some combined estimators although the details are omitted here. ■

We shall prove Theorem 4.1, which is relatively hard to show. For the purpose, we need the inequalities in Lemma 4.1.

Lemma 4.1 *Let* $\boldsymbol{F} = \mathrm{diag}\,(f_1, \ldots, f_p)$ *with* $f_i > 0$ *for* $i = 1, \ldots, p$. *Then the following inequalities hold:*

Table 1: Values of (c_0, c_1, c_2) in $g_p(\varepsilon)$ for $p = 2$.

n	$m = 3$	$m = 4$	$m = 5$	$m = 7$
10	$(-0.31, -0.68, -0.36)$	$(-0.63, -1.39, -1.04)$	$(-1.56, -2.16, -1.71)$	$(-5.25, -3.89, -3.08)$
30	$(-0.12, -0.25, -0.13)$	$(-0.25, -0.51, -0.38)$	$(-0.61, -0.77, -0.63)$	$(-2.07, -1.33, -1.13)$
50	$(-0.08, -0.15, -0.08)$	$(-0.15, -0.31, -0.23)$	$(-0.38, -0.47, -0.39)$	$(-1.29, -0.80, -0.69)$
70	$(-0.06, -0.11, -0.06)$	$(-0.11, -0.22, -0.17)$	$(-0.28, -0.34, -0.28)$	$(-0.93, -0.57, -0.50)$
100	$(-0.04, -0.08, -0.04)$	$(-0.08, -0.16, -0.12)$	$(-0.19, -0.24, -0.20)$	$(-0.66, -0.40, -0.35)$
500	$(-0.01, -0.02, -0.01)$	$(-0.02, -0.03, -0.02)$	$(-0.04, -0.05, -0.04)$	$(-0.14, -0.08, -0.07)$

Table 2: Values of (c_0, c_1, c_2) in $g_p(\varepsilon)$ for $p = 5$.

n	$m = 6$	$m = 7$	$m = 8$
10	$(-7.40, -39.06, -39.82)$	$(-8.66, -50.50, -60.12)$	$(-12.95, -62.62, -82.05)$
30	$(-2.32, -11.43, -13.21)$	$(-3.03, -14.52, -19.86)$	$(-5.13, -17.71, -26.73)$
50	$(-1.35, -6.67, -7.95)$	$(-1.81, -8.43, -11.94)$	$(-3.16, -10.22, -16.01)$
70	$(-0.95, -4.71, -5.69)$	$(-1.29, -5.93, -8.54)$	$(-2.28, -7.18, -11.43)$
100	$(-0.66, -3.27, -3.99)$	$(-0.90, -4.11, -5.98)$	$(-1.61, -4.96, -8.00)$
500	$(-0.13, -0.64, -0.80)$	$(-0.18, -0.80, -1.20)$	$(-0.33, -0.97, -1.60)$

Table 3: Values of (c_0, c_1, c_2) in $g_p(\varepsilon)$ for $p = 10$.

	$m = 11$	$m = 15$	$m = 20$
30	$(-14.98, -151.33, -314.60)$	$(-53.57, -231.07, -616.90)$	$(-211.08, -341.37, -1059.39)$
50	$(-8.04, -79.15, -176.36)$	$(-34.34, -118.89, -341.29)$	$(-140.91, -172.50, -571.28)$
70	$(-5.43, -53.37, -122.72)$	$(-25.25, -79.44, -236.00)$	$(-105.57, -114.04, -389.99)$
100	$(-3.63, -35.79, -84.32)$	$(-18.09, -52.85, -161.35)$	$(-76.70, -75.18, -263.81)$
500	$(-0.66, -6.61, -16.32)$	$(-3.80, -9.59, -30.92)$	$(-16.52, -13.37, -49.42)$

(i) $\operatorname{tr}\boldsymbol{F}/p \geq |\boldsymbol{F}|^{1/p} \geq p/\operatorname{tr}\boldsymbol{F}^{-1}$,

(ii) $\operatorname{tr}\boldsymbol{F}\operatorname{tr}\boldsymbol{F}^{-1} \geq p^2$,

(iii) $\operatorname{tr}\boldsymbol{F}^2 \leq (\operatorname{tr}\boldsymbol{F})^2 \leq p\operatorname{tr}\boldsymbol{F}^2$,

where the equalities hold when $f_1 = \cdots = f_p$.

Proof. From the concavity of the log-function , we can show that $\log(\sum_{i=1}^p f_i/p) - (1/p)\sum_{i=1}^p \log f_i \geq 0$ and $\log(\sum_{i=1}^p f_i^{-1}/p) - (1/p)\sum_{i=1}^p \log f_i^{-1} \geq 0$, which proves (i). The inequality in (ii) follows from (i). For (iii), it is trivial to obtain the lower bound of $(\operatorname{tr}\boldsymbol{F})^2$. Noting that

$$\begin{aligned}(\operatorname{tr}\boldsymbol{F})^2 &= \operatorname{tr}\boldsymbol{F}^2 + 2\sum_{i<j} f_i f_j = p\operatorname{tr}\boldsymbol{F}^2 - (p-1)\operatorname{tr}\boldsymbol{F}^2 + 2\sum_{i<j} f_i f_j \\ &= p\operatorname{tr}\boldsymbol{F}^2 - \sum_{i<j}(f_i^2 + f_j^2) + 2\sum_{i<j} f_i f_j = p\operatorname{tr}\boldsymbol{F}^2 - \sum_{i<j}(f_i - f_j)^2,\end{aligned}$$

we can get the upper bound of $(\operatorname{tr}\boldsymbol{F})^2$. ∎

Proof of Theorem 4.1. Let $\phi_i = d_i/f_i + \beta_1/\operatorname{tr}\boldsymbol{F}$. Then for $m \geq p$, we can express the combined estimator $\widehat{\boldsymbol{\Theta}}^{CM}$ as $\widehat{\boldsymbol{\Theta}}^{CM} = \boldsymbol{X}(\boldsymbol{I}_p - \boldsymbol{Q}\boldsymbol{\Psi}^{CM}\boldsymbol{Q}^{-1})$ where $\boldsymbol{\Psi}^{CM} = \operatorname{diag}(\psi_1^{CM}, \ldots, \psi_p^{CM})$ with $\psi_i^{CM} = (1-\varepsilon)\phi_i + \varepsilon\beta_0/\operatorname{tr}\boldsymbol{F}$. Thus, using Lemma 3.3 and expanding $(\psi_i^{CM})^2$, we can write the unbiased estimate of the risk difference of $\widehat{\boldsymbol{\Theta}}^{CM}$ and $\widehat{\boldsymbol{\Theta}}^{ML}$ and decompose it as

$$\widehat{\Delta}_{CM} = \widehat{R}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\Theta}}^{CM}) - mp = \widehat{\Delta}_1 + \widehat{\Delta}_2 + \widehat{\Delta}_3 + \widehat{\Delta}_4,$$

where $\widehat{\Delta}_1$, $\widehat{\Delta}_2$, $\widehat{\Delta}_3$ and $\widehat{\Delta}_4$, respectively, correspond to the risk of $\widehat{\boldsymbol{\Theta}}^{MST}$, the risk of $\widehat{\boldsymbol{\Theta}}^{JS}$, the cross product terms independent of $\partial\varepsilon/\partial f_i$ and the cross product terms involving $\partial\varepsilon/\partial f_i$, given by

$$\begin{aligned}\widehat{\Delta}_1 &= \sum_{i=1}^p \Big[(n+p-3)(1-\varepsilon)^2 f_i\phi_i^2 - 2(m-p+1)(1-\varepsilon)\phi_i - 4(1-\varepsilon)^2 f_i^2\phi_i\frac{\partial\phi_i}{\partial f_i} \\ &\quad - 4(1-\varepsilon)f_i\frac{\partial\phi_i}{\partial f_i} - 4(1-\varepsilon)\sum_{j>i}\frac{f_i\phi_i - f_j\phi_j}{f_i - f_j} - 2(1-\varepsilon)^2\sum_{j>i}\frac{f_i^2\phi_i^2 - f_j^2\phi_j^2}{f_i - f_j}\Big], \\ \widehat{\Delta}_2 &= \sum_{i=1}^p \Big[(n+p-3)f_i\varepsilon^2\frac{\beta_0^2}{(\operatorname{tr}\boldsymbol{F})^2} - 4\varepsilon^2 f_i^2\frac{\beta_0}{\operatorname{tr}\boldsymbol{F}}\frac{\partial}{\partial f_i}\frac{\beta_0}{\operatorname{tr}\boldsymbol{F}} - 2\varepsilon^2\frac{\beta_0^2}{(\operatorname{tr}\boldsymbol{F})^2}\sum_{j>i}\frac{f_i^2 - f_j^2}{f_i - f_j} \\ &\quad - 2(m-p+1)\varepsilon\frac{\beta_0}{\operatorname{tr}\boldsymbol{F}} - 4\varepsilon f_i\frac{\partial}{\partial f_i}\frac{\beta_0}{\operatorname{tr}\boldsymbol{F}} - 4\varepsilon\frac{\beta_0}{\operatorname{tr}\boldsymbol{F}}\sum_{j>i}\frac{f_i - f_j}{f_i - f_j}\Big], \\ \widehat{\Delta}_3 &= 2\frac{\beta_0(1-\varepsilon)\varepsilon}{\operatorname{tr}\boldsymbol{F}}\sum_{i=1}^p \Big[(n+p-3)f_i\phi_i + 2\frac{f_i^2\phi_i}{\operatorname{tr}\boldsymbol{F}} - 2f_i^2\frac{\partial\phi_i}{\partial f_i} - 2\sum_{j>i}\frac{f_i^2\phi_i - f_j^2\phi_j}{f_i - f_j}\Big], \\ \widehat{\Delta}_4 &= 4\sum_{i=1}^p \Big\{1 + f_i\phi_i(1-\varepsilon) + \frac{\beta_0\varepsilon f_i}{\operatorname{tr}\boldsymbol{F}}\Big\}\Big\{f_i\phi_i - \frac{\beta_0 f_i}{\operatorname{tr}\boldsymbol{F}}\Big\}\frac{\partial\varepsilon}{\partial f_i}.\end{aligned}$$

We first evaluate $\widehat{\Delta}_1$. Expanding $\phi_i^2 = (d_i/f_i + \beta_1/\mathrm{tr}\,\boldsymbol{F})^2$ gives that

$$\widehat{\Delta}_1 = \widehat{\Delta}_{1\cdot 1} + \widehat{\Delta}_{1\cdot 2} + \widehat{\Delta}_{1\cdot 3},$$

where

$$\begin{aligned}
\widehat{\Delta}_{1\cdot 1} =& \sum_{i=1}^{p} \Big[(n+p+1)(1-\varepsilon)^2 \frac{d_i^2}{f_i} - 2(m-p-1)(1-\varepsilon)\frac{d_i}{f_i} \\
& \qquad - 4(1-\varepsilon)\sum_{j>i} \frac{d_i - d_j}{f_i - f_j} - 2(1-\varepsilon)^2 \sum_{j>i} \frac{d_i^2 - d_j^2}{f_i - f_j} \Big], \\
\widehat{\Delta}_{1\cdot 2} =& \frac{1}{\mathrm{tr}\,\boldsymbol{F}} \Big\{ (n-p-1)\beta_1^2(1-\varepsilon)^2 - 2(mp-2)\beta_1(1-\varepsilon) \Big\} + \frac{4\beta_1^2(1-\varepsilon)^2 \mathrm{tr}\,\boldsymbol{F}^2}{(\mathrm{tr}\,\boldsymbol{F})^3}, \\
\widehat{\Delta}_{1\cdot 3} =& 2\frac{\beta_1(1-\varepsilon)^2}{\mathrm{tr}\,\boldsymbol{F}} \Big\{ (n+p-1)\mathrm{tr}\,\boldsymbol{D} + 2\frac{\mathrm{tr}\,\boldsymbol{F}\boldsymbol{D}}{\mathrm{tr}\,\boldsymbol{F}} - 2\sum_{i=1}^{p}\sum_{j>i} \frac{f_i d_i - f_j d_j}{f_i - f_j} \Big\}.
\end{aligned}$$

Since $f_1 > \cdots > f_p$ and $d_1 > \cdots > d_p$, we observe that

$$\begin{aligned}
& 4(1-\varepsilon)\sum_{i=1}^{p}\sum_{j>i} \frac{d_i - d_j}{f_i - f_j} + 2(1-\varepsilon)^2 \sum_{i=1}^{p}\sum_{j>i} \frac{d_i^2 - d_j^2}{f_i - f_j} \\
& \geq 4(1-\varepsilon)\sum_{i=1}^{p}\sum_{j>i} \frac{d_i - d_j}{f_i} + 2(1-\varepsilon)^2 \sum_{i=1}^{p}\sum_{j>i} \frac{d_i^2 - d_j^2}{f_i} \\
& = \sum_{i=1}^{p} \frac{1}{f_i} \Big[4(1-\varepsilon)(p-i)d_i + 2(1-\varepsilon)^2(p-i)d_i^2 - \sum_{j>i} \Big\{ 4(1-\varepsilon)d_j + 2(1-\varepsilon)^2 d_j^2 \Big\} \Big],
\end{aligned}$$

which implies that $\widehat{\Delta}_{1\cdot 1} \leq \sum_{i=1}^{p} f_i^{-1} \widehat{\Delta}_{1\cdot 1}(i)$, where

$$\begin{aligned}
\widehat{\Delta}_{1\cdot 1}(i) =& (n-p+2i+1)(1-\varepsilon)^2 d_i^2 - 2(m+p-2i-1)(1-\varepsilon)d_i \\
& + \sum_{j>i} \Big\{ 4(1-\varepsilon)d_j + 2(1-\varepsilon)^2 d_j^2 \Big\}.
\end{aligned}$$

Denote $h(x) = (n-p+2i+1)x^2 - 2(m+p-2i-1)x$. Since $h(x)$ is minimized at $x = (m+p-2i-1)/(n-p+2i+1) = d_i$, we can see that

$$h(d_i) \leq h((1-\varepsilon)d_i) \leq h((1-\varepsilon)d_{i+1}).$$

Then for each i,

$$\begin{aligned}
\widehat{\Delta}_{1\cdot 1}(i) \leq & (n-p+2i+1)(1-\varepsilon)^2 d_{i+1}^2 - 2(m+p-2i-1)(1-\varepsilon)d_{i+1} \\
& + 4(1-\varepsilon)d_{i+1} + 2(1-\varepsilon)^2 d_{i+1}^2 + \sum_{j>i+1} \Big\{ 4(1-\varepsilon)d_j + 2(1-\varepsilon)^2 d_j^2 \Big\} \\
=& (n-p+2(i+1)+1)(1-\varepsilon)^2 d_{i+1}^2 - 2(m+p-2(i+1)-1)(1-\varepsilon)d_{i+1} \\
& + \sum_{j>i+1} \Big\{ 4(1-\varepsilon)d_j + 2(1-\varepsilon)^2 d_j^2 \Big\}.
\end{aligned}$$

Repeating this argument yields the inequality

$$\widehat{\Delta}_{1\cdot 1}(i) \le (n+p+1)(1-\varepsilon)^2 d_p^2 - 2(m-p-1)(1-\varepsilon)d_p,$$

which gives that

$$\begin{aligned}\widehat{\Delta}_{1\cdot 1} &\le \sum_{i=1}^{p} f_i^{-1}\Big[(n+p+1)(1-\varepsilon)^2 d_p^2 - 2(m-p-1)(1-\varepsilon)d_p\Big]\\ &= \mathrm{tr}\,\boldsymbol{F}^{-1}\Big[(m-p-1)(1-\varepsilon)^2 d_p - 2(m-p-1)(1-\varepsilon)d_p\Big].\end{aligned}$$

Since the last rhs of the above inequality is negative, it holds from Lemma 4.1 (ii) that

$$\widehat{\Delta}_{1\cdot 1} \le \frac{1}{\mathrm{tr}\,\boldsymbol{F}}\Big[p^2(m-p-1)(1-\varepsilon)^2 d_p - 2p^2(m-p-1)(1-\varepsilon)d_p\Big]. \tag{4.3}$$

For $\widehat{\Delta}_{1\cdot 2}$, it follows from the inequality $\mathrm{tr}\,\boldsymbol{F}^2 \le (\mathrm{tr}\,\boldsymbol{F})^2$ that

$$\widehat{\Delta}_{1\cdot 2} \le \frac{1}{\mathrm{tr}\,\boldsymbol{F}}\Big\{(n-p+3)\beta_1^2(1-\varepsilon)^2 - 2(mp-2)\beta_1(1-\varepsilon)\Big\}. \tag{4.4}$$

The same arguments as in the proof of Theorem 3.1 can be used to get that

$$\widehat{\Delta}_{1\cdot 3} \le 2\frac{\beta_1(1-\varepsilon)^2}{\mathrm{tr}\,\boldsymbol{F}}\Big\{mp - 2p - 2\sum_{i=2}^{p} d_i\Big\}. \tag{4.5}$$

From the definition of β_1, note that $2(p-1+\sum_{i=2}^{p} d_i) = (n-p+3)\beta_1$. Thus, combining (4.3), (4.4) and (4.5) gives that

$$\begin{aligned}\widehat{\Delta}_1 \le{}& (\mathrm{tr}\,\boldsymbol{F})^{-1}\big\{p^2(m-p-1)(1-\varepsilon)^2 d_p - 2p^2(m-p-1)(1-\varepsilon)d_p\big\}\\ &+ (\mathrm{tr}\,\boldsymbol{F})^{-1}\big\{(n-p+3)\beta_1^2(1-\varepsilon)^2 - 2(mp-2)\beta_1(1-\varepsilon)\big\}\\ &+ 2\frac{\beta_1(1-\varepsilon)^2}{\mathrm{tr}\,\boldsymbol{F}}\Big\{mp-2p-2\sum_{i=2}^{p} d_i\Big\}\\ ={}& (\mathrm{tr}\,\boldsymbol{F})^{-1}\big[-p^2(m-p-1)d_p - \beta_1^2(n-p+3)\\ &\quad + 2\varepsilon\{\beta_1^2(n-p+3) - (mp-2)\beta_1\}\\ &\quad + \varepsilon^2\{p^2(m-p-1)d_p + 2(mp-2)\beta_1 - \beta_1^2(n-p+3)\}\big].\end{aligned} \tag{4.6}$$

Using the fact that $\mathrm{tr}\,\boldsymbol{F}^2 \le (\mathrm{tr}\,\boldsymbol{F})^2$, we can evaluate $\widehat{\Delta}_2$ as

$$\begin{aligned}\widehat{\Delta}_2 &= (\mathrm{tr}\,\boldsymbol{F})^{-1}\big\{(n-p-1)\beta_0^2\varepsilon^2 - 2(mp-2)\beta_0\varepsilon\big\} + \frac{4\beta_0^2\varepsilon^2\mathrm{tr}\,\boldsymbol{F}^2}{(\mathrm{tr}\,\boldsymbol{F})^3}\\ &\le (\mathrm{tr}\,\boldsymbol{F})^{-1}\big\{(n-p+3)\beta_0^2\varepsilon^2 - 2(mp-2)\beta_0\varepsilon\big\}\\ &= (\mathrm{tr}\,\boldsymbol{F})^{-1}\big\{(mp-2)\beta_0\varepsilon^2 - 2(mp-2)\beta_0\varepsilon\big\}.\end{aligned} \tag{4.7}$$

By the same manner as in the proof of Theorem 3.1, $\widehat{\Delta}_3$ can be evaluated as

$$\begin{aligned}\widehat{\Delta}_3 =& 2\frac{\beta_0(1-\varepsilon)\varepsilon}{\mathrm{tr}\,\boldsymbol{F}}\Big\{(n+p-1)\mathrm{tr}\,\boldsymbol{D} + 2\frac{\mathrm{tr}\,\boldsymbol{F}\boldsymbol{D}}{\mathrm{tr}\,\boldsymbol{F}} - 2\sum_{i=1}^{p}\sum_{j>i}\frac{f_i d_i - f_j d_j}{f_i - f_j}\Big\}\\ &+ 2\frac{\beta_0(1-\varepsilon)\varepsilon}{\mathrm{tr}\,\boldsymbol{F}}\Big\{(n-p-1)\beta_1 + 4\beta_1\frac{\mathrm{tr}\,\boldsymbol{F}^2}{(\mathrm{tr}\,\boldsymbol{F})^2}\Big]\\ \le& 2\frac{\beta_0(1-\varepsilon)\varepsilon}{\mathrm{tr}\,\boldsymbol{F}}\Big[mp - 2p - 2\sum_{i=2}^{p} d_i + (n-p+3)\beta_1\Big\}\\ =& 2(mp-2)\frac{\beta_0(1-\varepsilon)\varepsilon}{\mathrm{tr}\,\boldsymbol{F}}. \end{aligned} \tag{4.8}$$

Combining (4.6), (4.7) and (4.8) gives that

$$\widehat{\Delta}_1 + \widehat{\Delta}_2 + \widehat{\Delta}_3 = (\mathrm{tr}\,\boldsymbol{F})^{-1} \times h_p(\varepsilon),$$

where $h_p(\varepsilon)$ is a quadratic function of $\varepsilon \in [0, 1]$ defined by

$$\begin{aligned}h_p(\varepsilon) =& -p^2(m-p-1)d_p - \beta_1^2(n-p+3) + \varepsilon\left\{2\beta_1^2(n-p+3) - 2(mp-2)\beta_1\right\}\\ &+ \varepsilon^2\left\{p^2(m-p-1)d_p + 2(mp-2)\beta_1 - \beta_1^2(n-p+3) - (mp-2)\beta_0\right\}. \end{aligned} \tag{4.9}$$

Noting that $\beta_0 - \beta_1 > 0$, we observe that $h_p(0) < 0$, $h_p(1) < 0$ and $h'_p(0) < 0$ for $h'_p(\varepsilon) = (\mathrm{d}/\mathrm{d}\varepsilon)h_p(\varepsilon)$. These facts imply that

$$\sup_{0<\varepsilon<1} h_p(\varepsilon) < 0. \tag{4.10}$$

It is noted that $\widehat{\Delta}_1 + \widehat{\Delta}_2 + \widehat{\Delta}_3 \le 0$ irrespective of the specific form of the function $\varepsilon = \varepsilon(\boldsymbol{F})$. The function ε affects the term $\widehat{\Delta}_4$. Since $\partial\varepsilon(\boldsymbol{F})/\partial f_i = -\gamma \times \varepsilon(\boldsymbol{F})$, $\widehat{\Delta}_4$ is rewritten by

$$\widehat{\Delta}_4 = 4\gamma\varepsilon\sum_{i=1}^{p}\left\{1 + (1-\varepsilon)\left(d_i + \frac{\beta_1 f_i}{\mathrm{tr}\,\boldsymbol{F}}\right) + \varepsilon\frac{\beta_0 f_i}{\mathrm{tr}\,\boldsymbol{F}}\right\} \times \left\{(\beta_0 - \beta_1)\frac{f_i}{\mathrm{tr}\,\boldsymbol{F}} - d_i\right\},$$

which can be expressed by

$$\begin{aligned}\widehat{\Delta}_4 = 4\gamma\varepsilon\Big[&\beta_0 - \beta_1 - \sum_{i=1}^{p} d_i\\ &+ (1-\varepsilon)\Big\{(\beta_0-\beta_1)\frac{\mathrm{tr}\,\boldsymbol{F}\boldsymbol{D}}{\mathrm{tr}\,\boldsymbol{F}} - \sum_{i=1}^{p} d_i^2 + \beta_1(\beta_0-\beta_1)\frac{\mathrm{tr}\,\boldsymbol{F}^2}{(\mathrm{tr}\,\boldsymbol{F})^2} - \beta_1\frac{\mathrm{tr}\,\boldsymbol{F}\boldsymbol{D}}{\mathrm{tr}\,\boldsymbol{F}}\Big\}\\ &+ \varepsilon\Big\{\beta_0(\beta_0-\beta_1)\frac{\mathrm{tr}\,\boldsymbol{F}^2}{(\mathrm{tr}\,\boldsymbol{F})^2} - \beta_0\frac{\mathrm{tr}\,\boldsymbol{F}\boldsymbol{D}}{\mathrm{tr}\,\boldsymbol{F}}\Big\}\Big]. \end{aligned}$$

Since $\beta_0 - \beta_1 \geq 0$, $\mathrm{tr}\, \boldsymbol{F}^2/(\mathrm{tr}\, \boldsymbol{F})^2 \leq 1$ and $d_p \leq \mathrm{tr}\, \boldsymbol{F}\boldsymbol{D}/\mathrm{tr}\, \boldsymbol{F} \leq d_1$, we can show that

$$\begin{aligned}
\widehat{\Delta}_4 \leq\, & 4\gamma\varepsilon\Big[\beta_0 - \beta_1 - \sum_{i=1}^{p} d_i + (1-\varepsilon)\Big\{(\beta_0 - \beta_1)d_1 - \sum_{i=1}^{p} d_i^2 - \beta_1 d_p + \beta_1(\beta_0 - \beta_1)\Big\} \\
& + \varepsilon\Big\{\beta_0(\beta_0 - \beta_1) - \beta_0 d_p\Big\}\Big] \\
=\, & 4\gamma\varepsilon \times K(\varepsilon),
\end{aligned}$$

where $K(\varepsilon) = (\beta_0 - \beta_1)(1 + d_1 + \beta_1) - \sum_{i=1}^{p} d_i - \sum_{i=1}^{p} d_i^2 - \beta_1 d_p + \varepsilon\{(\beta_0 - \beta_1)^2 - (\beta_0 - \beta_1)(d_1 + d_p) + \sum_{i=1}^{p} d_i^2\}$. If $K(\varepsilon)$ is negative, then from (4.10), it follows that $\widehat{\Delta}_{CM} \leq 0$. If $K(\varepsilon)$ is non-negative, from the fact that $e^{-x} \leq e^{-1}x^{-1}$ for $x > 0$, we observe that

$$\widehat{\Delta}_4 \leq \frac{(4/e)}{\mathrm{tr}\, \boldsymbol{F}} \times K(\varepsilon). \tag{4.11}$$

Combining (4.9) and (4.11), we can see that

$$\widehat{\Delta}_{CM} \leq (\mathrm{tr}\, \boldsymbol{F})^{-1}\left[h_p(\varepsilon) + (4/e)K(\varepsilon)\right] = (\mathrm{tr}\, \boldsymbol{F})^{-1} \times g_p(\varepsilon),$$

where

$$\begin{aligned}
g_p(\varepsilon) =& -p^2(m-p-1)d_p - \beta_1^2(n-p+3) \\
& \qquad + (4/e)\Big\{(\beta_0 - \beta_1)(1 + d_1 + \beta_1) - \sum_{i=1}^{p} d_i - \sum_{i=1}^{p} d_i^2 - \beta_1 d_p\Big\} \\
& + \varepsilon\Big[2\beta_1^2(n-p+3) - 2(mp-2)\beta_1 \\
& \qquad + (4/e)\Big\{(\beta_0 - \beta_1)^2 - (\beta_0 - \beta_1)(d_1 + d_p) + \sum_{i=1}^{p} d_i^2\Big\}\Big] \\
& + \varepsilon^2\big\{p^2(m-p-1)d_p + 2(mp-2)\beta_1 - \beta_1^2(n-p+3) - (mp-2)\beta_0\big\},
\end{aligned}$$

which is rewritten by the expression (4.1). Therefore, if

$$\sup_{0<\varepsilon<1} g_p(\varepsilon) < 0$$

for a pair of (n, m, p), then the estimator $\widehat{\boldsymbol{\Theta}}^{CM}$ is minimax. ■

5 Monte Carlo studies

We now investigate the risk performances of several minimax estimators derived in the previous sections. The values of the risks are estimated through a Monte-Carlo simulation for $m \geq p$.

The estimators we shall investigate are
[1] the James-Stein estimator $\widehat{\boldsymbol{\Theta}}^{JS} = (1 - \beta_0/\mathrm{tr}\, \boldsymbol{F})\boldsymbol{X}$ for $\beta_0 = (mp-2)/(n-p+3)$,
[2] the modified Efron-Morris estimator $\widehat{\boldsymbol{\Theta}}^{MEM} = \widehat{\boldsymbol{\Theta}}^{EM} - (\beta_0'/\mathrm{tr}\, \boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1})\boldsymbol{X}$

for $\beta'_0 = (p-1)(p+2)(m+n)/\{(n+p+1)(n-p+3)\}$,

[3] the modified Stein estimator $\widehat{\mathbf{\Theta}}^{MST} = \widehat{\mathbf{\Theta}}^{ST} - (\beta_1/\mathrm{tr}\, \boldsymbol{X}^t\boldsymbol{X}\boldsymbol{S}^{-1})\boldsymbol{X}$ for $\beta_1 = 2(p-1+\sum_{i=2}^p d_i)/(n-p+3)$,

[4] the modified Dey estimator $\widehat{\mathbf{\Theta}}^{MDY} = \boldsymbol{X}(\boldsymbol{I}_p - \alpha_1\boldsymbol{Q}\boldsymbol{F}\boldsymbol{Q}^{-1}/\mathrm{tr}\, \boldsymbol{F}^2) - (\beta_2/\mathrm{tr}\, \boldsymbol{F})\boldsymbol{X}$ for $\alpha_1 = (m+p-3)/(n-p+3)$ and $\beta_2 = (m-1)(p-1)/(n-p+3)$,

[5] the modified Haff estimator $\widehat{\mathbf{\Theta}}^{MHF} = \boldsymbol{X}(\boldsymbol{I}_p - \alpha_0\boldsymbol{Q}(\boldsymbol{F}+\delta/\mathrm{tr}\, \boldsymbol{F}^{-1})^{-1}\boldsymbol{Q}^{-1}) - (\beta_3/\mathrm{tr}\, \boldsymbol{F})\boldsymbol{X}$ for $\alpha_0 = (m-p-1)/(n+p+1)$, $\beta_3 = (p-1)(p+2)/(n-p+3)$ and $\delta = 0.01$,

[6] the combined estimator $\widehat{\mathbf{\Theta}}^{CM} = (1-\varepsilon)\widehat{\mathbf{\Theta}}^{MST} + \varepsilon\widehat{\mathbf{\Theta}}^{JS}$ for $\varepsilon = \exp(-\gamma \mathrm{tr}\, \boldsymbol{F})$ and $\gamma = (n-p-1)/mp$, and

[7] the empirical Bayes estimator recommended by Shieh (1993):

$$\widehat{\mathbf{\Theta}}^{SH} = \boldsymbol{X}\{\boldsymbol{I}_p - \boldsymbol{T}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{S}/(n+p+1)\},$$
$$\boldsymbol{T} = (1-\alpha_2)\Big\{(m-p-1)\boldsymbol{I}_p + \frac{p-1}{\mathrm{tr}\, \boldsymbol{X}^t\boldsymbol{X}}\boldsymbol{X}^t\boldsymbol{X}\Big\} + \alpha_2\frac{mp-2}{\mathrm{tr}\, \boldsymbol{X}^t\boldsymbol{X}}\boldsymbol{X}^t\boldsymbol{X},$$

for $\alpha_2 = (m-p-1)/\{m^2(p-1)\}$.

The simulation experiments are done based on 50,000 independent replications generated from (1.1). The risk functions are estimated by the average of the simulated values of the risks, and their estimated risks are reported by Table 4 for $m = 4$, 8, 12 and $p = 2$, and by Table 5 for $m = 6$, 12, 18 and $p = 4$, where $n = 5m$. Since the risk functions of the above estimators are functions of $\mathbf{\Theta}^t\mathbf{\Theta}\mathbf{\Sigma}^{-1}$, we look into the two cases of eigenvalues of $\mathbf{\Theta}^t\mathbf{\Theta}\mathbf{\Sigma}^{-1}$, namely, we choose $(0,0)$ and $(100,1)$ for $p = 2$ and $(0,0,0,0)$ and $(100,10,1,0)$ for $p = 4$ as the eigenvalues.

The risk behaviors of the five estimators $\widehat{\mathbf{\Theta}}^{ML}$, $\widehat{\mathbf{\Theta}}^{MEM}$, $\widehat{\mathbf{\Theta}}^{MST}$, $\widehat{\mathbf{\Theta}}^{JS}$, $\widehat{\mathbf{\Theta}}^{CM}$ are drawn in Figure 1 for $(n,m,p) = (10,8,4)$, where the eigenvalues of $\mathbf{\Theta}^t\mathbf{\Theta}\mathbf{\Sigma}^{-1}$ take the values of (4^c, 2^c, 1, 0) for $0 \le c \le 5$. In the tables and the figure, $\mathrm{Ch}(\mathbf{\Theta}^t\mathbf{\Theta}\mathbf{\Sigma}^{-1})$ denotes the eigenvalues of $\mathbf{\Theta}^t\mathbf{\Theta}\mathbf{\Sigma}^{-1}$, and for the simplicity the estimators $\widehat{\mathbf{\Theta}}^{MEM}$, $\widehat{\mathbf{\Theta}}^{MDY}$, $\widehat{\mathbf{\Theta}}^{MHF}$ $\widehat{\mathbf{\Theta}}^{MST}$, $\widehat{\mathbf{\Theta}}^{JS}$, $\widehat{\mathbf{\Theta}}^{CM}$ and $\widehat{\mathbf{\Theta}}^{SH}$ are denoted by MEM, MDY, MHF, MST, JS, CM and SH, respectively. Also $\widehat{\mathbf{\Theta}}^{ML} = \boldsymbol{X}$ is denoted by ML.

It is noted that the values of (n,m,p) in the above studies satisfy the minimaxity of the estimators. Especially the condition of Theorem 4.1 is satisfied and the minimaxity of the combined estimator $\widehat{\mathbf{\Theta}}^{CM}$ is guaranteed.

The numerical results given in Tables 4 and 5 and Figure 1 illustrate several important observations.

(1) When the eigenvalues of $\mathbf{\Theta}^t\mathbf{\Theta}\mathbf{\Sigma}^{-1}$ are zeros, $\widehat{\mathbf{\Theta}}^{JS}$, $\widehat{\mathbf{\Theta}}^{MDY}$ and $\widehat{\mathbf{\Theta}}^{CM}$ are more favorable than the others. When the eigenvalues of $\mathbf{\Theta}^t\mathbf{\Theta}\mathbf{\Sigma}^{-1}$ are dispersed, on the other hand, $\widehat{\mathbf{\Theta}}^{MEM}$, $\widehat{\mathbf{\Theta}}^{MST}$, $\widehat{\mathbf{\Theta}}^{CM}$ and $\widehat{\mathbf{\Theta}}^{SH}$ are better.

(2) The risk behavior of $\widehat{\mathbf{\Theta}}^{MDY}$ and $\widehat{\mathbf{\Theta}}^{MHF}$ are similar to that of $\widehat{\mathbf{\Theta}}^{JS}$. The risk of $\widehat{\mathbf{\Theta}}^{SH}$ is not favorable than the others when the eigenvalues of $\mathbf{\Theta}^t\mathbf{\Theta}\mathbf{\Sigma}^{-1}$ are zeros.

(3) $\widehat{\mathbf{\Theta}}^{CM}$ is superior to either $\widehat{\mathbf{\Theta}}^{JS}$ or $\widehat{\mathbf{\Theta}}^{MST}$.

(4) On the whole, for a fixed p, the savings in risk increase with m (and $n = 5m$).

(5) Figure 1 indicates that for small sample size n, $\widehat{\mathbf{\Theta}}^{MST}$ is better than $\widehat{\mathbf{\Theta}}^{MEM}$. Also, $\widehat{\mathbf{\Theta}}^{CM}$ has a smaller risk than both $\widehat{\mathbf{\Theta}}^{JS}$ and $\widehat{\mathbf{\Theta}}^{MST}$ when the eigenvalues of $\mathbf{\Theta}^t\mathbf{\Theta}\mathbf{\Sigma}^{-1}$ are close together.

Through the Monte Carlo simulation studies, we come to the conclusions that the combined estimator $\widehat{\mathbf{\Theta}}^{CM}$ has an excellent risk behavior such that $\widehat{\mathbf{\Theta}}^{CM}$ inherits the nice risk properties of both the estimators $\widehat{\mathbf{\Theta}}^{JS}$ and $\widehat{\mathbf{\Theta}}^{MST}$. Of course, there is no estimator which has the best risk behavior over the whole parameter space.

Remark 5.1 For small sample case, we also carried out Monte Calro studies when $(n, m, p) = (10, 6, 4)$, $(8, 6, 4)$ and $(6, 4, 2)$, and others. In such cases, we observed that $\widehat{\mathbf{\Theta}}^{CM}$ has smaller risks than both $\widehat{\mathbf{\Theta}}^{JS}$ and $\widehat{\mathbf{\Theta}}^{MST}$.

Table 4: Simulated risks in estimation of mean matrix ($p = 2$ and $n = 5m$).

Ch ($\mathbf{\Theta}^t\mathbf{\Theta}\mathbf{\Sigma}^{-1}$)	$(0, 0)$			$(100, 1)$		
(m, p)	$(4, 2)$	$(8, 2)$	$(12, 2)$	$(4, 2)$	$(8, 2)$	$(12, 2)$
ML	7.96	16.01	23.99	7.96	16.01	23.99
MEM	3.45	5.32	5.93	7.16	11.52	15.15
MDY	2.57	2.80	2.92	7.78	15.03	21.58
MHF	3.05	7.83	12.93	7.64	14.33	20.57
MST	3.24	4.60	5.13	7.23	11.59	15.22
JS	2.41	2.45	2.44	7.65	14.37	20.12
CM	2.48	2.67	2.76	7.23	11.57	15.18
SH	4.84	5.75	6.13	7.24	11.59	15.21

Table 5: Simulated risks in estimation of mean matrix ($p = 4$ and $n = 5m$).

Ch($\mathbf{\Theta}^t\mathbf{\Theta}\mathbf{\Sigma}^{-1}$)	$(0, 0, 0, 0)$			$(100, 10, 1, 0)$		
(m, p)	$(6, 4)$	$(12, 4)$	$(18, 4)$	$(6, 4)$	$(12, 4)$	$(18, 4)$
ML	23.99	47.95	71.93	23.99	47.95	71.93
MEM	5.50	13.63	16.82	19.76	29.29	36.44
MDY	3.35	3.35	3.41	21.45	37.93	50.86
MHF	3.79	15.66	30.15	20.47	36.35	51.77
MST	6.29	11.65	13.91	18.28	28.00	35.13
JS	2.84	2.65	2.58	20.50	34.65	45.00
CM	3.23	3.87	4.19	18.23	27.18	33.75
SH	16.41	19.66	20.96	21.55	31.06	38.06

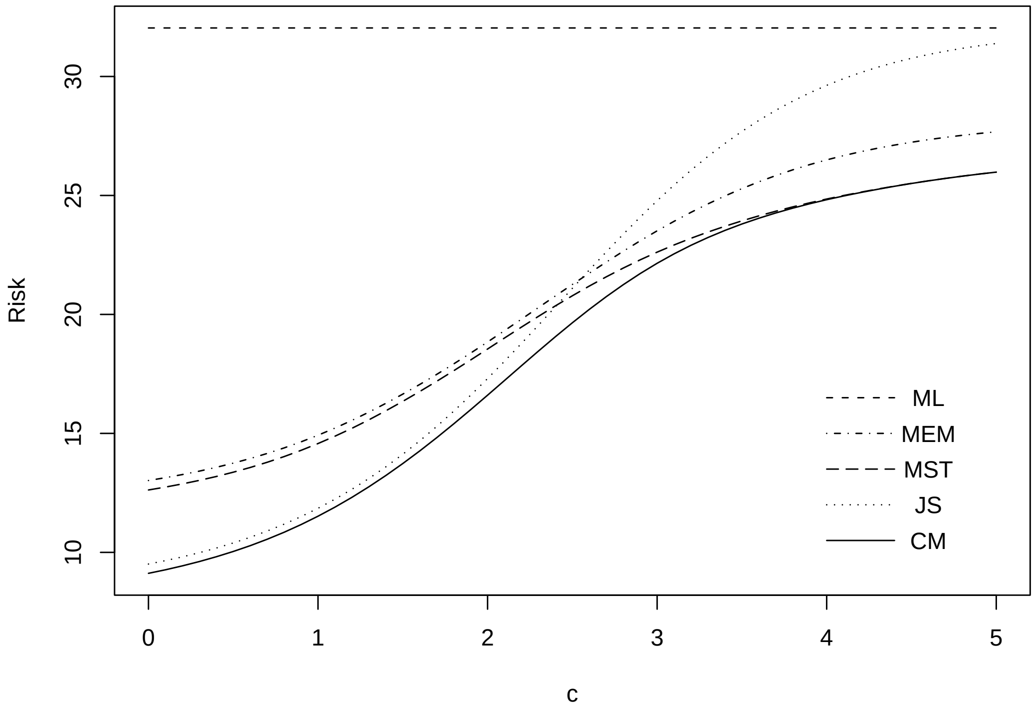

Figure 1: Simulated risks in estimation of mean matrix where $(n,\, m,\, p) = (10,\, 8,\, 4)$ and the eigenvalues of $\mathbf{\Sigma}^{-1}\mathbf{\Theta}^t\mathbf{\Theta}$ are $(4^c,\, 2^c,\, 1,\, 0)$ for $0 \le c \le 5$.

Acknowledgments. The research of the first author was supported in part by Grant-in-Aid for Scientific Research No. 1610018. The research of the second author was supported in part by Grant-in-Aid for Scientific Research Nos. 15200021, 15200022 and 16500172 and in part by a grant from the 21st Century COE Program at Faculty of Economics, University of Tokyo.

REFERENCES

Berger, J.O., Strawderman, W., and Tang, D. (2005). Posterior propriety and admissibility of hyperpriors in normal hierarchical models. *Ann. Statist.*, **33**, 606-646.

Bilodeau, M. and Kariya, T. (1989). Minimax estimators in the normal MANOVA model. *J. Multivariate Anal.*, **28**, 260–270.

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

Efron, B. and Morris, C. (1972). Empirical Bayes on vector observations: an extension of Stein's method. *Biometrika*, **59**, 335–347.

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

Ghosh, M. and Shieh, G. (1991). Empirical Bayes minimax estimators of matrix normal means. *J. Multivariate Anal.*, **38**, 306–318.

Ghosh, M. and Shieh, G. (1992). Empirical Bayes minimax estimators of matrix normal means -II. In *Nonparametric Statistics and Related Topics* (A.K.Md.E. Saleh, ed.), 231–251.

Gupta, A.K. and Nagar, D.K. (1999). *Matrix variate distributions*, Chapman & Hall/CRC.

Haff, L.R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.*, **8**, 586–597.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Sympos. Math. Statist. and Prob.* Vol.I, pp. 361–379, Univ. California Press, Berkeley.

Kiefer, J. and Schwartz, R. (1965). Admissible Bayes character of T^2-, R^2-, and other fully invariant tests for classical multivariate normal problems. *Ann. Math. Statist.*, **36**, 747–770.

Konno, Y. (1990). Families of minimax estimators of matrix of normal means with unknown covariance matrix. *J. Japan Statist. Soc.*, **20**, 191–201.

Konno, Y. (1991). On estimation of a matrix of normal means with unknown covariance matrix. *J. Multivariate Anal.*, **36**, 44–55.

Konno, Y. (1992). Improved estimation of matrix of normal mean and eigenvalues in the multivariate F-distribution. Doctoral dissertation, Institute of Mathematics, Univ. Tsukuba.

Loh, W.-L. (1988). Estimating covariance matrices. Ph. D. dissertation, Dept. Statist., Stanford Univ.

Loh, W.-L. (1991). Estimating covariance matrices II. *J. Multivariate Anal.*, **36**, 163–174.

Shieh, G. (1993). Empirical Bayes minimax estimators of matrix normal means for arbitrary quadratic loss and unknown covariance matrix. *Statist. Decisions*, **11**, 317–341.