CIRJE-F-937

Unified Improvements in Estimation of a Normal Covariance Matrix in High and Low Dimensions

Hisayuki Tsukuma
Toho University

Tatsuya Kubokawa
The University of Tokyo

August 2014

CIRJE Discussion Papers can be downloaded without charge from:

http://www.cirje.e.u-tokyo.ac.jp/research/03research02dp.html

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.

Unified Improvements in Estimation of a Normal Covariance Matrix in High and Low Dimensions

Hisayuki Tsukuma* and Tatsuya Kubokawa†

August 19, 2014

Abstract

The problem of estimating a covariance matrix in multivariate linear regression models is addressed in a decision-theoretic framework. Although a standard loss function is the Stein loss, it is not available in the case of a high dimension. In this paper, a new type of a quadratic loss function, called the intrinsic loss, is suggested, and unified dominance results are derived under the loss, irrespective of order of the dimension, the sample size and the rank of the regression coefficients matrix. Especially, using the Stein-Haff identity, we develop a key inequality which is useful for constructing a truncated and improved estimator based on the information contained in the sample means or the ordinary least squares estimator of the regression coefficients.

AMS 2010 subject classifications: Primary 62F11, 62J12, Secondary 62C15, 62C20.

Key words and phrases: high dimension, inadmissibility, invariant loss, Moore-Penrose inverse, statistical decision theory.

1 Introduction

The problems of estimating the covariance matrix in multivariate linear regression models are addressed in a decision-theoretic framework. The dominance properties of truncated estimators over non-truncated and unbiased estimators have been studied in Sinha and Ghosh (1987), Kubokawa and Srivastava (2003) and Kubokawa and Tsai (2006). These are multivariate extensions of Stein (1964) who established that the best location-scale equivariant estimator of a normal variance is dominated by the truncated estimator using the information contained in a sample mean. All the dominance results have been derived

*Faculty of Medicine, Toho University, 5-21-16 Omori-nishi, Ota-ku, Tokyo 143-8540, Japan, E-Mail: tsukuma@med.toho-u.ac.jp

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

when the dimension p of the covariance matrix is less than the degrees of freedom n. In this paper, we want to establish unified dominance results which cover both cases of $p > n$ and $n \geq p$.

To explain the problem specifically, let us consider a canonical model of the multivariate linear regression model. Let $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m)^t$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)^t$ be, respectively, $m \times p$ and $n \times p$ random matrices, where $\boldsymbol{X}_i$'s and $\boldsymbol{Y}_i$'s are mutually and independently distributed as

$$\begin{aligned} \boldsymbol{X}_i &\sim \mathcal{N}_p(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}), \quad i = 1, \ldots, m, \\ \boldsymbol{Y}_j &\sim \mathcal{N}_p(\boldsymbol{0}_p, \boldsymbol{\Sigma}), \quad j = 1, \ldots, n. \end{aligned} \tag{1.1}$$

Suppose that $\boldsymbol{\theta}_i$'s are unknown mean vectors and that $\boldsymbol{\Sigma}$ is an unknown positive definite matrix.

Let $\boldsymbol{V} = \boldsymbol{Y}^t\boldsymbol{Y} = \sum_{i=1}^n \boldsymbol{Y}_i\boldsymbol{Y}_i^t$. Then, $\boldsymbol{V}$ has a Wishart distribution $\mathcal{W}_p(n, \boldsymbol{\Sigma})$ for $n \geq p$, but a singular Wishart distribution for $p > n$ (see Srivastava (2003)). Our primary interest is in estimation of the covariance matrix $\boldsymbol{\Sigma}$ based on $(\boldsymbol{V}, \boldsymbol{X})$ and in derivation of unified dominance results irrespective of order of n, p and m in a decision-theoretic framework. In the case of $n \geq p$, a standard loss function is the Stein loss given by

$$L_S(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{tr}\, \widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \log |\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p, \tag{1.2}$$

which is easier to handle than a quadratic loss $\mathrm{tr}[(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_p)^2]$. Also, the unbiased estimator $n^{-1}\boldsymbol{V}$ is the best among estimators $c\boldsymbol{V}$ for positive constant c. In the case of $p > n$, however, the Stein loss is not available, since $n^{-1}\boldsymbol{V}$ is singular. Thus, in this paper, we suggest a new intrinsic loss function given by

$$L_V(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{tr}[\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}} - \boldsymbol{I}_p)^2\boldsymbol{V}^+] = \mathrm{tr}[\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{V}^+\widehat{\boldsymbol{\Sigma}}] - 2\mathrm{tr}[\widehat{\boldsymbol{\Sigma}}\boldsymbol{V}^+] + \mathrm{tr}[\boldsymbol{\Sigma}\boldsymbol{V}^+], \tag{1.3}$$

where $\boldsymbol{V}^+$ is the Moore-Penrose inverse of $\boldsymbol{V}$. The intrinsic loss corresponds to the loss derive by substituting $\boldsymbol{V}^+$ into one of two $\boldsymbol{\Sigma}^{-1}$ in the quadratic loss. It is interesting to point out the following properties of the intrinsic loss (1.3).

(1) In the case of $n \geq p$, there are several similar properties between the losses (1.2) and (1.3). First, the unbiased estimator $n^{-1}\boldsymbol{V}$ is the best location-equivariant under the two losses. Secondly, the unbiased estimator can be improved on by the same James-Stein (1961) estimator under the two losses. Thirdly, the Bayes estimator of $\boldsymbol{\Sigma}$ is of the same form $(E[\boldsymbol{\Sigma}^{-1}|\boldsymbol{V}])^{-1}$ under the two losses.

(2) The decision-theoretic results derived in the case of $n \geq p$ can be extended to the case of $p > n$ under the intrinsic loss with exchanging n and p.

(3) The terms which we need to evaluate analytically under the two loss functions (1.2) and (1.3) are $E[\mathrm{tr}[\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}]]$ and $E[\mathrm{tr}[\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{V}^+\widehat{\boldsymbol{\Sigma}}]]$, while we need to evaluate the term $E[\mathrm{tr}[\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}]]$ for the quadratic loss. This shows that the two losses are easier to treat analytically than the quadratic loss.

The main objective of this paper is the derivation of unified dominance results that estimators of $\boldsymbol{\Sigma}$ can be improved on by truncated estimators based on the information contained in $\boldsymbol{X}$, irrespective of order among n, p and m. Such a dominance result was first established by Stein (1964), and several extensions to the multivariate models were studied by Sinha and Ghosh (1987), Perron (1990), Kubokawa, Robert and Saleh (1992) and Kubokawa and Srivastava (2003) in the case of $n \geq p$. These articles applied conditional arguments to deriving the dominance results. Kubokawa and Tsai (2006) suggested a new method based on the Stein-Haff identity developed by Stein (1977) and Haff (1980) for $n \geq p$. In this paper, we use the same method to extend the dominance results to the case of $p > n$ under the intrinsic loss.

The paper is organized as follows: In Section 2, we illustrate several important points on how similar the intrinsic loss (1.3) is to the Stein loss (1.2). In the univariate case of $p = 1$, the unbiased estimator of σ^2 can be improved on by a common estimator under the same conditions relative to the two losses. In the multivariate case, the unified James-Stein type estimator is developed for the two cases of $n \geq p$ and $p > n$ relative to the intrinsic loss. This estimator is identical to the James-Stein (1961) estimator under the Stein loss for $n \geq p$.

In Section 3, we analytically derive unified dominance results that estimators of $\boldsymbol{\Sigma}$ can be improved on by truncated estimators based on the information contained in $\boldsymbol{X}$, irrespective of order among n, p and m. The main issue in Section 3 from a technical point of view is the derivation of a key inequality to showing the dominance. Also, some numerical results of simulation studies are provided for the risk functions of several truncated estimators. The numerical results show nice performances of the truncated estimators for various n, p and m.

In Section 4, we extend the results to the estimation of the covariance matrix in linear mixed models and to the estimation of the precision matrix. Concerning the former issue, the covariance matrix $\boldsymbol{\Sigma}$ corresponds to the 'within' component of variance. Although the estimation of variance components in univariate random effects models have been studies in many articles, multivariate cases have been discussed in several articles including Amemiya (1985), Calvin and Dykstra (1991), Mathew, Niyogi and Sinha (1994) and Srivastava and Kubokawa (1999). The results given in Section 3 can be applied to this problem.

2 Similarity between the Intrinsic and the Stein Losses

2.1 A univariate case

In the univariate case of $p = 1$, let $V = \sum_{i=1}^{n} Y_i^2$ and $\boldsymbol{X} = (X_1, \ldots, X_m)^t$ in the model (1.1). Then, $V/\sigma^2 \sim \chi_n^2$ and $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\theta}, \sigma^2 \boldsymbol{I}_m)$. The Stein loss and the intrinsic loss functions are described as

$$
\begin{aligned}
L_S(\hat{\sigma}^2, \sigma^2) &= \hat{\sigma}^2/\sigma^2 - \log(\hat{\sigma}^2/\sigma^2) - 1,\\
L_V(\hat{\sigma}^2, \sigma^2) &= \sigma^2(\hat{\sigma}^2/\sigma^2 - 1)^2/V = \frac{(\hat{\sigma}^2)^2}{\sigma^2 V} - 2\frac{\hat{\sigma}^2}{V} + \frac{\sigma^2}{V},
\end{aligned}
$$

both of which are invariant under scale transformations. Since a class of location-scale equivariant estimators is of the form cV for positive constant c, the corresponding loss functions for the estimator cV are given by

$$
\begin{aligned}
L_S(cV, \sigma^2) =& \frac{cV}{\sigma^2} - \log\frac{cV}{\sigma^2} - 1,\\
L_V(cV, \sigma^2) =& c\Big\{\frac{cV}{\sigma^2} + \frac{\sigma^2}{cV} - 2\Big\},
\end{aligned}
$$

both of which are zero at $cV/\sigma^2 = 1$ and diverge when $cV/\sigma^2 \to 0$ or $cV/\sigma^2 \to \infty$. Although a standard loss function is the scale-invariant quadratic loss $L_Q(\hat{\sigma}^2, \sigma^2) = (\hat{\sigma}^2/\sigma^2 - 1)^2$, the penalties are extremely unbalanced for the two cases of $cV/\sigma^2 < 1$ and $cV/\sigma^2 > 1$, since $L_Q(cV, \sigma^2)$ converges to 1 as $cV/\sigma^2 \to 0$, while it diverges when $cV/\sigma^2 \to \infty$. This undesirable property may be relaxed in the losses $L_S(cV, \sigma^2)$ and $L_V(cV, \sigma^2)$.

It is interesting to demonstrate that the losses $L_S(cV, \sigma^2)$ and $L_V(cV, \sigma^2)$ provide the same minimax and unbiased estimator, the same Bayes estimator and the same class of improved minimax estimators.

(1) It is seen that the unbiased estimator $\hat{\sigma}_0^2 = n^{-1}V$ is the best location-invariant and minimax under the two loss functions.

(2) Concerning the Bayes estimation, the Bayes estimator of σ^2 is given by the posterior harmonic mean $\hat{\sigma}^{2BAY} = (E[(\sigma^2)^{-1}|V])^{-1}$ under the two loss functions.

(3) Concerning the improvement over $\hat{\sigma}_0^2$, the Stein-type truncated estimator

$$
\hat{\sigma}^{2TR} = \min\big\{n^{-1}V,\ (n+m)^{-1}(V + \|\boldsymbol{X}\|^2)\big\} \tag{2.1}
$$

dominates $\hat{\sigma}_0^2$ relative to the two loss functions.

This type of truncated estimator (2.1) was first established by Stein (1964) under the quadratic loss. This dominance result can be verified below for a general class of scale-equivariant estimators given by

$$
\hat{\sigma}_\phi^2 = \phi(W)V, \quad \text{for} \quad W = \|\boldsymbol{X}\|^2/V,
$$

where $\|\boldsymbol{X}\|^2 = \boldsymbol{X}^t\boldsymbol{X}$.

Theorem 2.1 *Assume that $\phi(w)$ satisfies the following conditions:*
(a) *$\phi(w)$ is non-decreasing and* $\lim_{w\to\infty}\phi(w) = n^{-1}$.

(b) $\phi(w) \geq \phi_0(w)$ *for*

$$\phi_0(w) = \int_0^\infty F_m(wv) f_n(v) dv \Big/ \int_0^\infty v F_m(wv) f_n(v) dv,$$

where $f_m(v)$ and $F_m(v)$ denote density and distribution functions of a central chi-square distribution with m degrees of freedom.

Then, the scale-equivariant estimator $\hat{\sigma}^2_\phi$ dominates $\hat{\sigma}^2_0$ under the intrinsic loss and Stein loss functions.

Proof. We first show the dominance result for the intrinsic loss $L_V(\hat{\sigma}^2, \sigma^2)$. Since $\lim_{w\to\infty} \phi(w) = n^{-1}$, it can be seen that

$$\begin{aligned}
\Delta_V(\lambda) =& E[L_V(\hat{\sigma}^2_0, \sigma^2) - L_V(\hat{\sigma}^2_\phi, \sigma^2)] \\
=& E[\int_1^\infty \frac{\mathrm{d}}{\mathrm{d}t} L_V(\phi(tW)V, \sigma^2)\mathrm{d}t] \\
=& 2E[\int_1^\infty \Big\{\frac{V}{\sigma^2}\phi(tW) - 1\Big\} W\phi'(tW)\mathrm{d}t] \\
=& 2\int_0^\infty \int_0^\infty \int_1^\infty \Big\{v\phi(\frac{tu}{v}) - 1\Big\} \frac{u}{v}\phi'(tu/v)\mathrm{d}t f_m(u;\lambda) f_n(v) du dv,
\end{aligned}$$

where $f_m(v;\lambda)$ denotes the density function of a non-central chi-square distribution with m degrees of freedom and non-centrality parameter $\lambda = \|\boldsymbol{\theta}\|^2/(2\sigma^2)$. Making the transformations $w = (t/v)u$ and $z = vw/t$ with $\mathrm{d}w = (t/v)\mathrm{d}u$ and $\mathrm{d}z = (vw/t^2)\mathrm{d}t$, we can rewrite $\Delta(\lambda)$ as

$$\Delta_V(\lambda) = 2\int_0^\infty \int_0^\infty \{v\phi(w) - 1\} \int_0^{vw} f_m(z;\lambda)\mathrm{d}z f_n(v) dv \phi'(w)\mathrm{d}w. \tag{2.2}$$

Since $\phi'(w) \geq 0$, it is seen that $\Delta_V(\lambda) \geq 0$ if

$$\phi(w) \geq \frac{\int_0^\infty f_n(v) F_m(vw;\lambda)\mathrm{d}v}{\int_0^\infty v f_n(v) F_m(vw;\lambda)\mathrm{d}v},$$

for $F_m(vw;\lambda) = \int_0^{vw} f_m(z;\lambda)\mathrm{d}z$. It can be verified that

$$\frac{\int_0^\infty f_n(v) F_m(vw)\mathrm{d}v}{\int_0^\infty v f_n(v) F_m(vw)\mathrm{d}v} \geq \frac{\int_0^\infty f_n(v) F_m(vw;\lambda)\mathrm{d}v}{\int_0^\infty v f_n(v) F_m(vw;\lambda)\mathrm{d}v},$$

which proves the part of the intrinsic loss in Theorem 2.1.

For the Stein loss $L_S(\hat{\sigma}^2, \sigma^2)$, it is seen that

$$\begin{aligned}
\Delta_S(\lambda) =& E[L_S(\hat{\sigma}^2_0, \sigma^2) - L_S(\hat{\sigma}^2_\phi, \sigma^2)] \\
=& E[\int_1^\infty \frac{\mathrm{d}}{\mathrm{d}t} L_S(\phi(tW)V, \sigma^2)\mathrm{d}t] \\
=& E[\int_1^\infty \Big\{\frac{V}{\sigma^2}\phi(tW) - 1\Big\} W \frac{\phi'(tW)}{\phi(tW)}\mathrm{d}t],
\end{aligned}$$

which can be verified to be non-negative if $\phi(w)$ satisfies the conditions (a) and (b) in Theorem 2.1. Therefore, the proof is complete. $\square$

Let $\phi_{TR}(w) = \min\{n^{-1}, (n+m)^{-1}(1+w)\}$. Then, $\phi_{TR}(w)$ satisfies the conditions (a) and (b) since $\phi_0(w) \le \phi_{TR}(w)$, and it yields the Stein type truncated estimator $\hat{\sigma}^{2ST}$. Also, $\phi_0(w)$ satisfies the conditions (a) and (b). Thus, the estimators $\hat{\sigma}^{2ST}$ and $\hat{\sigma}^2_{\phi_0}$ dominate $\hat{\sigma}^2_0$ under the intrinsic and Stein loss functions. Especially, $\hat{\sigma}^2_{\phi_0}$ is the generalized Bayes estimator of σ^2 relative to the two loss functions.

2.2 A multivariate case

We next treat the estimation of the covariance matrix $\boldsymbol{\Sigma}$ in the model (1.1). Let $\boldsymbol{V} = \sum_{i=1}^n \boldsymbol{Y}_i \boldsymbol{Y}_i^t$. In the case $n \ge p$, $\boldsymbol{V}$ has a Wishart distribution $\mathcal{W}_p(n, \boldsymbol{\Sigma})$ with $E[\boldsymbol{V}] = n\boldsymbol{\Sigma}$, and the unbiased estimator of $\boldsymbol{\Sigma}$ is $\widehat{\boldsymbol{\Sigma}}^{UB} = \boldsymbol{V}/n$, which is neither admissible nor minimax, however. In fact, James and Stein (1961) showed that $\widehat{\boldsymbol{\Sigma}}^{UB}$ is dominated by the minimax estimator $\widehat{\boldsymbol{\Sigma}}_p^{JS} = \boldsymbol{T}\boldsymbol{D}_p^{JS}\boldsymbol{T}^t$ relative to the Stein loss function

$$L_S(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{tr}[\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}] - \log|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}| - p,$$

where $\boldsymbol{T}$ is a $p \times p$ lower triangular matrix with positive diagonal elements satisfying $\boldsymbol{V} = \boldsymbol{T}\boldsymbol{T}^t$ and $\boldsymbol{D}_p^{JS}$ is the diagonal matrix of order p with the i-th diagonal element being $(n+p-2i+1)^{-1}$.

A drawback of the Stein loss is that it is not available when $p > n$. As an alternative loss, we here use the intrinsic loss function

$$L_V(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \mathrm{tr}[\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}} - \boldsymbol{I}_p)^2\boldsymbol{V}^+] = \mathrm{tr}[\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{V}^+\widehat{\boldsymbol{\Sigma}}] - 2\mathrm{tr}[\widehat{\boldsymbol{\Sigma}}\boldsymbol{V}^+] + \mathrm{tr}[\boldsymbol{\Sigma}\boldsymbol{V}^+], \quad (2.3)$$

where $\boldsymbol{V}^+$ is the Moore-Penrose inverse of $\boldsymbol{V}$. It is interesting to point out that the intrinsic loss $L_V(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$ not only produces the same minimax estimator $\widehat{\boldsymbol{\Sigma}}_p^{JS}$ as given under the Stein loss for $n \ge p$, but also extends the dominance result to the case of $p > n$.

In the case $p > n$, the James-Stein type estimator is constructed as follows: Let $\boldsymbol{T}$ be a $p \times n$ matrix such that $\boldsymbol{V} = \boldsymbol{Y}^t\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{T}^t$ and

$$\boldsymbol{T} = (t_{ij}) = \begin{pmatrix} \boldsymbol{T}_1 \\ \boldsymbol{T}_2 \end{pmatrix},$$

where $\boldsymbol{T}_1$ is an $n \times n$ lower triangular matrix with positive diagonal elements and $\boldsymbol{T}_2$ is a $(p-n) \times n$ matrix. Then the James-Stein type estimator is given by

$$\widehat{\boldsymbol{\Sigma}}_n^{JS} = \boldsymbol{T}\boldsymbol{D}_n^{JS}\boldsymbol{T}^t,$$

where $\boldsymbol{D}_n^{JS} = \mathrm{diag}(d_1^{JS}, \ldots, d_n^{JS})$ for $d_i^{JS} = (n+p-2i+1)^{-1}$.

Theorem 2.2 *For real numbers a and b, denote $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. Let $\boldsymbol{V} = \boldsymbol{T}\boldsymbol{T}^t$, where $\boldsymbol{T}$ is a $p \times (n \wedge p)$ matrix such that all the diagonal elements are positive and all the off-diagonal elements above the diagonals are zeros. Let $\boldsymbol{D}^{JS} = \mathrm{diag}(d_1^{JS}, \ldots, d_{n\wedge p}^{JS})$ with $d_i^{JS} = (n + p - 2i + 1)^{-1}$. Then the James-Stein type estimator*

$$\widehat{\boldsymbol{\Sigma}}^{JS} = \boldsymbol{T}\boldsymbol{D}^{JS}\boldsymbol{T}^t \tag{2.4}$$

dominates $\widehat{\boldsymbol{\Sigma}}^{US}$ relative to the loss (2.3), where the estimator

$$\widehat{\boldsymbol{\Sigma}}^{US} = c_0\boldsymbol{V}, \qquad c_0 = \frac{n \wedge p}{np} = \frac{1}{n \vee p}, \tag{2.5}$$

is the best among estimators $c\boldsymbol{V}$ for positive constants c under the loss (2.3).

Proof. Since the case of $n \geq p$ can be easily verified, we here treat the case of $p > n$. Let us consider a class of estimators

$$\widehat{\boldsymbol{\Sigma}}_n^T = \boldsymbol{T}\boldsymbol{D}_n\boldsymbol{T}^t,$$

where the size of $\boldsymbol{T}$ is $p \times n$ and $\boldsymbol{D}_n = \mathrm{diag}(d_1, \ldots, d_n)$ is an $n \times n$ diagonal matrix with constant diagonals. We shall evaluate the risk of $\widehat{\boldsymbol{\Sigma}}_n^T$ relative to the intrinsic loss. Using Corollary 3.1 of Srivastava (2003), we can express the p.d.f. of $\boldsymbol{T}$ as

$$\frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\Big(-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{T}\boldsymbol{T}^t\Big) \times \frac{2^n\pi^{n^2/2}}{\Gamma_n[n/2]}\prod_{i=1}^n t_{ii}^{n-i},$$

where $\Gamma_n[n/2] = \pi^{n(n-1)/4}\prod_{i=1}^n \Gamma[(n-i+1)/2]$. Let $\boldsymbol{\Sigma}^{-1} = \boldsymbol{B}^t\boldsymbol{B}$, where $\boldsymbol{B} = (b_{ij})$ is a $p \times p$ lower triangular matrix with positive diagonal elements. Since the Jacobian of transformation $\boldsymbol{A} = (a_{ij}) = \boldsymbol{B}\boldsymbol{T}$ is given by

$$J[\boldsymbol{T} \to \boldsymbol{A}] = \Big(\prod_{i=1}^n b_{ii}^{-i}\Big)\Big(\prod_{i=n+1}^p b_{ii}^{-n}\Big),$$

the p.d.f. of $\boldsymbol{A}$ is written as

$$\begin{aligned}
&\frac{1}{(2\pi)^{np/2}}\frac{2^n\pi^{n^2/2}}{\Gamma_n(n/2)}\exp\Big(-\frac{1}{2}\mathrm{tr}\boldsymbol{A}\boldsymbol{A}^t\Big)\prod_{i=1}^n a_{ii}^{n-i} \\
&= \Big(\prod_{i=1}^n \frac{a_{ii}^{n-i}e^{-a_{ii}^2/2}}{2^{(n-i-1)/2}\Gamma[(n-i+1)/2]}\Big)\Big(\prod_{i=2}^p\prod_{j=1}^{(i-1)\wedge n}\frac{e^{-a_{ij}^2/2}}{\sqrt{2\pi}}\Big),
\end{aligned}$$

namely, $a_{ii}^2 \sim \chi_{n-i+1}^2$ for $i = 1, \ldots, n$, and $a_{ij} \sim \mathcal{N}(0, 1)$ for $2 \leq i \leq p$ and $1 \leq j \leq (i-1) \wedge n$. All the a_{ij}'s are mutually independent.

The risk of $\widehat{\boldsymbol{\Sigma}}_n^T$ is expressed by

$$\begin{aligned}
R(\widehat{\boldsymbol{\Sigma}}_n^T, \boldsymbol{\Sigma}) &= E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{T}\boldsymbol{D}_n\boldsymbol{T}^t(\boldsymbol{T}\boldsymbol{T}^t)^+\boldsymbol{T}\boldsymbol{D}_n\boldsymbol{T}^t - 2\mathrm{tr}\boldsymbol{T}\boldsymbol{D}_n\boldsymbol{T}^t(\boldsymbol{T}\boldsymbol{T}^t)^+ + \mathrm{tr}\boldsymbol{\Sigma}\boldsymbol{V}^+] \\
&= E[\mathrm{tr}\boldsymbol{A}\boldsymbol{D}_n^2\boldsymbol{A}^t] - 2\mathrm{tr}\boldsymbol{D}_n + E[\mathrm{tr}\boldsymbol{\Sigma}\boldsymbol{V}^+].
\end{aligned}$$

Denote $\boldsymbol{A} = (\boldsymbol{A}_1^t, \boldsymbol{A}_2^t)^t$, where $\boldsymbol{A}_1$ is the $n \times n$ matrix. It follows that

$$\mathrm{tr}\boldsymbol{A}\boldsymbol{D}_n^2\boldsymbol{A}^t = \mathrm{tr}\boldsymbol{D}_n^2\boldsymbol{A}^t\boldsymbol{A} = \mathrm{tr}\boldsymbol{D}_n^2\boldsymbol{A}_1^t\boldsymbol{A}_1 + \mathrm{tr}\boldsymbol{D}_n^2\boldsymbol{A}_2^t\boldsymbol{A}_2 = \sum_{i\geq j}^{n} a_{ij}^2 d_j^2 + \sum_{i=n+1}^{p}\sum_{j=1}^{n} a_{ij}^2 d_j^2,$$

which yields

$$E[\mathrm{tr}\boldsymbol{A}\boldsymbol{D}_n^2\boldsymbol{A}^t] = \sum_{i=1}^{n}(n-i+1)d_i^2 + \sum_{i>j}^{n} d_j^2 + \sum_{i=n+1}^{p}\sum_{j=1}^{n} d_j^2 = \sum_{i=1}^{n}(n+p-2i+1)d_i^2.$$

Hence the risk of $\widehat{\boldsymbol{\Sigma}}_n^T$ is rewritten by

$$R(\widehat{\boldsymbol{\Sigma}}_n^T, \boldsymbol{\Sigma}) = \sum_{i=1}^{n}\{(n+p-2i+1)d_i^2 - 2d_i\} + E[\mathrm{tr}\boldsymbol{\Sigma}\boldsymbol{V}^+]. \tag{2.6}$$

The best constant for d_i minimizing the risk is given by

$$d_i^{JS} = (n+p-2i+1)^{-1} \qquad (i = 1, \ldots, n),$$

which yields the James-Stein type estimator $\widehat{\boldsymbol{\Sigma}}^{JS}$ for $p > n$.

Concerning estimators $\widehat{\boldsymbol{\Sigma}}_c = c\boldsymbol{V}$ for positive constant c, the best c is n^{-1} under the loss (2.3) in the case of $n \geq p$, while in the case of $p > n$, the best c is p^{-1} under the loss (2.3), since $R(c\boldsymbol{V}, \boldsymbol{\Sigma}) = npc^2 - 2nc + E[\mathrm{tr}\boldsymbol{\Sigma}\boldsymbol{V}^+]$. In any of these cases, the estimator $c\boldsymbol{V}$ with the best c can be improved on by the James-Stein type estimator $\widehat{\boldsymbol{\Sigma}}^{JS}$ relative to the loss (2.3). □

3 Dominance Results in Estimation of the Covariance Matrix

3.1 Notations and preliminaries

We begin by giving some notations. Let $\mathcal{O}(r)$ be the group of $r \times r$ orthogonal matrices. For $r \geq q$, let $\mathcal{V}_{r,q}$ be the Stiefel manifold, namely the set of $r \times q$ matrices $\boldsymbol{M}$ such that $\boldsymbol{M}^t\boldsymbol{M} = \boldsymbol{I}_q$. It is noted that $\mathcal{O}(r) = \mathcal{V}_{r,r}$. Define $\mathbb{D}_r^+$ as the set of $r \times r$ diagonal matrices $\mathrm{diag}(d_1, \ldots, d_r)$ such that $d_1 > \cdots > d_r > 0$.

Let $\ell = m \wedge p \wedge n$. The eigenvalue decomposition of $\boldsymbol{V} = \boldsymbol{Y}^t\boldsymbol{Y}$ is written as

$$\boldsymbol{V} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t, \quad \text{for} \;\; \boldsymbol{H} \in \mathcal{V}_{p,n\wedge p} \;\; \text{and} \;\; \boldsymbol{L} \in \mathbb{D}_{n\wedge p}^+.$$

The nonsingular part of the singular value decomposition of $\boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1/2}$ is defined as

$$\boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1/2} = \boldsymbol{R}\boldsymbol{F}^{1/2}\boldsymbol{P}^t,$$

where $\boldsymbol{R} \in \mathcal{V}_{m,\ell}$, $\boldsymbol{P} \in \mathcal{V}_{n\wedge p,\ell}$ and $\boldsymbol{F}^{1/2} = \mathrm{diag}(f_1^{1/2}, \ldots, f_\ell^{1/2}) \in \mathbb{D}_\ell^+$. Let $\boldsymbol{V}^+$ be the Moore-Penrose inverse of $\boldsymbol{V}$. It is noted that $\boldsymbol{V}^+ = \boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{H}^t$ and

$$\boldsymbol{X}\boldsymbol{V}^+\boldsymbol{X}^t = \boldsymbol{X}\boldsymbol{H}\boldsymbol{L}^{-1}\boldsymbol{H}^t\boldsymbol{X}^t = \boldsymbol{R}\boldsymbol{F}\boldsymbol{R}^t.$$

Note also that $\boldsymbol{R}$ is orthogonal if $\ell = m$ and otherwise $\boldsymbol{P}$ is orthogonal.

A class of estimators treated in this section is of the form

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}) = \widehat{\boldsymbol{\Sigma}}^{US} + c_0\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t = c_0\{\boldsymbol{V} + \boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t\} \tag{3.1}$$

where $\boldsymbol{Q} = \boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{P}$ is a $p \times \ell$ matrix and $\boldsymbol{\Psi}(\boldsymbol{F})$ is an $\ell \times \ell$ diagonal matrix such that the diagonal elements are absolutely continuous functions of $\boldsymbol{F}$. The class (3.1) can be rewritten by

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}) = \widehat{\boldsymbol{\Sigma}}^{US} + c_0\boldsymbol{V}\boldsymbol{V}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{F}^{-1}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{V}\boldsymbol{V}^+. \tag{3.2}$$

Let $\boldsymbol{Q}^- = \boldsymbol{P}^t\boldsymbol{L}^{-1/2}\boldsymbol{H}^t$. Then $\boldsymbol{Q}^-$ is the generalized inverse of $\boldsymbol{Q}$ because

$$\boldsymbol{Q}\boldsymbol{Q}^-\boldsymbol{Q} = \boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{P}\boldsymbol{P}^t\boldsymbol{L}^{-1/2}\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{P} = \boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{P} = \boldsymbol{Q}.$$

It follows that

$$\boldsymbol{Q}^-\boldsymbol{V}(\boldsymbol{Q}^-)^t = \boldsymbol{I}_\ell, \quad \boldsymbol{Q}^-\boldsymbol{X}^t\boldsymbol{X}(\boldsymbol{Q}^-)^t = \boldsymbol{F}.$$

However, it is noted that

$$\boldsymbol{Q}\boldsymbol{Q}^t = \begin{cases} \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^t = \boldsymbol{V} & \text{for } m > n \wedge p, \\ \boldsymbol{V}\boldsymbol{V}^+\boldsymbol{X}^t(\boldsymbol{X}\boldsymbol{V}^+\boldsymbol{X}^t)^+\boldsymbol{X}\boldsymbol{V}\boldsymbol{V}^+ & \text{for } m \le n \wedge p, \end{cases}$$

and

$$\begin{aligned} \boldsymbol{Q}\boldsymbol{F}\boldsymbol{Q}^t &= \boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{P}\boldsymbol{F}^{1/2}\boldsymbol{R}^t\boldsymbol{R}\boldsymbol{F}^{1/2}\boldsymbol{P}^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t \\ &= \boldsymbol{H}\boldsymbol{H}^t\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t \\ &= \boldsymbol{V}\boldsymbol{V}^+\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{V}\boldsymbol{V}^+. \end{aligned}$$

To evaluate risk properties of the estimator (3.1), we here give some calculus and lemmas which will be used in the next subsection.

For an $m \times q$ rectangular matrix $\boldsymbol{Z} = (z_{ij})$, define an $m \times q$ rectangular matrix of differential operators with respect to $\boldsymbol{Z}$ as

$$\nabla_Z = \Big(\frac{\partial}{\partial z_{ij}}\Big).$$

The operation in terms of ∇_Z is defined as follows: For a differentiable and scalar-valued function $g(\boldsymbol{Z})$, $\nabla_Z g(\boldsymbol{Z})$ indicates an $m \times q$ rectangular matrix such that the (i,j)-th element is $(\partial/\partial z_{ij})g(\boldsymbol{Z})$. Also for a $q \times r$ matrix-valued function $\boldsymbol{G}(\boldsymbol{Z}) = (g_{ij}(\boldsymbol{Z}))$, $\nabla_Z\boldsymbol{G}(\boldsymbol{Z})$ is an $m \times r$ matrix whose (i,j)-th element is given by $\sum_{k=1}^q(\partial/\partial z_{ik})g_{kj}(\boldsymbol{Z})$.

Let $\boldsymbol{Z}$ be an $m \times q$ matrix and $\boldsymbol{L}$ a $q \times q$ diagonal matrix. Let $\boldsymbol{W} = \boldsymbol{Z}\boldsymbol{L}^{-1}\boldsymbol{Z}^t$. Denote by $\mathcal{D}_W = (\mathrm{d}_{ij}^W)$ the symmetric matrix of differential operators with respect to $\boldsymbol{W} = (w_{ij})$, where

$$\mathrm{d}_{ij}^W = \frac{1+\delta_{ij}}{2}\frac{\partial}{\partial w_{ij}}$$

with $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$. The operation in terms of $\mathcal{D}_W$ is defined in the same way as for ∇_Z.

The Stein (1973) identity, which is given in the following lemma, is a key tool to evaluating the risk function. For details, see Bilodeau and Kariya (1989) and Konno (1992).

Lemma 3.1 *Let $\boldsymbol{Z} = (z_{ij}) \sim \mathcal{N}_{m\times q}(\boldsymbol{0}_{m\times q}, \boldsymbol{I}_m \otimes \boldsymbol{\Omega})$. Let $\boldsymbol{G} = (g_{ij})$ be a $q \times m$ matrix such that all the elements g_{ij} are absolutely continuous functions of $\boldsymbol{Z}$ and satisfy $E[|g_{ab}z_{cd}|] < \infty$ and $E[|\partial g_{ab}/\partial z_{cd}|] < \infty$ for $a, d = 1, \ldots, q$ and $b, c = 1, \ldots, m$. It then follows that*

$$E[\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{G}] = E[\mathrm{tr}\nabla_Z\boldsymbol{G}].$$

The following two lemmas are useful for showing Theorem 3.1 given in the next subsection. The lemmas are easily proved by the same arguments as in Konno (1992, Lemma 2.1.9) and in Tsukuma and Kubokawa (2014, equation (6.18)), respectively, and the proofs are omitted.

Lemma 3.2 *Let $\boldsymbol{Z}$ be an $m \times q$ matrix and $\boldsymbol{L}$ a $q \times q$ diagonal matrix. Let $g(\boldsymbol{Z})$ be a differentiable and scalar-valued function of $\boldsymbol{Z}$. Define $\boldsymbol{G}(\boldsymbol{W})$ as an $m \times m$ symmetric matrix such that all the elements are differentiable functions of $\boldsymbol{W} = \boldsymbol{Z}\boldsymbol{L}^{-1}\boldsymbol{Z}^t$. Then we have*

$$\begin{aligned}\mathrm{tr}\nabla_Z\{\boldsymbol{Z}^t\boldsymbol{G}(\boldsymbol{W})g(\boldsymbol{Z})\} &= g(\boldsymbol{Z})(q-m-1)\mathrm{tr}\boldsymbol{G}(\boldsymbol{W}) + 2g(\boldsymbol{Z})\mathrm{tr}\mathcal{D}_W\{\boldsymbol{W}\boldsymbol{G}(\boldsymbol{W})\} \\ &\quad + \mathrm{tr}\boldsymbol{Z}^t\boldsymbol{G}(\boldsymbol{W})\{\nabla_Z g(\boldsymbol{Z})\},\end{aligned}$$

Lemma 3.3 *Let the notation be as in* (3.1) *and* (3.2). *Denote $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{V}^+\boldsymbol{X}^t$. Let $\boldsymbol{\Phi}(\boldsymbol{F})$ be an $\ell \times \ell$ diagonal matrix such that the diagonal elements are absolutely continuous functions of $\boldsymbol{F}$. Then we have*

$$E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{Q}^t] = E[a\mathrm{tr}\boldsymbol{\Phi}(\boldsymbol{F}) - 2\mathrm{tr}\mathcal{D}_W\boldsymbol{R}\boldsymbol{F}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{R}^t],$$

where $a = 2(n \vee p) - p - n + 2m + 1$.

3.2 A key inequality to improvement

We now prove the following theorem which will be used as a key tool to showing the Stein-type dominance results in the next subsection.

Theorem 3.1 *Let $\boldsymbol{\Phi}(\boldsymbol{F})$ be an $\ell\times\ell$ diagonal matrix such that the diagonal elements are absolutely continuous and nonnegative functions of $\boldsymbol{F}$. Then we have*

$$E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}(\boldsymbol{I}_\ell+\boldsymbol{F})\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{Q}^t]\geq E[(n\vee p+m)\mathrm{tr}\boldsymbol{\Phi}(\boldsymbol{F})].$$

Proof. Abbreviate $\boldsymbol{\Phi}(\boldsymbol{F})$ to $\boldsymbol{\Phi}$. Define

$$I_1=E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\boldsymbol{\Phi}\boldsymbol{Q}^t],\qquad I_2=E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\boldsymbol{F}\boldsymbol{\Phi}\boldsymbol{Q}^t].$$

The probability density function (p.d.f.) of $\boldsymbol{X}$ is proportional to

$$f(\boldsymbol{X}|\boldsymbol{\Theta},\boldsymbol{\Sigma})\propto\exp\Big(-\frac{1}{2}\mathrm{tr}(\boldsymbol{X}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\Theta})^t\Big),$$

where a normalizing constant is omitted. Take $\boldsymbol{H}_0$ as a $p\times(p-n\wedge p)$ matrix such that $\boldsymbol{H}_0\in\mathcal{V}_{p\times(p-n\wedge p)}$ and $\boldsymbol{H}_0^t\boldsymbol{H}=\boldsymbol{0}_{(p-n\wedge p)\times(n\wedge p)}$. Note that $[\boldsymbol{H},\boldsymbol{H}_0]\in\mathcal{O}(p)$ and

$$\begin{aligned}\mathrm{tr}(\boldsymbol{X}-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\Theta})^t&=\mathrm{tr}\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{X}^t+2\mathrm{tr}\boldsymbol{X}\boldsymbol{H}\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{H}_0\boldsymbol{H}_0^t-\boldsymbol{\Theta})^t\\&\quad+\mathrm{tr}(\boldsymbol{X}\boldsymbol{H}_0\boldsymbol{H}_0^t-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{H}_0\boldsymbol{H}_0^t-\boldsymbol{\Theta})^t.\end{aligned}$$

Making the orthogonal transformation $(\boldsymbol{Z},\boldsymbol{Z}_0)=(\boldsymbol{X}\boldsymbol{H},\boldsymbol{X}\boldsymbol{H}_0)$, we get the joint p.d.f. of $(\boldsymbol{Z},\boldsymbol{Z}_0)$, which is proportional to

$$\exp\Big(-\frac{1}{2}\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}^t+\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Xi}^t-\frac{1}{2}\mathrm{tr}(\boldsymbol{Z}_0\boldsymbol{H}_0^t-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}_0\boldsymbol{H}_0^t-\boldsymbol{\Theta})^t\Big),$$

where $\boldsymbol{\Omega}^{-1}=\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H}$ and $\boldsymbol{\Xi}=-(\boldsymbol{Z}_0\boldsymbol{H}_0^t-\boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}\boldsymbol{H}$. This implies that $\boldsymbol{Z}|\boldsymbol{Z}_0,\boldsymbol{Y}\sim\mathcal{N}_{m\times(n\wedge p)}(\boldsymbol{\Xi}\boldsymbol{\Omega},\boldsymbol{I}_m\otimes\boldsymbol{\Omega})$. Thus I_2 is expressed as

$$I_2=E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{V}^+\boldsymbol{X}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\boldsymbol{X}\boldsymbol{V}\boldsymbol{V}^+]=E^{\boldsymbol{Z}_0,\boldsymbol{Y}}[E^{\boldsymbol{Z}|\boldsymbol{Z}_0,\boldsymbol{Y}}[\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t]],$$

where $E^{\boldsymbol{Z}_0,\boldsymbol{Y}}$ denotes expectations with respect to $(\boldsymbol{Z}_0,\boldsymbol{Y})$ and $E^{\boldsymbol{Z}|\boldsymbol{Z}_0,\boldsymbol{Y}}$ denotes conditional expectation with respect to $\boldsymbol{Z}$ given $(\boldsymbol{Z}_0,\boldsymbol{Y})$. It is noted that

$$\boldsymbol{R}\boldsymbol{F}\boldsymbol{R}^t=\boldsymbol{X}\boldsymbol{V}^+\boldsymbol{X}^t=\boldsymbol{Z}\boldsymbol{L}^{-1}\boldsymbol{Z}^t$$

and

$$\begin{aligned}&E^{\boldsymbol{Z}|\boldsymbol{Z}_0,\boldsymbol{Y}}[\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t]\\&=K\exp\Big(-\frac{1}{2}\mathrm{tr}\boldsymbol{\Xi}\boldsymbol{\Omega}\boldsymbol{\Xi}^t\Big)\int\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\exp\Big(-\frac{1}{2}\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}^t+\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Xi}^t\Big)\mathrm{d}\boldsymbol{Z},\end{aligned}$$

where K is a normalizing constant.

Using the same arguments as in the proof of Theorem 3.3 of Díaz-García *et al*. (1997), we obtain

$$\exp(\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Xi}^t)={}_0F_1\Big(\frac{1}{2}\ell;\frac{1}{4}\boldsymbol{\Xi}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{\Xi}^t\Big),$$

where ${}_0F_1(\cdot)$ is a hypergeometric function with matrix argument. For details of a hypergeometric function, see Muirhead (1982, Section 7.3). Hence it is observed that

$$E^{\boldsymbol{Z}|\boldsymbol{Z}_0,\boldsymbol{Y}}[\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t] = E_*\Big[\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\cdot {}_0F_1\Big(\frac{1}{2}\ell;\frac{1}{4}\boldsymbol{\Xi}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{\Xi}^t\Big)\Big],$$

where E_* denotes expectation with respect to $\boldsymbol{Z}|\boldsymbol{Y}\sim\mathcal{N}_{m\times(n\wedge p)}(\boldsymbol{0}_{m\times(n\wedge p)},\boldsymbol{I}_m\otimes\boldsymbol{\Omega})$. Using Lemmas 3.1 and 3.2 gives that

$$\begin{aligned}
&E_*\Big[\mathrm{tr}\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\cdot {}_0F_1\Big(\frac{1}{2}\ell;\frac{1}{4}\boldsymbol{\Xi}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{\Xi}^t\Big)\Big]\\
&= E_*\Big[\mathrm{tr}\nabla_Z\boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\cdot {}_0F_1\Big(\frac{1}{2}\ell;\frac{1}{4}\boldsymbol{\Xi}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{\Xi}^t\Big)\Big]\\
&= E_*\Big[{}_0F_1\Big(\frac{1}{2}\ell;\frac{1}{4}\boldsymbol{\Xi}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{\Xi}^t\Big)\{a_2\mathrm{tr}\boldsymbol{\Phi}+2\mathrm{tr}\mathcal{D}_W\boldsymbol{R}\boldsymbol{F}\boldsymbol{\Phi}\boldsymbol{R}^t\}\Big]\\
&\quad + E_*\Big[\mathrm{tr}\boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\Big\{\nabla_Z\cdot {}_0F_1\Big(\frac{1}{2}\ell;\frac{1}{4}\boldsymbol{\Xi}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{\Xi}^t\Big)\Big\}\Big],
\end{aligned}\tag{3.3}$$

for $a_2 = n\wedge p - m - 1$ and $\boldsymbol{W} = \boldsymbol{Z}\boldsymbol{L}^{-1}\boldsymbol{Z}^t = \boldsymbol{R}\boldsymbol{F}\boldsymbol{R}^t$. Using Lemma 3.4 given below, we see that the second term in the last r.h.s. of (3.3) is nonnegative, which implies that

$$\begin{aligned}
I_2 &\ge E^{\boldsymbol{Z}_0,\boldsymbol{Y}}\Big[E_*\Big[{}_0F_1\Big(\frac{1}{2}\ell;\frac{1}{4}\boldsymbol{\Xi}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{\Xi}^t\Big)\{a_2\mathrm{tr}\boldsymbol{\Phi}+2\mathrm{tr}\mathcal{D}_W\boldsymbol{R}\boldsymbol{F}\boldsymbol{\Phi}\boldsymbol{R}^t\}\Big]\Big]\\
&= E[a_2\mathrm{tr}\boldsymbol{\Phi}+2\mathrm{tr}\mathcal{D}_W\boldsymbol{R}\boldsymbol{F}\boldsymbol{\Phi}\boldsymbol{R}^t].
\end{aligned}\tag{3.4}$$

Applying Lemma 3.3 to I_1, we get

$$I_1 = E[a_1\mathrm{tr}\boldsymbol{\Phi} - 2\mathrm{tr}\mathcal{D}_W\boldsymbol{R}\boldsymbol{F}\boldsymbol{\Phi}\boldsymbol{R}^t],\tag{3.5}$$

where $a_1 = 2(n\vee p) - p - n + 2m + 1$. It is here observed that

$$\begin{aligned}
a_1 + a_2 &= n\vee p + \{n\vee p + n\wedge p\} - p - n + m\\
&= n\vee p + m.
\end{aligned}\tag{3.6}$$

Combining (3.4), (3.5) and (3.6) gives that

$$E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}(\boldsymbol{I}_\ell+\boldsymbol{F})\boldsymbol{\Phi}\boldsymbol{Q}^t] = I_1 + I_2 \ge E[(n\vee p+m)\mathrm{tr}\boldsymbol{\Phi}].$$

Hence the proof is complete. □

Lemma 3.4 *Let $\boldsymbol{\Phi}$ be a diagonal matrix such that the diagonal elements are nonnegative functions of $\boldsymbol{F}$. Then we observe that*

$$\mathrm{tr}\boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\Big\{\nabla_Z\cdot {}_0F_1\Big(\frac{1}{2}\ell;\frac{1}{4}\boldsymbol{\Xi}\boldsymbol{Z}^t\boldsymbol{Z}\boldsymbol{\Xi}^t\Big)\Big\}\ge 0.$$

Proof. The proof will be proved by the same way as in Kubokawa and Tsai (2006). For a nonnegative integer k, let $\kappa = \{k_1, \ldots, k_\ell\}$ be a partition of k, namely $k_1 + \cdots + k_\ell = k$, where $k_i \geq 0$ for $i = 1, \ldots, \ell$. Denote by $\sum_\kappa$ the summation over all partitions $\kappa = \{k_1, \ldots, k_\ell\}$ of k, where $k_1 \geq \ldots \geq k_\ell \geq 0$. It follows from the definition of hypergeometric function (Muirhead (1982, p.258)) that

$$
{}_0F_1\Big(\frac{1}{2}\ell; \frac{1}{4}\Xi\boldsymbol{Z}^t\boldsymbol{Z}\Xi^t\Big) = \sum_{k=0}^{\infty}\sum_{\kappa} \alpha_\kappa^{(\ell)} \mathcal{C}_\kappa(\Xi\boldsymbol{Z}^t\boldsymbol{Z}\Xi^t),
$$

where $\alpha_\kappa^{(\ell)}$ are positive constants and $\mathcal{C}_\kappa(\Xi\boldsymbol{Z}^t\boldsymbol{Z}\Xi^t)$ are the zonal polynomials. For details of the hypergeometric function and the zonal polynomial, see Muirhead (1982) and also Takemura (1984).

Denote $q = n \wedge p$. It follows that

$$
\mathrm{tr}\boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\nabla_Z\mathcal{C}_\kappa(\Xi\boldsymbol{Z}^t\boldsymbol{Z}\Xi^t) = \sum_{i=1}^{m}\sum_{j=1}^{q}\{\boldsymbol{Z}^t\boldsymbol{R}\boldsymbol{\Phi}\boldsymbol{R}^t\}_{ji}\frac{\partial}{\partial z_{ij}}\mathcal{C}_\kappa(\Xi\boldsymbol{Z}^t\boldsymbol{Z}\Xi^t).
$$

Let $\boldsymbol{U} = (u_{ab}) = \boldsymbol{Z}\Xi^t$ and $\Xi = (\xi_{ab})$. Since the (a,b)-th element of $\boldsymbol{U}$ is given by $u_{ab} = \sum_{c=1}^{q} z_{ac}\xi_{bc}$, it is observed that

$$
\begin{aligned}
\frac{\partial}{\partial z_{ij}}\mathcal{C}_\kappa(\Xi\boldsymbol{Z}^t\boldsymbol{Z}\Xi^t) &= \sum_{a,b}^{m}\frac{\partial u_{ab}}{\partial z_{ij}}\cdot\frac{\partial}{\partial u_{ab}}\mathcal{C}_\kappa(\boldsymbol{U}^t\boldsymbol{U}) \\
&= \sum_{a,b}^{m}\sum_{c=1}^{q}\xi_{bc}\delta_{ia}\delta_{jc}\frac{\partial}{\partial u_{ab}}\mathcal{C}_\kappa(\boldsymbol{U}^t\boldsymbol{U}) \\
&= \sum_{b=1}^{m}\xi_{bj}\frac{\partial}{\partial u_{ib}}\mathcal{C}_\kappa(\boldsymbol{U}^t\boldsymbol{U}).
\end{aligned}
$$

Write the eigenvalue decomposition of $\boldsymbol{U}^t\boldsymbol{U}$ as $\boldsymbol{U}^t\boldsymbol{U} = \boldsymbol{O}\mathrm{diag}(d_1, \ldots, d_\ell)\boldsymbol{O}^t$, where $\boldsymbol{O} = (o_{ab}) \in \mathcal{V}_{m,\ell}$ and $d_1, \ldots, d_\ell$ are the nonzero eigenvalues of $\boldsymbol{U}^t\boldsymbol{U}$. Let

$$
\beta_{\kappa,a} = \frac{\partial}{\partial d_a}\mathcal{C}_\kappa(\boldsymbol{U}^t\boldsymbol{U}).
$$

It is noted that $\beta_{\kappa,a} \geq 0$ for every κ and a because the zonal polynomial $\mathcal{C}_\kappa(\boldsymbol{U}^t\boldsymbol{U})$ is a symmetric homogeneous polynomial of the nonzero eigenvalues of $\boldsymbol{U}^t\boldsymbol{U}$ with positive coefficients. Hence the chain rule gives that

$$
\frac{\partial}{\partial z_{ij}}\mathcal{C}_\kappa(\Xi\boldsymbol{Z}^t\boldsymbol{Z}\Xi^t) = \sum_{b=1}^{m}\xi_{bj}\frac{\partial}{\partial u_{ib}}\mathcal{C}_\kappa(\boldsymbol{U}^t\boldsymbol{U}) = \sum_{a=1}^{\ell}\sum_{b=1}^{m}\xi_{bj}\beta_{\kappa,a}\frac{\partial d_a}{\partial u_{ib}}.
$$

Using Lemma 4.1 of Konno (2009) yields that

$$
\frac{\partial d_a}{\partial u_{ib}} = 2\sum_{c=1}^{m} o_{ca}u_{ic}o_{ba},
$$

so that

$$\frac{\partial}{\partial z_{ij}}\mathcal{C}_\kappa(\Xi \boldsymbol{Z}^t\boldsymbol{Z}\Xi^t) = 2\sum_{a=1}^{\ell}\sum_{b,c}^{m}\xi_{bj}\beta_{\kappa,a}o_{ca}u_{ic}o_{ba} = 2\sum_{a=1}^{\ell}\{\boldsymbol{U}\boldsymbol{O}\}_{ia}\beta_{\kappa,a}\{\boldsymbol{O}^t\Xi\}_{aj}.$$

Thus we get

$$\mathrm{tr}\boldsymbol{Z}^t\boldsymbol{R}\Phi\boldsymbol{R}^t\nabla_Z\mathcal{C}_\kappa(\Xi\boldsymbol{Z}^t\boldsymbol{Z}\Xi^t) = 2\sum_{a=1}^{\ell}\beta_{\kappa,a}\{\boldsymbol{O}^t\boldsymbol{U}^t\boldsymbol{R}\Phi\boldsymbol{R}^t\boldsymbol{U}\boldsymbol{O}\}_{aa} \geq 0,$$

which completes the proof. □

3.3 Methods for improvements

We here present some kinds of improvements. Consider first the class of estimators (3.1), given by

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}) = \widehat{\boldsymbol{\Sigma}}^{US} + c_0\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t = c_0(\boldsymbol{V} + \boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t),$$

where $c_0 = (n \vee p)^{-1}$. We derive conditions for improvements over the James-Stein estimator $\widehat{\boldsymbol{\Sigma}}^{JS}$ given in (2.4) and the estimator $\widehat{\boldsymbol{\Sigma}}^{US} = c_0\boldsymbol{V}$ given in (2.5).

Theorem 3.2 *Let $\ell = n \wedge p \wedge m$ and $\boldsymbol{\Psi} = \mathrm{diag}(\psi_1, \ldots, \psi_\ell)$. For any order among n, p and m, the risk function of the estimator $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ given in* (3.1) *relative to the intrinsic loss* (2.3) *is expressed as*

$$\begin{aligned} R(\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}),\boldsymbol{\Sigma}) =& R(\widehat{\boldsymbol{\Sigma}}^{US},\boldsymbol{\Sigma}) \\ &+ c_0^2E\Big[\sum_{i=1}^{\ell}\{\alpha_i\psi_i^2 - 2(c_0^{-1}-\alpha_i)\psi_i\} - 4g_1(\boldsymbol{\Psi}) - 2g_2(\boldsymbol{\Psi})\Big], \end{aligned} \tag{3.7}$$

where $\alpha_i = |n-p| + 2i - 1$ for $i = 1, \ldots, \ell$ and

$$g_1(\boldsymbol{\Psi}) = \sum_{i=1}^{\ell} f_i(1+\psi_i)\frac{\partial\psi_i}{\partial f_i}, \quad g_2(\boldsymbol{\Psi}) = \sum_{i=1}^{\ell}\sum_{j>i}^{\ell}\frac{\psi_i^2 + 2\psi_i - \psi_j^2 - 2\psi_j}{f_i - f_j}f_j.$$

Proof. It is observed that

$$\begin{aligned} R(\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}),\boldsymbol{\Sigma}) &= E[L_V(\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}),\boldsymbol{\Sigma})] = E[\mathrm{tr}\boldsymbol{\Sigma}(c_0\boldsymbol{\Sigma}^{-1}\boldsymbol{V} - \boldsymbol{I}_p + c_0\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t)^2\boldsymbol{V}^+] \\ &= R(\widehat{\boldsymbol{\Sigma}}^{US},\boldsymbol{\Sigma}) + 2c_0E[c_0\mathrm{tr}\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t\boldsymbol{V}^+ - \mathrm{tr}\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t\boldsymbol{V}^+] \\ &\qquad + c_0^2E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t\boldsymbol{V}^+\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t] \\ &= R(\widehat{\boldsymbol{\Sigma}}^{US},\boldsymbol{\Sigma}) + c_0^2(2E_1 - 2c_0^{-1}E_2 + E_3), \end{aligned} \tag{3.8}$$

where $E_1 = E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t\boldsymbol{V}^+\boldsymbol{V}]$, $E_2 = E[\mathrm{tr}\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t\boldsymbol{V}^+]$ and

$$E_3 = E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t\boldsymbol{V}^+\boldsymbol{Q}\boldsymbol{\Psi}(\boldsymbol{F})\boldsymbol{Q}^t].$$

Since $\boldsymbol{Q}^t\boldsymbol{V}^+\boldsymbol{Q} = \boldsymbol{I}_\ell$, it is seen that

$$E_2 = E[\mathrm{tr}\boldsymbol{\Psi}(\boldsymbol{F})] = E\Big[\sum_{i=1}^{\ell}\psi_i\Big]. \tag{3.9}$$

Note that $\boldsymbol{Q}^t\boldsymbol{V}^+\boldsymbol{V} = \boldsymbol{P}^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t\boldsymbol{H}\boldsymbol{H}^t = \boldsymbol{P}^t\boldsymbol{L}^{1/2}\boldsymbol{H}^t = \boldsymbol{Q}^t$. Applying Lemma 3.3 to E_1 leads to

$$E_1 = E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\boldsymbol{\Psi}\boldsymbol{Q}^t] = E[a_1\mathrm{tr}\boldsymbol{\Psi} - 2\mathrm{tr}\mathcal{D}_W\boldsymbol{R}\boldsymbol{F}\boldsymbol{\Psi}\boldsymbol{R}^t],$$

where $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{F})$ and $a_1 = 2(n\vee p) - p - n + 2m + 1$. Using Lemma 6.4 of Tsukuma and Kubokawa (2014) gives that

$$\begin{aligned}\mathrm{tr}\mathcal{D}_W\boldsymbol{R}\boldsymbol{F}\boldsymbol{\Psi}\boldsymbol{R}^t &= \sum_{i=1}^{\ell}\Big\{(m-\ell+1)\psi_i + f_i\frac{\partial\psi_i}{\partial f_i} + \sum_{j>i}^{\ell}\frac{f_i\psi_i - f_j\psi_j}{f_i - f_j}\Big\}\\ &= \sum_{i=1}^{\ell}\Big\{(m-i+1)\psi_i + f_i\frac{\partial\psi_i}{\partial f_i} + \sum_{j>i}^{\ell}\frac{\psi_i - \psi_j}{f_i - f_j}f_j\Big\},\end{aligned}$$

which implies that

$$E_1 = E\Big[\sum_{i=1}^{\ell}\Big\{\alpha_i\psi_i - 2f_i\frac{\partial\psi_i}{\partial f_i} - 2\sum_{j>i}^{\ell}\frac{\psi_i - \psi_j}{f_i - f_j}f_j\Big\}\Big]. \tag{3.10}$$

where $\alpha_i = 2(n\vee p) - p - n + 2m + 1 - 2(m-i+1) = |n-p| + 2i - 1$. Similarly, E_3 can be expressed as

$$E_3 = E[\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\boldsymbol{\Psi}^2(\boldsymbol{F})\boldsymbol{Q}^t] = \sum_{i=1}^{\ell}\Big\{\alpha_i\psi_i^2 - 4f_i\psi_i\frac{\partial\psi_i}{\partial f_i} - 2\sum_{j>i}^{\ell}\frac{\psi_i^2 - \psi_j^2}{f_i - f_j}f_j\Big\}. \tag{3.11}$$

Combining (3.8), (3.9), (3.10) and (3.11), we obtain (3.7). Thus the proof is complete. □

Using Theorem 3.2, we can investigate dominance properties for a couple of estimators. A Stein-type estimator is described by

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^{ST}) = c_0\{\boldsymbol{V} + \boldsymbol{Q}\boldsymbol{\Psi}^{ST}(\boldsymbol{F})\boldsymbol{Q}^t\}, \quad \boldsymbol{\Psi}^{ST}(\boldsymbol{F}) = \mathrm{diag}(\psi_1^{ST}, \ldots, \psi_\ell^{ST}),$$

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

$$\psi_i^{ST} = \frac{c_0^{-1} - \alpha_i}{\alpha_i} = \frac{n\wedge p - 2i + 1}{|n-p| + 2i - 1}.$$

Then from Theorem 3.2, it follows that $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^{ST})$ dominates $\widehat{\boldsymbol{\Sigma}}^{US}$ for any order of n, p and m, and that it further dominates the James-Stein type estimator $\widehat{\boldsymbol{\Sigma}}^{JS}$ if $m > n\wedge p$. In fact, the risk function of $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^{ST})$ under the loss (2.3) is expressed as

$$R(\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^{ST}), \boldsymbol{\Sigma}) = R(\widehat{\boldsymbol{\Sigma}}^{US}, \boldsymbol{\Sigma}) - c_0^2\sum_{i=1}^{\ell}\frac{(c_0^{-1} - \alpha_i)^2}{\alpha_i} - 2c_0^2E\big[g_2(\boldsymbol{\Psi}^{ST})\big],$$

which is less than $R(\widehat{\mathbf{\Sigma}}^{US}, \mathbf{\Sigma})$ since, for $j > i$,

$$(\psi_i^{ST})^2 + 2\psi_i^{ST} - (\psi_j^{ST})^2 - 2\psi_j^{ST} = c_0^{-2}(\alpha_i^{-2} - \alpha_j^{-2}) > 0,$$

namely $g_2(\mathbf{\Psi}^{ST}) > 0$. Furthermore, when $\ell = n \wedge p$, namely, $m > n \wedge p$, it then follows that

$$R(\widehat{\mathbf{\Sigma}}^{US}, \mathbf{\Sigma}) - c_0^2 \sum_{i=1}^{n\wedge p} \frac{(c_0^{-1} - \alpha_i)^2}{\alpha_i} = -\sum_{i=1}^{n\wedge p} \frac{1}{\alpha_i} + E[\mathrm{tr}\mathbf{\Sigma}\boldsymbol{V}^+] = R(\widehat{\mathbf{\Sigma}}^{JS}, \mathbf{\Sigma}).$$

This shows that if $\ell = n \wedge p$ then $\widehat{\mathbf{\Sigma}}(\mathbf{\Psi}^{ST})$ dominates $\widehat{\mathbf{\Sigma}}^{JS}$ relative to the loss (2.3).

Another reasonable estimator is the Haff (1980) type estimator

$$\begin{aligned}
\widehat{\mathbf{\Sigma}}(\mathbf{\Psi}^{HF}) &= c_0(\boldsymbol{V} + \boldsymbol{Q}\mathbf{\Psi}^{HF}\boldsymbol{Q}^t), \quad \mathbf{\Psi}^{HF}(\boldsymbol{F}) = \mathrm{diag}(\psi_1^{HF}, \ldots, \psi_\ell^{HF}),\\
\psi_i^{HF} &= \frac{a}{\mathrm{tr}\boldsymbol{F}} f_i \quad (i = 1, \ldots, \ell).
\end{aligned}$$

Using Theorem 3.2, we can show that the Haff type estimator $\widehat{\mathbf{\Sigma}}(\mathbf{\Psi}^{HF})$ dominates $\widehat{\mathbf{\Sigma}}^{US}$ if constant a satisfies the inequality $0 < a \leq 2(n \wedge p - 1)/(|n - p| + 1)$ for $n \wedge p > 1$. In fact, it is noted that

$$\sum_{i=1}^{\ell}\sum_{j>i}^{\ell} \frac{\psi_i^{HF} - \psi_j^{HF}}{f_i - f_j} f_j = \frac{a}{\mathrm{tr}\boldsymbol{F}} \sum_{i=1}^{\ell}\sum_{j>i}^{\ell} f_j = \frac{a}{\mathrm{tr}\boldsymbol{F}} \sum_{i=1}^{\ell}(i-1) f_i \tag{3.12}$$

and also

$$\begin{aligned}
\sum_{i=1}^{\ell}\sum_{j>i}^{\ell} \frac{(\psi_i^{HF})^2 - (\psi_j^{HF})^2}{f_i - f_j} f_j &= \frac{a^2}{(\mathrm{tr}\boldsymbol{F})^2} \sum_{i=1}^{\ell}\sum_{j>i}^{\ell} (f_i + f_j) f_j\\
&= \frac{a^2}{2(\mathrm{tr}\boldsymbol{F})^2} \sum_{i=1}^{\ell}\sum_{j\neq i}^{\ell} f_i f_j + \frac{a^2}{(\mathrm{tr}\boldsymbol{F})^2} \sum_{i=1}^{\ell}\sum_{j>i}^{\ell} f_j^2\\
&= \frac{a^2}{2(\mathrm{tr}\boldsymbol{F})^2} \sum_{i=1}^{\ell}\left(\sum_{j=1}^{\ell} f_i f_j - f_i^2\right) + \frac{a^2}{(\mathrm{tr}\boldsymbol{F})^2} \sum_{i=1}^{\ell}(i-1) f_i^2\\
&= \frac{a^2}{2(\mathrm{tr}\boldsymbol{F})^2} \sum_{i=1}^{\ell} \left\{ f_i \mathrm{tr}\boldsymbol{F} + (2i-3) f_i^2 \right\}. \qquad (3.13)
\end{aligned}$$

Combining these identities (3.12) and (3.13) gives

$$g_2(\mathbf{\Psi}^{HF}) = \frac{a^2}{2} + \frac{1}{2}\sum_{i=1}^{\ell}\left\{ \frac{a^2}{(\mathrm{tr}\boldsymbol{F})^2}(2i-3) f_i^2 + \frac{4a}{\mathrm{tr}\boldsymbol{F}}(i-1) f_i \right\}.$$

Thus the difference in risk of $\widehat{\mathbf{\Sigma}}(\mathbf{\Psi}^{HF})$ and $\widehat{\mathbf{\Sigma}}^{US}$ is written as

$$\begin{aligned}
&R(\widehat{\mathbf{\Sigma}}(\mathbf{\Psi}^{HF}), \mathbf{\Sigma}) - R(\widehat{\mathbf{\Sigma}}^{US}, \mathbf{\Sigma})\\
&= c_0^2 \left\{ (|n-p| + 2) a^2 E\left[\frac{\mathrm{tr}\boldsymbol{F}^2}{(\mathrm{tr}\boldsymbol{F})^2}\right] - a^2 - 2(n \wedge p - 1)a - 4E[g_1(\mathbf{\Psi}^{HF})] \right\}.
\end{aligned}$$

Since $\mathrm{tr}\boldsymbol{F}^2/(\mathrm{tr}\boldsymbol{F})^2 \leq 1$ and

$$g_1(\boldsymbol{\Psi}^{HF}) = \frac{a}{\mathrm{tr}\boldsymbol{F}}\sum_{i=1}^{\ell} f_i\Big(1+\frac{af_i}{\mathrm{tr}\boldsymbol{F}}\Big)\Big(1-\frac{f_i}{\mathrm{tr}\boldsymbol{F}}\Big) \geq 0,$$

it is observed that

$$R(\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^{HF}),\boldsymbol{\Sigma}) - R(\widehat{\boldsymbol{\Sigma}}^{US},\boldsymbol{\Sigma}) \leq c_0^2\left\{(|n-p|+1)a^2 - 2(n\wedge p-1)a\right\},$$

which shows the dominance result.

Next, we consider improvement on $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^{ST})$ and $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^{HF})$ by using Theorem 3.1. Let $[\boldsymbol{\Psi}]^{TR} = \mathrm{diag}(\psi_1^{TR}(\boldsymbol{F}),\ldots,\psi_\ell^{TR}(\boldsymbol{F}))$ be a diagonal matrix of order ℓ such that the i-th diagonal element is given by

$$\psi_i^{TR}(\boldsymbol{F}) = \min\Big\{\psi_i(\boldsymbol{F}), \frac{n\vee p}{n\vee p+m}(1+f_i)-1\Big\},$$

where $\boldsymbol{\Psi}(\boldsymbol{F}) = \mathrm{diag}(\psi_1(\boldsymbol{F}),\ldots,\psi_\ell(\boldsymbol{F}))$. Then we obtain a general dominance result for improvement on the class (3.1).

Theorem 3.3 *Assume that each diagonal element of $\boldsymbol{\Psi}(\boldsymbol{F})+\boldsymbol{I}_\ell$ is larger than or equal to zero. For any tuple of positive integers n, p and m, the truncated estimator $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR})$ dominates $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ relative to the loss* (2.3) *if* $\Pr([\boldsymbol{\Psi}]^{TR} \neq \boldsymbol{\Psi}) > 0$.

Proof. Abbreviate $\boldsymbol{\Psi}(\boldsymbol{F})$ to $\boldsymbol{\Psi}$. The difference in risk of $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ and $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR})$ is expressed by

$$\begin{aligned}
&R(\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}),\boldsymbol{\Sigma}) - R(\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR}),\boldsymbol{\Sigma})\\
&= E[c_0^2\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\{\boldsymbol{\Psi}^2+2\boldsymbol{\Psi}-([\boldsymbol{\Psi}]^{TR})^2-2[\boldsymbol{\Psi}]^{TR}\}\boldsymbol{Q}^t - 2c_0\mathrm{tr}(\boldsymbol{\Psi}-[\boldsymbol{\Psi}]^{TR})]\\
&= c_0E[c_0\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\{(\boldsymbol{\Psi}+\boldsymbol{I}_\ell)^2-([\boldsymbol{\Psi}]^{TR}+\boldsymbol{I}_\ell)^2\}\boldsymbol{Q}^t - 2\mathrm{tr}(\boldsymbol{\Psi}-[\boldsymbol{\Psi}]^{TR})].
\end{aligned}$$

Denote $\boldsymbol{\Psi} = \mathrm{diag}(\psi_1,\ldots,\psi_\ell)$ and $[\boldsymbol{\Psi}]^{TR} = \mathrm{diag}(\psi_1^{TR},\ldots,\psi_\ell^{TR})$. From the given assumption and the definition of $[\boldsymbol{\Psi}]^{TR}$, it is seen that $\psi_i \geq \psi_i^{TR}$ and $\psi_i^{TR}+1 \geq 0$ for each i. Thus using Theorem 3.1 verifies

$$\begin{aligned}
&R(\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}),\boldsymbol{\Sigma}) - R(\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}]^{TR}),\boldsymbol{\Sigma})\\
&\geq c_0E[c_0(n\vee p+m)\mathrm{tr}(\boldsymbol{I}_\ell+\boldsymbol{F})^{-1}\{(\boldsymbol{\Psi}+\boldsymbol{I}_\ell)^2-([\boldsymbol{\Psi}]^{TR}+\boldsymbol{I}_\ell)^2\} - 2\mathrm{tr}(\boldsymbol{\Psi}-[\boldsymbol{\Psi}]^{TR})].
\end{aligned} \tag{3.14}$$

The r.h.s. of (3.14) is rewritten as

$$c_0E\Big[\sum_{i=1}^{\ell}(\psi_i-\psi_i^{TR})\Big\{c_0\frac{n\vee p+m}{1+f_i}(\psi_i+\psi_i^{TR}+2)-2\Big\}\Big]. \tag{3.15}$$

Recall that $\psi_i^{TR} = \min\{\psi_i, c_0^{-1}(n \vee p + m)^{-1}(1 + f_i) - 1\} \le \psi_i$. The summation inside the square brackets in (3.15) is bounded below by

$$\begin{aligned} &\sum_{i=1}^{\ell}(\psi_i - \psi_i^{TR})\Big\{c_0 \frac{n \vee p + m}{1 + f_i}(\psi_i^{TR} + \psi_i^{TR} + 2) - 2\Big\} \\ &= 2\sum_{i=1}^{\ell}(\psi_i - \psi_i^{TR})\Big\{c_0 \frac{n \vee p + m}{1 + f_i}(\psi_i^{TR} + 1) - 1\Big\}, \end{aligned}$$

which is equal to zero. Hence the proof is complete. $\square$

It is obviously seen that $\psi_i^{ST} + 1 \ge 0$ and $\psi_i^{HF} + 1 \ge 0$ for $i = 1, \ldots, \ell$. Hence the following corollary is given from Proposition 3.3.

Corollary 3.1 *The truncated estimator $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^{ST}]^{TR})$ dominates $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^{ST})$ relative to the loss* (2.3)*. Also, $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^{HF}]^{TR})$ dominates $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^{HF})$ relative to the loss* (2.3).

We conclude this section with providing other types of estimators. As shown in Theorem 2.2, the James-Stein estimator $\widehat{\boldsymbol{\Sigma}}^{JS}$ dominates $\widehat{\boldsymbol{\Sigma}}^{US}$ under the intrinsic loss (2.3). Using Theorem 2.2 of Konno (2009), we can show that $\widehat{\boldsymbol{\Sigma}}^{JS}$ is further dominated by the orthogonally invariant estimator $\widehat{\boldsymbol{\Sigma}}^{DS} = \boldsymbol{HLD}^{JS}\boldsymbol{H}^t$ relative to the loss (2.3). The estimator $\widehat{\boldsymbol{\Sigma}}^{DS}$ is called the Dey-Srinivasan (1985) type estimator. It is recalled that the truncated Stein type estimator is given by $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^{ST}]^{TR}) = \widehat{\boldsymbol{\Sigma}}^{US} + c_0\boldsymbol{Q}[\boldsymbol{\Psi}^{ST}]^{TR}\boldsymbol{Q}^t$. In this estimator, we can suggest replacing $\widehat{\boldsymbol{\Sigma}}^{US}$ with $\widehat{\boldsymbol{\Sigma}}^{JS}$ and $\widehat{\boldsymbol{\Sigma}}^{DS}$, which gives

$$\widehat{\boldsymbol{\Sigma}}^{JSTR} = \widehat{\boldsymbol{\Sigma}}^{JS} + c_0\boldsymbol{Q}[\boldsymbol{\Psi}^{ST}]^{TR}\boldsymbol{Q}^t, \tag{3.16}$$

$$\widehat{\boldsymbol{\Sigma}}^{DSTR} = \widehat{\boldsymbol{\Sigma}}^{DS} + c_0\boldsymbol{Q}[\boldsymbol{\Psi}^{ST}]^{TR}\boldsymbol{Q}^t. \tag{3.17}$$

It is, however, difficult to show that $\widehat{\boldsymbol{\Sigma}}^{JSTR}$ and $\widehat{\boldsymbol{\Sigma}}^{DSTR}$ dominate $\widehat{\boldsymbol{\Sigma}}^{JS}$ and $\widehat{\boldsymbol{\Sigma}}^{DS}$, respectively, relative to the loss (2.3). In the next subsection, we will investigate the performances of $\widehat{\boldsymbol{\Sigma}}^{JSTR}$ and $\widehat{\boldsymbol{\Sigma}}^{DSTR}$ through the Monte Carlo simulations.

3.4 Simulation studies

We here briefly report risk performances of estimators suggested in this section by simulation. Especially, it is interesting to investigate whether $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^{ST})$ or $\widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^{ST}]^{TR})$ dominates the James-Stein estimator $\widehat{\boldsymbol{\Sigma}}^{JS}$ in the case of $m < p$, because this dominance result can not be shown analytically. Since dominance properties of the estimators $\widehat{\boldsymbol{\Sigma}}^{JSTR}$ and $\widehat{\boldsymbol{\Sigma}}^{DSTR}$ given in (3.16) and (3.17) cannot be shown analytically, it is also interesting to examine their risk performances numerically.

In our simulation studies, we consider the following two cases: (A) $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ and (B) $\boldsymbol{\Sigma} = \mathrm{diag}(1, 2, \ldots, p)$ as the true covariance matrix $\boldsymbol{\Sigma}$. The true mean matrix $\boldsymbol{\Theta} = (\theta_{ij})$ is supposed as (a) $\boldsymbol{\Theta} = \boldsymbol{0}_{m\times p}$, (b) $\theta_{ij} = 3\sin(i^2 + j)$ and (c) $\theta_{ij} = 10\sin(i^2 + j)$ for

$i = 1, \ldots, m$ and $j = 1, \ldots, p$. Values of m, p and n are taken as combinations of 5, 10 and 15.

Through the simulation experiments, we investigate the risk performances of the estimators suggested by the previous subsection, given by

1. $\widehat{\boldsymbol{\Sigma}}^{HF} = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^{HF})$ with $a = (n \wedge p - 1)/(|n - p| + 1)$;
2. $\widehat{\boldsymbol{\Sigma}}^{HFTR} = \widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^{HF}]^{TR})$ with $a = (n \wedge p - 1)/(|n - p| + 1)$;
3. $\widehat{\boldsymbol{\Sigma}}^{ST} = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Psi}^{ST})$;
4. $\widehat{\boldsymbol{\Sigma}}^{STTR} = \widehat{\boldsymbol{\Sigma}}([\boldsymbol{\Psi}^{ST}]^{TR})$;
5. $\widehat{\boldsymbol{\Sigma}}^{JSTR} = \widehat{\boldsymbol{\Sigma}}^{JS} + c_0 \boldsymbol{Q}[\boldsymbol{\Psi}^{ST}]^{TR}\boldsymbol{Q}^t$;
6. $\widehat{\boldsymbol{\Sigma}}^{DS} = \boldsymbol{HLD}^{JS}\boldsymbol{H}^t$;
7. $\widehat{\boldsymbol{\Sigma}}^{DSTR} = \widehat{\boldsymbol{\Sigma}}^{DS} + c_0 \boldsymbol{Q}[\boldsymbol{\Psi}^{ST}]^{TR}\boldsymbol{Q}^t$.

The improvements of the above estimators over $\widehat{\boldsymbol{\Sigma}}^{JS}$ are measured by the percentage relative improvement in risk (PRIR), which is defined as

$$100 \times \frac{R(\widehat{\boldsymbol{\Sigma}}^{JS}, \boldsymbol{\Sigma}) - R(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})}{R(\widehat{\boldsymbol{\Sigma}}^{JS}, \boldsymbol{\Sigma})},$$

where $\widehat{\boldsymbol{\Sigma}}$ is any of the above estimators from 1 to 7. The risk function of each estimator is calculated by average of the loss function (2.3) based on 50,000 replications. The estimated PRIRs are reported in Table 1.

Table 1 indicates several interesting observations.

(1) The negative PRIRs imply that the corresponding estimators are inferior to $\widehat{\boldsymbol{\Sigma}}^{JS}$. Such PRIRs frequently appear in the cases of $m < p$.

(2) In the cases satisfying $m < p$, $\widehat{\boldsymbol{\Sigma}}^{ST}$ and $\widehat{\boldsymbol{\Sigma}}^{STTR}$ sometimes improve on $\widehat{\boldsymbol{\Sigma}}^{JS}$ and sometimes do not.

(3) When $(n, p, m) = (15, 10, 5)$ and $(10, 15, 5)$, namely $\ell = m$ $(m < n \wedge p)$, the improvements over $\widehat{\boldsymbol{\Sigma}}^{JS}$ are generally poor since the information is not much available on the sample mean matrix $\boldsymbol{X}$ with small m.

(4) The risk functions of $\widehat{\boldsymbol{\Sigma}}^{HF}$, $\widehat{\boldsymbol{\Sigma}}^{HFTR}$, $\widehat{\boldsymbol{\Sigma}}^{ST}$, $\widehat{\boldsymbol{\Sigma}}^{STTR}$ and $\widehat{\boldsymbol{\Sigma}}^{JSTR}$ do not depend on $\boldsymbol{\Sigma}$ when $n \geq p$ and $\boldsymbol{\Theta} = \boldsymbol{0}_{m \times p}$. This fact follows from invariance of these estimators under a scale transformation. On the other hand, the risk function of $\widehat{\boldsymbol{\Sigma}}^{DS}$ is invariant under a location transformation on $\boldsymbol{X}$ for any fixed (n, m, p) because $\widehat{\boldsymbol{\Sigma}}^{DS}$ is independent of $\boldsymbol{X}$.

(5) $\widehat{\mathbf{\Sigma}}^{JSTR}$ and $\widehat{\mathbf{\Sigma}}^{DSTR}$ substantially reduce the risks of $\widehat{\mathbf{\Sigma}}^{JS}$ and $\widehat{\mathbf{\Sigma}}^{DS}$, respectively.

(6) In our simulation studies, the excellent estimator is $\widehat{\mathbf{\Sigma}}^{DSTR}$ among the estimators considered here, though it is difficult to establish the improvement of $\widehat{\mathbf{\Sigma}}^{DSTR}$ over $\widehat{\mathbf{\Sigma}}^{DS}$ analytically.

4 Extensions

The results given in the previous sections will be here extended to the two directions: Estimation of a component of covariance and estimation of the precision matrix.

4.1 Estimation of the covariance matrix in linear mixed models

A canonical form of multivariate linear mixed models can be provided by the marginal distribution of the model (1.1) under the assumption that $\boldsymbol{\theta}_i \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{\Sigma}_A)$ for $i = 1, \ldots, m$. Then, the canonical model is expressed as

$$\begin{aligned} \boldsymbol{X}_i &\sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{\Sigma} + \mathbf{\Sigma}_A), \quad i = 1, \ldots, m, \\ \boldsymbol{Y}_j &\sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{\Sigma}), \quad j = 1, \ldots, n, \end{aligned} \tag{4.1}$$

where the covariance matrices $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_A$ are referred to as the 'within' and 'between' multivariate components of variance, respectively. Let $\boldsymbol{V} = \sum_{i=1}^n \boldsymbol{Y}_i \boldsymbol{Y}_i^t$ and $\boldsymbol{V}_2 = \sum_{i=1}^m \boldsymbol{X}_i \boldsymbol{X}_i^t$. The problem is that we want to construct truncated estimators improving the unbiased estimator $n^{-1}\boldsymbol{V}$ using the statistic $\boldsymbol{V}_2$. This is known to be a hard issue, and in the case of $n \geq p$ and $m \geq p$, Srivastava and Kubokawa (1999) used the conditional distribution of $\boldsymbol{V}$ given $\boldsymbol{V}_2^{-1/2}\boldsymbol{V}\boldsymbol{V}_2^{-1/2}$ to get the improvement. Kubokawa and Tsai (2006) derived the dominance result using the Stein-Haff identity in the case of $n \geq p$ without any constraints on m and p.

It is noted that the risk function of estimator $\widehat{\mathbf{\Sigma}}$ relative to the intrinsic loss is

$$R(\mathbf{\Sigma}, \mathbf{\Sigma}_A, \widehat{\mathbf{\Sigma}}) = E[L_V(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma})] = E[E[L_V(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma})|\mathbf{\Theta}]],$$

where $\mathbf{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m)^t$. Since $E[L_V(\widehat{\mathbf{\Sigma}}, \mathbf{\Sigma})|\mathbf{\Theta}]$ is the corresponding risk in the original model (1.1), it can be seen that all the dominance results given in the model (1.1) still hold in the covariance components model (4.1). Thus, one gets unified dominance results in the model (4.1), irrespective of n, p and m.

4.2 Estimation of the precision matrix

Let $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m)^t$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)^t$ be random matrices having distributions given in (1.1) with unknown $\boldsymbol{\theta}_i$'s and unknown $\mathbf{\Sigma}$. Consider the problem of estimating the precision matrix $\boldsymbol{\Sigma}^{-1}$ under a weighted invariant quadratic loss given by

$$\begin{aligned} L(\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}) &= \mathrm{tr}[\{(\boldsymbol{\delta} - \boldsymbol{\Sigma}^{-1})\boldsymbol{V}\}^2] \\ &= \mathrm{tr}[\boldsymbol{\delta}\boldsymbol{V}\boldsymbol{\delta}\boldsymbol{V}] - 2\mathrm{tr}[\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\delta}\boldsymbol{V}] + \mathrm{tr}[\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}], \end{aligned} \tag{4.2}$$

where $\boldsymbol{\delta}$ is an estimator of $\boldsymbol{\Sigma}^{-1}$. The merit of this loss function is that we can use the key inequality given in Theorem 3.1.

Consider first the simple class of estimators $c\boldsymbol{V}^+$, where c is a positive constant. The risk function of $c\boldsymbol{V}^+$ is calculated as

$$R(c\boldsymbol{V}^+, \boldsymbol{\Sigma}^{-1}) = E[L(c\boldsymbol{V}^+, \boldsymbol{\Sigma}^{-1})] = (n \wedge p)c^2 - 2npc + E[\mathrm{tr}[\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}]].$$

Thus the best estimator among the class is given by

$$\boldsymbol{\delta}^{BU} = \{np/(n \wedge p)\}\boldsymbol{V}^+ = (n \vee p)\boldsymbol{V}^+ = c_0^{-1}\boldsymbol{V}^+.$$

We next consider the class of estimators $(\boldsymbol{T}^t)^+\boldsymbol{D}\boldsymbol{T}^+$, where $\boldsymbol{T}$ is defined in Theorem 2.2, $\boldsymbol{T}^+ = (\boldsymbol{T}^t\boldsymbol{T})^{-1}\boldsymbol{T}^t$ is the Moore-Penrose inverse of $\boldsymbol{T}$ and $\boldsymbol{D}$ is an $(n \wedge p) \times (n \wedge p)$ diagonal matrix with constant diagonal elements. Denote $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_{n\wedge p})$. Using the same lines as in the proof of Theorem 2.2, we evaluate the risk function of $(\boldsymbol{T}^t)^+\boldsymbol{D}\boldsymbol{T}^+$ as

$$R((\boldsymbol{T}^t)^+\boldsymbol{D}\boldsymbol{T}^+, \boldsymbol{\Sigma}^{-1}) = \sum_{i=1}^{n\wedge p}\{d_i^2 - 2(n + p - 2i + 1)d_i\} + E[\mathrm{tr}[\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}]].$$

Hence the best estimator among the class $(\boldsymbol{T}^t)^+\boldsymbol{D}\boldsymbol{T}^+$ is

$$\boldsymbol{\delta}^{JS} = (\boldsymbol{T}^t)^+(\boldsymbol{D}^{JS})^{-1}\boldsymbol{T}^+,$$

where $\boldsymbol{D}^{JS}$ is defined in Theorem 2.2. It is here noted that $\boldsymbol{\delta}^{JS} = \{\widehat{\boldsymbol{\Sigma}}^{JS}\}^+$ is better than $\boldsymbol{\delta}^{BU} = \{\widehat{\boldsymbol{\Sigma}}^{US}\}^+$ relative to the loss (4.2).

We derive alternative estimators to $\boldsymbol{\delta}^{BU}$ and $\boldsymbol{\delta}^{JS}$. For $\boldsymbol{Q} = \boldsymbol{H}\boldsymbol{L}^{1/2}\boldsymbol{P}$ given below (3.1), we have $\boldsymbol{Q}^- = \boldsymbol{P}^t\boldsymbol{L}^{-1/2}\boldsymbol{H}^t$. It is noted that $\boldsymbol{Q}^-$ is the generalized inverse of $\boldsymbol{Q}$. Consider the class of estimators of the form

$$\boldsymbol{\delta}(\boldsymbol{\Psi}) = \boldsymbol{\delta}^{BU} + c_0^{-1}(\boldsymbol{Q}^-)^t\boldsymbol{\Psi}\boldsymbol{Q}^- = c_0^{-1}(\boldsymbol{V}^+ + (\boldsymbol{Q}^-)^t\boldsymbol{\Psi}\boldsymbol{Q}^-),$$

where $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{F})$ is an $\ell\times\ell$ diagonal matrix such that the diagonal elements are absolutely continuous functions of $\boldsymbol{F}$.

Theorem 4.1 *The risk difference between $\boldsymbol{\delta}(\boldsymbol{\Psi})$ and $\boldsymbol{\delta}^{BU}$ is given by*

$$\begin{aligned} &R(\boldsymbol{\delta}(\boldsymbol{\Psi}), \boldsymbol{\Sigma}^{-1}) - R(\boldsymbol{\delta}^{BU}, \boldsymbol{\Sigma}^{-1}) \\ &= c_0^{-2}E\left[\sum_{i=1}^{\ell}\left\{\psi_i^2 - 2(c_0\alpha_i - 1)\psi_i + 4c_0 f_i\frac{\partial\psi_i}{\partial f_i} + 4c_0\sum_{j>i}^{\ell}\frac{\psi_i - \psi_j}{f_i - f_j}f_j\right\}\right], \end{aligned}$$

where $\alpha_i = |n - p| + 2i - 1$ for $i = 1, \ldots, \ell$.

Proof. It is noted that $\boldsymbol{Q}^- \boldsymbol{V} = \boldsymbol{Q}^t$ and $\boldsymbol{Q}^- \boldsymbol{V}(\boldsymbol{Q}^t)^- = \boldsymbol{I}_\ell$. The risk difference between $\boldsymbol{\delta}(\boldsymbol{\Psi})$ and $\boldsymbol{\delta}^{BU}$ is expressed as

$$
\begin{aligned}
&R(\boldsymbol{\delta}(\boldsymbol{\Psi}), \boldsymbol{\Sigma}^{-1}) - R(\boldsymbol{\delta}^{BU}, \boldsymbol{\Sigma}^{-1}) \\
&= E[2c_0^{-1}\mathrm{tr}(\boldsymbol{\delta}^{BU} - \boldsymbol{\Sigma}^{-1})\boldsymbol{V}(\boldsymbol{Q}^t)^- \boldsymbol{\Psi}\boldsymbol{Q}^-\boldsymbol{V} + c_0^{-2}\mathrm{tr}(\boldsymbol{Q}^t)^-\boldsymbol{\Psi}\boldsymbol{Q}^-\boldsymbol{V}(\boldsymbol{Q}^t)^-\boldsymbol{\Psi}\boldsymbol{Q}^-\boldsymbol{V}] \\
&= c_0^{-2}E[2\mathrm{tr}\boldsymbol{\Psi} - 2c_0\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\boldsymbol{\Psi}\boldsymbol{Q}^t + \mathrm{tr}\boldsymbol{\Psi}^2].
\end{aligned}
$$

Using the same arguments as (3.10) gives the result of this theorem. □

For examples of $\boldsymbol{\delta}(\boldsymbol{\Psi})$ improving on $\boldsymbol{\delta}^{BU}$, the Stein type estimator $\boldsymbol{\delta}(\boldsymbol{\Psi}^{ST})$ is defined as

$$
\boldsymbol{\Psi}^{ST} = \mathrm{diag}(\psi_1^{ST}, \ldots, \psi_\ell^{ST}), \quad \psi_i^{ST} = c_0\alpha_i - 1 \quad (i = 1, \ldots, \ell).
$$

It can easily be verified that $\boldsymbol{\delta}(\boldsymbol{\Psi}^{ST})$ dominates $\boldsymbol{\delta}^{BU}$ relative to the loss (4.2) and also, if $\ell = n \wedge p$, then $\boldsymbol{\delta}(\boldsymbol{\Psi}^{ST})$ is superior to $\boldsymbol{\delta}^{JS}$.

We next consider the Haff type estimator $\boldsymbol{\delta}(\boldsymbol{\Psi}^{HF})$, where

$$
\boldsymbol{\Psi}^{HF} = \mathrm{diag}(\psi_1^{HF}, \ldots, \psi_\ell^{HF}), \quad \psi_i^{HF} = \frac{a}{\mathrm{tr}\boldsymbol{F}} f_i \quad (i = 1, \ldots, \ell),
$$

where a is a nonpositive constant. Applying Theorem 4.1 to $\boldsymbol{\delta}(\boldsymbol{\Psi}^{HF})$, we obtain

$$
\begin{aligned}
&R(\boldsymbol{\delta}(\boldsymbol{\Psi}^{HF}), \boldsymbol{\Sigma}^{-1}) - R(\boldsymbol{\delta}^{BU}, \boldsymbol{\Sigma}^{-1}) \\
&= a^2 E\left[\frac{\mathrm{tr}\boldsymbol{F}^2}{(\mathrm{tr}\boldsymbol{F})^2}\right] + 2c_0(n \wedge p - 1)a + 4c_0 a E\left[\sum_{i=1}^{\ell} \frac{f_i}{\mathrm{tr}\boldsymbol{F}}\left(1 - \frac{f_i}{\mathrm{tr}\boldsymbol{F}}\right)\right],
\end{aligned}
$$

which is bounded above by $a^2 + 2c_0(n \wedge p - 1)a$. Thus, when $-2c_0(n \wedge p - 1) \le a < 0$, $\boldsymbol{\delta}(\boldsymbol{\Psi}^{HF})$ dominates $\boldsymbol{\delta}^{BU}$ relative to the loss (4.2).

For improvement on estimators $\boldsymbol{\delta}(\boldsymbol{\Psi})$ with $\boldsymbol{\Psi} = \mathrm{diag}(\psi_1, \ldots, \psi_\ell)$, we can apply the truncation method. Let $[\boldsymbol{\Psi}]^{TR} = \mathrm{diag}(\psi_1^{TR}, \ldots, \psi_\ell^{TR})$ be a diagonal matrix of order ℓ such that the i-th diagonal element is given by $\psi_i^{TR} = \max\{\psi_i,\ c_0(n \vee p + m)(1 + f_i)^{-1} - 1\}$. Theorem 3.1 provides the following proposition.

Proposition 4.1 *If* $\Pr([\boldsymbol{\Psi}]^{TR} \neq \boldsymbol{\Psi}) > 0$*, then, for any tuple of positive integers* n*,* p *and* m*, the truncated estimator* $\boldsymbol{\delta}([\boldsymbol{\Psi}]^{TR})$ *dominates* $\boldsymbol{\delta}(\boldsymbol{\Psi})$ *relative to the loss* (4.2)*.*

Proof. Since

$$
\psi_i^{TR} = \max\{\psi_i,\ c_0(n \vee p + m)(1 + f_i)^{-1} - 1\} \ge \psi_i \tag{4.3}
$$

for each i, it is observed that

$$
\begin{aligned}
&R(\boldsymbol{\delta}(\boldsymbol{\Psi}), \boldsymbol{\Sigma}^{-1}) - R(\boldsymbol{\delta}([\boldsymbol{\Psi}]^{TR}), \boldsymbol{\Sigma}^{-1}) \\
&= c_0^{-2}E[\mathrm{tr}\{\boldsymbol{\Psi}^2 + 2\boldsymbol{\Psi} - ([\boldsymbol{\Psi}]^{TR})^2 - 2[\boldsymbol{\Psi}]^{TR}\} - 2c_0\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}\{\boldsymbol{\Psi} - [\boldsymbol{\Psi}]^{TR}\}\boldsymbol{Q}^t] \\
&\ge c_0^{-2}E[\mathrm{tr}\{\boldsymbol{\Psi}^2 + 2\boldsymbol{\Psi} - ([\boldsymbol{\Psi}]^{TR})^2 - 2[\boldsymbol{\Psi}]^{TR}\} - 2c_0(n \vee p + m)\mathrm{tr}(\boldsymbol{I}_\ell + \boldsymbol{F})^{-1}\{\boldsymbol{\Psi} - [\boldsymbol{\Psi}]^{TR}\}] \\
&= c_0^{-2}E[\mathrm{tr}\{\boldsymbol{\Psi} - [\boldsymbol{\Psi}]^{TR}\}\{\boldsymbol{\Psi} + [\boldsymbol{\Psi}]^{TR} + 2\boldsymbol{I}_\ell - 2c_0(n \vee p + m)(\boldsymbol{I}_\ell + \boldsymbol{F})^{-1}\}],
\end{aligned} \tag{4.4}
$$

where the inequality is verified by Theorem 3.1. The last r.h.s. of (4.4) is expressed by

$$c_0^{-2} E\Big[\sum_{i=1}^{\ell}(\psi_i-\psi_i^{TR})\Big\{\psi_i+\psi_i^{TR}+2-2c_0\frac{n\vee p+m}{1+f_i}\Big\}\Big]. \tag{4.5}$$

Once again using (4.3), we can see that the summation in (4.5) is bounded below by

$$\begin{aligned}&\sum_{i=1}^{\ell}(\psi_i-\psi_i^{TR})\Big\{\psi_i^{TR}+\psi_i^{TR}+2-2c_0\frac{n\vee p+m}{1+f_i}\Big\}\\&=2\sum_{i=1}^{\ell}(\psi_i-\psi_i^{TR})\Big\{\psi_i^{TR}+1-c_0\frac{n\vee p+m}{1+f_i}\Big\},\end{aligned}$$

which is equal to zero. Hence the proof is complete. □

Using Proposition 4.1 immediately yields the following corollary.

Corollary 4.1 *The truncated estimator* $\boldsymbol{\delta}([\boldsymbol{\Psi}^{ST}]^{TR})$ *dominates* $\boldsymbol{\delta}(\boldsymbol{\Psi}^{ST})$ *relative to the loss* (4.2)*. Also,* $\boldsymbol{\delta}([\boldsymbol{\Psi}^{HF}]^{TR})$ *dominates* $\boldsymbol{\delta}(\boldsymbol{\Psi}^{HF})$ *relative to the loss* (4.2)*.*

References

[1] Amemiya, Y. (1985). What should be done when an estimated between-group covariance matrix is not nonnegative definite ? *Amer. Statist.*, **39**, 112–117.

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

[3] Calvin, J.A., and Dykstra, R.L. (1991). Maximum likelihood estimation of a set of covariance matrices under Loewner order restrictions with applications to balanced multivariate variance components models. *Ann. Statist.*, **19**, 850–869.

[4] Chételat, D. and Wells, M.T. (2012). Improved multivariate normal mean estimation with unknown covariance when p is greater than n, *Ann. Statist.*, **40**, 3137–3160.

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

[6] Díaz-García, J.A., Gutierrez-Jáimez, R. and Mardia, K.V. (1997). Wishart and pseudo-Wishart distributions and some applications to shape theory, *J. Multivariate Anal.*, **63**, 73–87.

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

[8] James, W. and Stein, C. (1961). Estimation with quadratic loss, In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, **1**, 361–380, Univ. of California Press, Berkeley.

[9] Konno, Y. (1992). Improved estimation of matrix of normal mean and eigenvalues in the multivariate F-distribution, Doctoral dissertation, Institute of Mathematics, University of Tsukuba. This is downloadable from his website (http://mcm-www.jwu.ac.jp/˜konno/).

[10] Konno, Y. (2009). Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss, *J. Multivariate Anal.*, **100**, 2237–2253.

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

[12] Kubokawa, T. and Srivastava, M.S. (2003). Estimating the covariance matrix: a new approach. *J. Multivariate Anal.*, **86**, 28–47.

[13] Kubokawa, T. and Tsai, M.-T. (2006). Estimation of covariance matrices in fixed and mixed effects linear models, *J. Multivariate Anal.*, **97**, 2242–2261.

[14] Mathew, T., Niyogi, A., and Sinha, B.K. (1994). Improved nonnegative estimation of variance components in balanced multivariate mixed models. *J. Multivariate Anal.*, **51**, 83–101.

[15] Muirhead, R.J. (1982). *Aspects of multivariate statistical theory*, John Wiley & Sons, New York.

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

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

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

[19] Srivastava, M.S. and Kubokawa, T. (1999). Improved nonnegative estimation of multivariate components of variance. *Ann. Statist.*, **27**, 2008–2032.

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

[21] Stein, C. (1973). Estimation of the mean of a multivariate normal distribution, In *Proc. Prague Symp. Asymptotic Statist.*, 345–381.

[22] Stein, C. (1977). Lectures on the theory of estimation of many parameters. In *Studies in the Statistical Theory of Estimation, Part I* (I.A. Ibragimov and M.S. Nikulin, eds.). *Proceedings of Scientific Seminars of the Steklov Institute, Leningrad Division* **74**, 4–65. (In Russian)

[23] Takemura, A. (1984). *Zonal polynomials*, Lecture Notes Monograph Series Vol.4, Institute of Mathematical Statistics, Hayward.

[24] Tsukuma, H. and Kubokawa, T. (2014). A unified approach to estimating a normal mean matrix in high and low dimensions, Discussion Paper Series, CIRJE-F-926, Faculty of Economics, The University of Tokyo. This is downloadable from http://www.cirje.e.u-tokyo.ac.jp/index.html.

Table 1: Estimated PRIR (%) in estimation of the covariance matrix

$\boldsymbol{\Sigma}$	$\boldsymbol{\Theta}$	n	(m,p)	$\widehat{\boldsymbol{\Sigma}}^{HF}$	$\widehat{\boldsymbol{\Sigma}}^{HFTR}$	$\widehat{\boldsymbol{\Sigma}}^{ST}$	$\widehat{\boldsymbol{\Sigma}}^{STTR}$	$\widehat{\boldsymbol{\Sigma}}^{JSTR}$	$\widehat{\boldsymbol{\Sigma}}^{DS}$	$\widehat{\boldsymbol{\Sigma}}^{DSTR}$
(A)	(a)	15	$(10,5)$	1.2	9.8	12.5	14.3	18.2	24.4	35.6
		10	$(15,5)$	2.3	10.4	13.6	15.1	18.7	18.9	30.3
		5	$(15,10)$	2.3	10.4	13.7	15.2	18.8	19.0	30.6
		15	$(5,10)$	−0.9	−0.8	3.0	3.5	10.0	17.4	27.2
		10	$(5,15)$	−0.9	−0.8	2.9	3.4	10.0	17.3	27.1
		5	$(10,15)$	1.2	9.8	12.5	14.3	18.2	24.4	35.6
(A)	(b)	15	$(10,5)$	0.4	6.2	6.3	7.2	11.7	24.4	31.6
		10	$(15,5)$	1.3	6.7	7.2	8.0	12.4	18.9	26.7
		5	$(15,10)$	1.1	6.5	6.8	7.6	11.8	19.0	26.6
		15	$(5,10)$	−1.8	−1.7	0.5	0.7	7.3	17.4	25.1
		10	$(5,15)$	−1.8	−1.8	0.5	0.7	7.2	17.3	25.0
		5	$(10,15)$	0.0	6.0	5.9	6.8	11.3	24.4	31.3
(A)	(c)	15	$(10,5)$	0.0	5.5	5.3	6.1	10.6	24.4	30.8
		10	$(15,5)$	0.8	5.8	5.7	6.4	11.0	18.9	25.7
		5	$(15,10)$	0.3	5.3	4.7	5.4	9.7	19.0	25.1
		15	$(5,10)$	−2.3	−2.2	−0.4	−0.1	6.5	17.4	24.4
		10	$(5,15)$	−2.4	−2.3	−0.4	−0.2	6.4	17.3	24.3
		5	$(10,15)$	−0.6	4.9	4.1	4.9	9.5	24.4	30.0
(B)	(a)	15	$(10,5)$	1.2	9.8	12.5	14.3	18.2	16.3	29.8
		10	$(15,5)$	2.3	10.4	13.6	15.1	18.7	14.2	27.6
		5	$(15,10)$	1.7	7.3	9.7	10.7	14.5	11.7	20.6
		15	$(5,10)$	−0.9	−0.8	3.0	3.5	10.0	13.0	22.9
		10	$(5,15)$	−0.7	−0.6	2.2	2.5	8.0	11.1	18.3
		5	$(10,15)$	0.7	5.5	7.1	8.1	11.3	12.0	18.8
(B)	(b)	15	$(10,5)$	0.6	6.8	7.6	8.5	16.2	16.3	23.4
		10	$(15,5)$	1.7	7.4	9.0	9.8	16.9	14.2	22.7
		5	$(15,10)$	1.4	5.9	8.0	8.6	12.9	11.7	19.3
		15	$(5,10)$	−1.4	−1.3	1.5	1.8	9.7	13.0	20.1
		10	$(5,15)$	−0.9	−0.8	1.7	1.9	7.7	11.1	17.6
		5	$(10,15)$	0.6	4.7	6.3	7.0	10.3	12.0	18.2
(B)	(c)	15	$(10,5)$	−0.1	5.4	4.6	5.4	14.0	16.3	20.6
		10	$(15,5)$	0.8	5.9	5.7	6.4	14.5	14.2	19.8
		5	$(15,10)$	0.4	4.0	4.0	4.5	9.0	11.7	16.3
		15	$(5,10)$	−2.4	−2.4	−0.4	−0.2	8.5	13.0	17.8
		10	$(5,15)$	−1.6	−1.5	0.1	0.2	6.3	11.1	16.0
		5	$(10,15)$	−0.0	3.2	3.2	3.7	7.1	12.0	15.8