CIRJE-F-62

Prediction in Multivariate Mixed Linear Models with Equal Replications

Tatsuya Kubokawa
University of Tokyo

M. S. Srivastava
University of Toronto

October 1999

Prediction in Multivariate Mixed Linear Models with Equal Replications

T. Kubokawa and M.S. Srivastava
University of Tokyo and *University of Toronto*

The multivariate mixed linear model or multivariate components of variance model with equal replications is considered. The paper addresses the problem of predicting the sum of the regression mean and the random effects. When the feasible best linear unbiased predictors or empirical Bayes predictors are used, this prediction problem reduces to the estimation of the ratio of two covariance matrices. For the estimation of the ratio matrix, the James-Stein type estimators based on the Bartlett's decomposition, the Stein type orthogonally equivariant estimators and the Efron-Morris type estimators are obtained. Their dominance properties over the usual estimators including the unbiased one are established, and further domination results are shown by using information of order restriction between the two covariance matrices. It is also demonstrated that the empirical Bayes predictors that employs these improved estimators of the ratio of the two covariance matrices have uniformly smaller risks than the crude Efron-Morris type estimator in the context of estimation of a matrix mean in a fixed effects linear regression model where the components are unknown parameters. Finally, an extension of the model is given.

Key words and phrases: Multivariate mixed linear model, multivariate components of variances, small area estimation, prediction, shrinkage estimation, empirical Bayes procedure, order restriction, orthogonal equivariance, covariance matrix, Bartlett's decomposition, decision-theory.

AMS subject classifications: Primary 62F11, 62J07, Secondary 62C15, 62C20.

Tatsuya Kubokawa
Faculty of Economics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, JAPAN
E-Mail: tatsuya@e.u-tokyo.ac.jp, Fax: +81-3-5841-5521, Tel: +81-3-5841-5656

M. S. Srivastava
Department of Statistics, University of Toronto, 100 St George Street, Toronto, Ontario, CANADA M5S 3G3
E-Mail: srivasta@utstat.utstat.toronto.edu, Fax: +1-416-978-5133,
Tel: +1-416-978-4450

1 Introduction

Mixed linear models or variance components models have been effectively and extensively employed in practical data-analysis when the response is univariate. For example, in estimation of small area means, they have been used as a method of pooling or smoothing data to strengthen the accuracy of the estimators of small area means. Mixed linear models are related to empirical Bayes models. Practical and theoretical studies of using these models have been given by Efron and Morris (1975), Fay and Herriot (1979), Casella (1985), Battese, Harter and Fuller (1988), Prasad and Rao (1990), Ghosh and Rao (1994) and others referred to in their papers.

In contrast to these activities in the univariate mixed linear models, the multivariate mixed linear models have received little attention. Fuller and Harter (1987) proposed a predictor of multivariate small area means and Amemiya (1994) extended the results to more general models. In predicting such quantity in high dimension, it is plausible to improve upon ordinary predictors through the Stein effects. However no exact results for the improvement have been studied.

In this paper, we treat the multivariate mixed linear models with equal replications and consider the problem of predicting multivariate small area means relative to a quadratic loss function. When the parameters of the model is known, the best linear unbiased predictor (BLUP) would be employed. The BLUP is interpreted as a Bayes rule. Since the parameters are unknown in the model, their estimators are substituted in the BLUP, and the substituted predictor is called the feasible BLUP or empirical Bayes predictor. The empirical Bayes predictor lies in between the predictor associated with each small area and the estimator pooling whole data. In the univariate case, the empirical Bayes predictor is the James-Stein estimator of a mean vector of a multivariate normal distribution (Stein (1956, 73, 81) and James and Stein (1961)). The relation between the predictor of random components and the James-Stein estimator has been pointed out by Efron and Morris (1972, 76), Peixoto and Harville (1986) and Sun (1996). For survey and account of the Stein problem, see Berger (1985), Brandwein and Strawderman (1990), Lehmann and Casella (1998) and Kubokawa (1998).

In Section 2, it is shown that the problem of predicting small area means is reduced to that of estimating the ratio $\boldsymbol{\Delta}$ of covariance matrices. This is an extension of the results given by Efron and Morris (1976). The estimator of the ratio matrix $\boldsymbol{\Delta}$ is important in the empirical Bayes predictor as it determines the extent to which the estimator associated with each small area should be shrunken towards the pooled estimator. The estimation of $\boldsymbol{\Delta}$ is related with the estimation of a scale matrix of a multivariate F-distribution, which was studied by Konno (1992b), Bilodeau and Srivastava (1992) among others.

In estimation of covariance matrix and of ratio of covariance matrices, several types of estimators improving upon usual estimators such as unbiased ones are available. One of them is the James-Stein type estimator based on the Bartlett's lower triangular decomposition. It is known that the James-Stein type estimator depends on the coordinate system. To eliminate this shortcoming, Stein (1977) and Dey and Srinivasan (1985) considered the Stein type orthogonally equivariant estimator and established its dominance property over the James-Stein type estimator. Haff (1980) showed that the Efron-Morris type orthogonally equivariant estimator dominates the usual one. Corresponding to a series of these dominance results, for the ratio matrix $\boldsymbol{\Delta}$, we obtain two James-Stein type estimators, three Stein type estimators and three Efron-Morris type estimators, given in Sections 3, 4 and 5, respectively, and show their dominance properties over the usual estimator including the unbiased one. It is noted in Section 4 that the estimation of $\boldsymbol{\Delta}$ in our setup has a story different from the estimation of the covariance matrix, that is, the Stein type estimators do not always dominate the corresponding James-Stein estimators.

In Section 6, the problem of estimating a matrix mean is addressed in a fixed effects linear regression model where the components for area effects are unknown parameters. In this context, it is shown that procedures improving upon the crude Efron-Morris (1976) estimator can be automatically produced by the empirical Bayes predictors that uses improved estimators of $\boldsymbol{\Delta}$. In the proofs of these results, the unbiased estimators of the risk functions are derived by using the same arguments as in Efron and Morris (1976) concerning the completeness of the Wishart distribution. These unbiased estimators of the risk functions are the same as given in Konno (1990, 92a). However, our derivation using the Efron-Morris' argument is much simpler in calculation. For other studies of estimation of the matrix mean, see Zidek (1978), Zheng (1986), Bilodeau and Kariya (1989), Konno (1991), Shieh (1993) and Kariya, Konno and Strawderman (1996, 99). Finally, an extension of the model is given in Section 7.

2 A Mixed Linear Model and Specification of the Estimation Problem

In this paper, we deal with the following multivariate mixed linear model with equal replications:

$$\boldsymbol{y}_{ij} = \boldsymbol{\beta}\boldsymbol{b}_i + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, r, \tag{2.1}$$

where $\boldsymbol{y}_{ij}$'s are p-variate observation vectors, $\boldsymbol{\beta}$ is a $p \times q$ common unknown regression coefficient, $\boldsymbol{b}_i$'s are $q \times 1$ covariates, $\boldsymbol{\alpha}_i$'s are $p \times 1$ random effects having a p-variate normal distribution $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}_A)$, and $\boldsymbol{\epsilon}_{ij}$'s are $p \times 1$ random error terms having $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. It is supposed that $\boldsymbol{\alpha}_i$'s and $\boldsymbol{\epsilon}_{ij}$'s are mutually independent and that $\boldsymbol{\Sigma}_A$ and $\boldsymbol{\Sigma}$ are unknown positive-definite dispersion matrices.

Let $\boldsymbol{\theta}_i = \boldsymbol{\beta}\boldsymbol{b}_i + \boldsymbol{\alpha}_i$ and $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k)$. Then the problem we consider in this paper is to predict the $p \times k$ matrix $\boldsymbol{\Theta}$ where estimator $\widehat{\boldsymbol{\Theta}}$ of $\boldsymbol{\Theta}$ is evaluated in terms of the risk function

$$R_m(\widehat{\boldsymbol{\Theta}}, \omega) = E_\omega\left[\mathrm{tr}\,\left(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\right)' \boldsymbol{\Sigma}^{-1}\left(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\right)\right] \tag{2.2}$$

for $\omega = (\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_A)$, unknown parameters.

The multivariate mixed linear model is also interpreted as an empirical Bayes model: $\boldsymbol{y}_{ij} \sim \mathcal{N}_p(\theta_i, \boldsymbol{\Sigma})$ and $\boldsymbol{\theta}_i$ having prior distribution $\mathcal{N}_p(\boldsymbol{\beta}\boldsymbol{b}_i, \boldsymbol{\Sigma}_A)$. In this situation, the Bayes estimator of $\boldsymbol{\theta}_i$ is given by

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_i^B &= \widehat{\boldsymbol{\theta}}_i^B(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_A) \\ &= \boldsymbol{\beta}\boldsymbol{b}_i + \left(\boldsymbol{\Sigma}^{-1} + r^{-1}\boldsymbol{\Sigma}_A^{-1}\right)^{-1} \boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{y}}_{i\cdot} - \boldsymbol{\beta}\boldsymbol{b}_i) \\ &= \overline{\boldsymbol{y}}_{i\cdot} - \boldsymbol{\Sigma}\left(\boldsymbol{\Sigma} + r\boldsymbol{\Sigma}_A\right)^{-1}(\overline{\boldsymbol{y}}_{i\cdot} - \boldsymbol{\beta}\boldsymbol{b}_i),\end{aligned}$$

for $\overline{\boldsymbol{y}}_{i\cdot} = r^{-1}\sum_{j=1}^{r} \boldsymbol{y}_{ij}$. This is the best linear unbiased predictor (BLUP) in our setup.

Since the Bayes estimator $\widehat{\boldsymbol{\theta}}_i^B$ depends on the unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_A$. they should be estimated from the marginal distribution. Let

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \sum_{i=1}^{k} \overline{\boldsymbol{y}}_{i\cdot}\boldsymbol{b}_i' \left(\sum_{i=1}^{k} \boldsymbol{b}_i\boldsymbol{b}_i'\right)^{-}, \\ \boldsymbol{S} &= \sum_{i=1}^{k}\sum_{j=1}^{r} \left(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_{i\cdot}\right)\left(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_{i\cdot}\right)', \\ \boldsymbol{W} &= r\sum_{i=1}^{k} \left(\overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\beta}}\boldsymbol{b}_i\right)\left(\overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\beta}}\boldsymbol{b}_i\right)',\end{aligned}$$

where $\boldsymbol{A}^-$ denotes the generalized inverse of matrix $\boldsymbol{A}$. Assume that the rank of $\sum_{i=1}^{k} \boldsymbol{b}_i\boldsymbol{b}_i'$ is q_1 with $q_1 \le q < p$. Then it is seen that

$$\begin{aligned}\boldsymbol{S} &\sim \mathcal{W}_p(\boldsymbol{\Sigma}, n), & n &= k(r-1), \\ \boldsymbol{W} &\sim \mathcal{W}_p(\boldsymbol{\Sigma}_2, m), & m &= k - q_1 \\ \boldsymbol{\Sigma}_2 &= \boldsymbol{\Sigma} + r\boldsymbol{\Sigma}_A,\end{aligned}$$

and that $\boldsymbol{S}$, $\boldsymbol{W}$ and $\widehat{\boldsymbol{\beta}}$ are mutually independent. The parameters $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_A$ (or $\boldsymbol{\Sigma}_2$) are estimated by $\widehat{\boldsymbol{\beta}}$ and estimators $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Sigma}}_A$ (or $\widehat{\boldsymbol{\Sigma}}_2$) based on $\widehat{\boldsymbol{\beta}}$, $\boldsymbol{S}$ and $\boldsymbol{W}$. This result in an empirical Bayes estimator (or feasible BLUP) of the form

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_i^{EB} &= \widehat{\boldsymbol{\theta}}_i^B(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\Sigma}}_A) \\ &= \overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\Sigma}} + r\widehat{\boldsymbol{\Sigma}}_A)^{-1}(\overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\beta}}\boldsymbol{b}_i) \\ &= \overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_2^{-1}(\overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\beta}}\boldsymbol{b}_i),\end{aligned} \tag{2.3}$$

which means that the sample mean of the i-th small area is shrunk towards the common value that uses all the data. It is known that $\overline{\boldsymbol{y}}_{i\cdot}$ has an unstable variance because of small data in the i-th small area. But this undesirable property can be avoided by using the estimator $\widehat{\boldsymbol{\theta}}_i^{EB}$ which borrows the data from the surrounding small area. The ratio

$\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_2^{-1}$ of the estimators of covariance matrices determines the extent to which $\overline{\boldsymbol{y}}_{i\cdot}$ should be shrunk. Since the parameter space is restricted as

$$\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Sigma}+r\boldsymbol{\Sigma}_A)^{-1}\boldsymbol{\Sigma}^{1/2}=\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}^{1/2}\le \boldsymbol{I}_p, \tag{2.4}$$

the ratio should be in $\mathbf{0}\le \widehat{\boldsymbol{\Sigma}}^{1/2}\widehat{\boldsymbol{\Sigma}}_2^{-1}\widehat{\boldsymbol{\Sigma}}^{1/2}\le \boldsymbol{I}_p$, where $\boldsymbol{A}^{1/2}$ denotes the factorizable of the symmetric matrix $\boldsymbol{A}$ and $\boldsymbol{A}\le\boldsymbol{B}$ means that $\boldsymbol{B}-\boldsymbol{A}$ is non-negative definite. Two extreme cases of taking $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}^{1/2}=\mathbf{0}$ and $=\boldsymbol{I}_p$ yield $\overline{\boldsymbol{y}}_{i\cdot}$ and $\widehat{\boldsymbol{\beta}}\boldsymbol{b}_i$, respectively, both of which are inappropriate. It is reasonable to choose $\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_2^{-1}$ such that $\widehat{\boldsymbol{\theta}}_i^{EB}$ has a uniformly smaller risk than existing estimators including $\overline{\boldsymbol{y}}_{i\cdot}$.

Let $\widehat{\boldsymbol{\Theta}}^{EB}=(\widehat{\boldsymbol{\theta}}_1^{EB},\dots,\widehat{\boldsymbol{\theta}}_k^{EB})$. Noting that

$$\begin{aligned}
E_\omega\left[\mathrm{tr}\left(\widehat{\boldsymbol{\Theta}}^B-\boldsymbol{\Theta}\right)'\boldsymbol{\Sigma}^{-1}\left(\widehat{\boldsymbol{\Theta}}^B-\boldsymbol{\Theta}\right)\right] &= k\mathrm{tr}\left(r\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Sigma}_A^{-1}\right)^{-1}\boldsymbol{\Sigma}^{-1}\\
&= \frac{k}{r}\mathrm{tr}\left(\boldsymbol{\Sigma}-\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}+r\boldsymbol{\Sigma}_A\right)^{-1}\boldsymbol{\Sigma}\right)\boldsymbol{\Sigma}^{-1}\\
&= \frac{pk}{r}-\frac{k}{r}\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1},
\end{aligned}$$

we see that the risk function of $\widehat{\boldsymbol{\Theta}}^{EB}$ is written by

$$\begin{aligned}
&R_m(\widehat{\boldsymbol{\Theta}}^{EB},\omega)\\
&= E_\omega\left[\mathrm{tr}\,(\widehat{\boldsymbol{\Theta}}^{EB}-\widehat{\boldsymbol{\Theta}}^B+\widehat{\boldsymbol{\Theta}}^B-\boldsymbol{\Theta})'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Theta}}^{EB}-\widehat{\boldsymbol{\Theta}}^B+\widehat{\boldsymbol{\Theta}}^B-\boldsymbol{\Theta})\right]\\
&= E_\omega\left[\mathrm{tr}\,(\widehat{\boldsymbol{\Theta}}^{EB}-\widehat{\boldsymbol{\Theta}}^B)'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Theta}}^{EB}-\widehat{\boldsymbol{\Theta}}^B)\right]+E_\omega\left[\mathrm{tr}\,(\widehat{\boldsymbol{\Theta}}^B-\boldsymbol{\Theta})'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Theta}}^B-\boldsymbol{\Theta})\right]\\
&= \sum_{i=1}^k E_\omega\left[\mathrm{tr}\left\{(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1})'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1})(\overline{\boldsymbol{y}}_{i\cdot}-\widehat{\boldsymbol{\beta}}\boldsymbol{b}_i)(\overline{\boldsymbol{y}}_{i\cdot}-\widehat{\boldsymbol{\beta}}\boldsymbol{b}_i)'\right\}\right]\\
&\quad +\sum_{i=1}^k E_\omega\left[\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\boldsymbol{b}_i\boldsymbol{b}_i'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\right]+\frac{pk}{r}-\frac{k}{r}\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1} \qquad (2.5)\\
&= r^{-1}E_\omega\left[\mathrm{tr}\left\{(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1})'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1})\boldsymbol{W}\right\}\right]\\
&\quad +r^{-1}\left(pk-m\mathrm{tr}\,\boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1}\right),
\end{aligned}$$

so that the risk of $\widehat{\boldsymbol{\Theta}}^{EB}$ depends on the risk of the estimators of the ratio of covariance matrices. We define this risk by

$$R(\widehat{\boldsymbol{\Delta}},\omega)=E_\omega\left[\mathrm{tr}\,(\widehat{\boldsymbol{\Delta}}-\boldsymbol{\Delta})'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Delta}}-\boldsymbol{\Delta})\boldsymbol{W}\right] \tag{2.6}$$

for $\boldsymbol{\Delta}=\boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1}$ and its estimator $\widehat{\boldsymbol{\Delta}}$. We shall look for an estimator $\boldsymbol{\Delta}$ that has a smaller risk $R(\widehat{\boldsymbol{\Delta}},\omega)$.

The usual unbiased estimator of $\boldsymbol{\Delta}$ is

$$\widehat{\boldsymbol{\Delta}}^{UB} = n^{-1}(m-p-1)\boldsymbol{S}\boldsymbol{W}^{-1}$$

with risk

$$R(\widehat{\boldsymbol{\Delta}}^{UB}, \omega) = \left\{n^{-1}(n-p-1)(m-p-1)+m\right\} \operatorname{tr} \boldsymbol{\Delta}.$$

Since the crude estimator $(\boldsymbol{y}_{1\cdot}, \ldots, \boldsymbol{y}_{k\cdot})$ of $\boldsymbol{\Theta}$ corresponds to the the estimator $\widehat{\boldsymbol{\Delta}} = \boldsymbol{0}$, its risk has $R(\boldsymbol{0}, \omega) = m \operatorname{tr} \boldsymbol{\Delta}$ and this shows that $\widehat{\boldsymbol{\Delta}}^{UB}$ is better than the crude estimator $\widehat{\boldsymbol{\Delta}} = \boldsymbol{0}$ for $m > p+1$ and $n > p+1$. More generally the estimator $\widehat{\boldsymbol{\Delta}}(a) = a\boldsymbol{S}\boldsymbol{W}^{-1}$, a multiple of $\boldsymbol{S}\boldsymbol{W}^{-1}$, has the risk

$$R(\widehat{\boldsymbol{\Delta}}(a), \omega) = \left\{n(n+p+1)(m-p-1)^{-1}a^2 - 2na + m\right\} \operatorname{tr} \boldsymbol{\Delta},$$

which can be minimized at $a = (m-p-1)/(n+p+1)$ with risk

$$R(\widehat{\boldsymbol{\Delta}}_0, \omega) = \left\{-n(n+p+1)^{-1}(m-p-1)+m\right\} \operatorname{tr} \boldsymbol{\Delta},$$

where

$$\begin{aligned} \widehat{\boldsymbol{\Delta}}_0 &= \widehat{\boldsymbol{\Delta}}((n+p+1)^{-1}(m-p-1)) \\ &= (n+p+1)^{-1}(m-p-1)\boldsymbol{S}\boldsymbol{W}^{-1}. \end{aligned} \tag{2.7}$$

In the following sections, we shall obtain several types of estimators of $\boldsymbol{\Delta}$ improving upon $\widehat{\boldsymbol{\Delta}}_0$ relative to the risk $R(\widehat{\boldsymbol{\Delta}}, \omega)$.

3 James-Stein Type Improved Estimators

We shall derive two types of James-Stein type estimators based on Bartlett's decomposition, and show that they improve upon the estimator $\widehat{\boldsymbol{\Delta}}_0$ with the best multiplier.

Let G_L^+ denote the group of lower triangular matrices with positive diagonal elements. Let $\boldsymbol{T}$ and $\boldsymbol{U}$ be matrices in G_L^+ such that $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{U}'$ and $\boldsymbol{W} = \boldsymbol{T}\boldsymbol{T}'$. Then we can consider two types of James-Stein estimators:

$$\widehat{\boldsymbol{\Delta}}_1^{JS} = (n+p+1)^{-1}\boldsymbol{S}\boldsymbol{T}'^{-1}\boldsymbol{C}\boldsymbol{T}^{-1}, \tag{3.1}$$

$$\widehat{\boldsymbol{\Delta}}_2^{JS} = (m-p-1)\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'\boldsymbol{W}^{-1}, \tag{3.2}$$

where $\boldsymbol{C} = \operatorname{diag}(c_1, \ldots, c_p)$ and $\boldsymbol{D} = \operatorname{diag}(d_1, \ldots, d_p)$ for

$$c_i = m-i-1, \qquad i = 1, \ldots, p, \tag{3.3}$$

$$d_i = \frac{1}{n+p+3-2i}\prod_{j=1}^{i-1}\left(1 - \frac{1}{n+p+3-2j}\right), \quad i = 2, \ldots, p, \tag{3.4}$$

and $d_1 = (n+p+1)^{-1}$. For convenience, we use the notations $\widehat{\boldsymbol{\Delta}}_1$ and $\widehat{\boldsymbol{\Delta}}_2$ for the estimators (3.1) and (3.2) with general constants c_i's and d_i's.

3.1 Dominance property of $\hat{\Delta}_1^{JS}$

We first evaluate the risk function of $\widehat{\boldsymbol{\Delta}}_1^{JS}$ relative to the risk (2.6). Let ξ_{ii} be the (i,i) element of $\boldsymbol{\Xi}$ where $\boldsymbol{\Xi} = \widetilde{\boldsymbol{B}}^{-1}\boldsymbol{\Sigma}\widetilde{\boldsymbol{B}}'^{-1}$ for $\widetilde{\boldsymbol{B}} \in G_L^+$ such that $\boldsymbol{\Sigma}_2 = \widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{B}}'$. Note that $\mathrm{tr}\,\boldsymbol{\Xi} = \mathrm{tr}\,\boldsymbol{\Delta}$.

Proposition 1. *For general constants c_i's, the risk function $R(\widehat{\boldsymbol{\Delta}}_1, \omega)$ of $\widehat{\boldsymbol{\Delta}}_1$ is expressed as*

$$
\begin{aligned}
&\left\{R(\widehat{\boldsymbol{\Delta}}_1, \omega) - m\mathrm{tr}\,\boldsymbol{\Delta}\right\}(n+p+1)/n \\
&= \sum_{i=1}^{p} \xi_{ii} \left\{ \sum_{j=i}^{p} \frac{c_j^2}{\prod_{k=i}^{j}(m-k-1)} - 2c_i \right\} && (3.5)\\
&= \sum_{i=1}^{p} \tau_i \left\{ c_i^2 - 2(m-i-1)c_i + 2c_{i+1} \right\}, && (3.6)
\end{aligned}
$$

where $\tau_i = (\tau_{i-1} + \xi_{ii})/(m-i-1)$, $\tau_0 = 0$ and $c_{p+1} = 0$. Also an unbiased estimator of the risk $R(\widehat{\boldsymbol{\Delta}}_1, \omega)$ is given by

$$
\begin{aligned}
\widehat{R}(\widehat{\boldsymbol{\Delta}}_1, \omega) &= m(m-p-1)n^{-1}\mathrm{tr}\,\boldsymbol{S}\boldsymbol{W}^{-1} \\
&+ (n+p+1)^{-1} \sum_{i=1}^{p} F_{ii} \left\{ c_i^2 - 2(m-i-1)c_i + 2c_{i+1} \right\}, \qquad (3.7)
\end{aligned}
$$

where F_{ii} is the (i,i) element of $\boldsymbol{T}^{-1}\boldsymbol{S}\boldsymbol{T}'^{-1}$.

This expression means that we could not choose any optimal c_i's except for c_1. The optimal for c_1 is $c_1 = m-2$. It is seen, however, that one reasonable choice for c_i is $c_i = m-i-1$, given in (3.3).

We now show that the James-Stein type estimator $\widehat{\boldsymbol{\Delta}}_1^{JS}$ for $c_i = m-i-1$ dominates $\widehat{\boldsymbol{\Delta}}_0$. From (3.7), it is sufficient to check that

$$
c_i^2 - 2(m-i-1)c_i + 2c_{i+1} \le c_p^2 - 2(m-i-1)c_p + 2c_p, \quad i = 1, \ldots, p-1 \qquad (3.8)
$$

for $c_i = m-i-1$, since the risk of $\widehat{\boldsymbol{\Delta}}_0$ can be given by Proposition 1 with putting $c_1 = \cdots = c_p = m-p-1$. The inequality (3.8) is equivalent to

$$
(c_i - c_p)(c_p - c_i + 2) - 2 = -(p-i)(p-i-2) - 2 \le 0, \quad i = 1, \ldots, p-1,
$$

which is easily checked, and we get

Corollary 1. *The James-Stein type estimator $\widehat{\boldsymbol{\Delta}}_1^{JS}$ with $c_i = m-i-1$ dominates $\widehat{\boldsymbol{\Delta}}_0$ relative to the risk (2.6).*

Remark 1. Another choice for c_i's that may be suggested from an analogy of the estimation of a covariance matrix (James and Stein(1961)) are given by

$$c_i = m + p - 2i - 1, \quad \text{for} \quad i = 1, \dots, p.$$

In the risk function of $\widehat{\boldsymbol{\Delta}}_1$ for $c_i = m+p-2i-1$, the coefficient of $\xi_{p-1,p-1}$ in the expression (3.5) is

$$\frac{(m-p+1)^2}{m-p} + \frac{(m-p-1)^2}{(m-p)(m-p-1)} - 2(m-p+1) + m = p + 1, \tag{3.9}$$

while the corresponding coefficient in $\widehat{\boldsymbol{\Delta}}_1^{JS}$ for $c_i = m - i - 1$ in (3.5) is

$$\frac{(m-p)^2}{m-p} + \frac{(m-p-1)^2}{(m-p)(m-p-1)} - 2(m-p) + m = p + 1 - \frac{1}{m-p},$$

which is smaller than (3.9). This means that $\widehat{\boldsymbol{\Delta}}_1$ for $c_i = m + p - 2i - 1$ is not better than $\widehat{\boldsymbol{\Delta}}_1^{JS}$. Moreover, the coefficient of $\xi_{p-2,p-2}$ in the risk expression (3.5) of $\widehat{\boldsymbol{\Delta}}_1$ for $c_i = m + p - 2i - 1$ is

$$p + 1 + \frac{5 - (m-p)}{m - p + 1},$$

which is greater than $p + 1$ if $m - p < 5$. In this case, the risk goes beyond the risk of $\widehat{\boldsymbol{\Delta}}_0$. Hence these choices are not recommended.

3.2 Dominance property of $\hat{\Delta}_2^{JS}$

We next evaluate the risk function of $\widehat{\boldsymbol{\Delta}}_2^{JS}$ and verify the dominance result of $\widehat{\boldsymbol{\Delta}}_2^{JS}$ over $\widehat{\boldsymbol{\Delta}}_0$. Let λ_{ii} be the (i,i) element of $\boldsymbol{\Lambda}$ where $\boldsymbol{\Lambda} = \widetilde{\boldsymbol{A}}'\boldsymbol{\Sigma}_2^{-1}\widetilde{\boldsymbol{A}}$ for $\widetilde{\boldsymbol{A}} \in G_L^+$ such that $\boldsymbol{\Sigma} = \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{A}}'$. Note that $\mathrm{tr}\,\boldsymbol{\Lambda} = \mathrm{tr}\,\boldsymbol{\Delta}$.

Proposition 2. *For general constants d_i's, the risk function $R(\widehat{\boldsymbol{\Delta}}_2, \omega)$ is expressed as*

$$\begin{aligned}&\left\{R(\widehat{\boldsymbol{\Delta}}_2, \omega) - m\mathrm{tr}\,\boldsymbol{\Delta}\right\}(m-p-1)^{-1} \\ &= \sum_{i=1}^{p} \tau_i^* \left\{(n+p+3-2i)d_i^2 - 2\left(1 - \sum_{j=1}^{i-1} d_j\right) d_i\right\},\end{aligned} \tag{3.10}$$

where $\tau_i^* = \lambda_{ii}(n - i + 1) + \sum_{j=i+1}^{p} \lambda_{jj}$. Also an unbiased estimator of the risk $R(\widehat{\boldsymbol{\Delta}}_2, \omega)$ is given by

$$\begin{aligned}\hat{R}(\widehat{\boldsymbol{\Delta}}_2, \omega) &= m(m-p-1)n^{-1}\mathrm{tr}\,\boldsymbol{S}\boldsymbol{W}^{-1} \\ &+ \sum_{i=1}^{p} G_{ii} \left\{(n+p+3-2i)d_i^2 - 2\left(1 - \sum_{j=1}^{i-1} d_j\right) d_i\right\},\end{aligned} \tag{3.11}$$

where G_{ii} is the (i,i) element of $\boldsymbol{U}'\boldsymbol{W}^{-1}\boldsymbol{U}$.

From Proposition 2, it is seen that the optimal d_i is not obtainable except for d_1, the best d_1 being $d_1 = (n+p+1)^{-1}$. One reasonable choice for d_i is

$$\begin{aligned} d_i &= \frac{1}{n+p+3-2i}\left(1-\sum_{j=1}^{i-1} d_j\right) \\ &= \frac{1}{n+p+3-2i}\left(1-\frac{1}{n+p+1}\right)\left(1-\frac{1}{n+p-1}\right)\cdots\left(1-\frac{1}{n+p+5-2i}\right). \end{aligned}$$

As indicated in (3.2), the estimator $\widehat{\boldsymbol{\Delta}}_2$ with these d_i's is denoted by $\widehat{\boldsymbol{\Delta}}_2^{JS}$. These d_i's have the order relations

$$\begin{aligned} d_i &= d_{i-1} \times \frac{n+p+4-2i}{n+p+3-2i} \\ &\geq d_{i-1} \\ &\geq \cdots \geq d_1 = (n+p+1)^{-1}. \end{aligned}$$

To establish the dominance result of $\widehat{\boldsymbol{\Delta}}_2^{JS}$ over $\widehat{\boldsymbol{\Delta}}_0$, the following inequality is essential:

$$\frac{(1-\sum_{j=1}^{i-1} d_j)^2}{n+p+3-2i} \geq \frac{1}{n+p+1}, \quad i = 2, \ldots, p, \tag{3.12}$$

which is equivalent to

$$\frac{(n+p)^2}{(n+p+1)^2}\frac{(n+p-2)^2}{(n+p-1)^2}\cdots\frac{(n+p+4-2i)^2}{(n+p+5-2i)^2}\frac{n+p+3-2i}{n+p+1},$$

or

$$\prod_{j=2}^{i}\left\{\frac{(n+p+4-2j)^2}{(n+p+5-2j)(n+p+3-2j)}\right\} \geq 1.$$

Here it is easily checked that for $j = 2, \ldots, i$,

$$(n+p+4-2j)^2 \geq (n+p+5-2j)(n+p+3-2j),$$

so that we get the inequality (3.12).

From Proposition 2, the risk of $\widehat{\boldsymbol{\Delta}}_2^{JS}$ is evaluated as

$$\begin{aligned} &\left\{R(\widehat{\boldsymbol{\Delta}}_2^{JS}, \omega) - m \operatorname{tr} \boldsymbol{\Delta}\right\}(m-p-1)^{-1} \\ &= -(\lambda_{11} n + \operatorname{tr} \boldsymbol{\Lambda}_{22})\, d_1 \\ &\quad - (\lambda_{22}(n-1) + \operatorname{tr} \boldsymbol{\Lambda}_{33})(1-d_1)^2/(n+p-1) - \cdots \end{aligned}$$

$$
\begin{aligned}
& -\left(\lambda_{p-1,p-1}(n-p+2)+\lambda_{pp}\right)\left(1-\sum_{j=1}^{p-2} d_j\right)^2 /(n+p+3-2(p-1)) \\
& -\lambda_{pp}(n-p+1)\left(1-\sum_{j=1}^{p-1} d_j\right)^2 /(n+p+3-2p) \\
\leq\; & -\left(\lambda_{11}n+\operatorname{tr}\boldsymbol{\Lambda}_{22}\right) d_1-\left(\lambda_{22}(n-1)+\operatorname{tr}\boldsymbol{\Lambda}_{33}\right) d_1-\cdots \\
& -\left(\lambda_{p-1,p-1}(n-p+2)+\lambda_{pp}\right) d_1 \\
& -\lambda_{pp}(n-p+1) d_1,
\end{aligned} \tag{3.13}
$$

where we have used (3.12) for the inequality in (3.13). On the other hand, the risk of $\widehat{\boldsymbol{\Delta}}_0$ corresponds to the case $d_1=\cdots=d_p=(n+p+1)^{-1}$ in Proposition 2, and then we observe that

$$
\begin{aligned}
& (n+p+3-2i)d_i^2-2\left(1-\sum_{j+1}^{i-1} d_j\right) d_i \\
=\; & \left\{(n+p+3-2i)d_1-2\left(1-(i-1)d_1\right)\right\} d_1 \\
=\; & -d_1.
\end{aligned}
$$

This implies that $\{R(\widehat{\boldsymbol{\Delta}}_0,\omega)-m\operatorname{tr}\boldsymbol{\Lambda}\}(m-p-1)^{-1}$ is given by the r.h.s. in the inequality (3.13), and we get

Corollary 2. *The James-Stein type estimator* $\widehat{\boldsymbol{\Delta}}_2^{JS}$ *dominates* $\widehat{\boldsymbol{\Delta}}_0$ *relative to the risk (2.6).*

3.3 Proofs

We here prove Propositions 1 and 2. For the convenience's sake, we define $R^*(\widehat{\boldsymbol{\Delta}},\omega)$ by

$$
\begin{aligned}
R^*(\widehat{\boldsymbol{\Delta}},\omega) &= R(\widehat{\boldsymbol{\Delta}},\omega)-m\operatorname{tr}\boldsymbol{\Delta} \\
&= E_\omega\left[\operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}_2^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_2^{-1}\boldsymbol{W}\right)-2\operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_2^{-1}\boldsymbol{W}\boldsymbol{\Sigma}_2^{-1}\right)\right].
\end{aligned} \tag{3.14}
$$

Proof of Proposition 1. Since $\widehat{\boldsymbol{\Sigma}}=(n+p+1)^{-1}\boldsymbol{S}$, we first observe that

$$
\begin{aligned}
R^*(\widehat{\boldsymbol{\Delta}}_1,\omega) &= \frac{n}{n+p+1}E_{\Sigma_2}\left[\operatorname{tr}\widehat{\boldsymbol{\Sigma}}_2^{-1}\boldsymbol{W}\widehat{\boldsymbol{\Sigma}}_2^{-1}\boldsymbol{\Sigma}-2\operatorname{tr}\widehat{\boldsymbol{\Sigma}}_2^{-1}\boldsymbol{W}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}\right] \\
&= \frac{n}{n+p+1}E_{\Sigma_2=I}\left[\operatorname{tr}\widehat{\boldsymbol{\Sigma}}_2^{-1}\boldsymbol{W}\widehat{\boldsymbol{\Sigma}}_2^{-1}\boldsymbol{\Xi}-2\operatorname{tr}\widehat{\boldsymbol{\Sigma}}_2^{-1}\boldsymbol{W}\boldsymbol{\Xi}\right].
\end{aligned} \tag{3.15}
$$

Let

$$
\boldsymbol{T}=\begin{pmatrix} t_{11} & \mathbf{0}' \\ \boldsymbol{t}_{21} & \boldsymbol{T}_{22}\end{pmatrix},\quad \boldsymbol{C}=\begin{pmatrix} c_1 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{C}_2\end{pmatrix} \quad\text{and}\quad \boldsymbol{\Xi}=\begin{pmatrix} \xi_{11} & \boldsymbol{\Xi}_{12} \\ \boldsymbol{\Xi}_{21} & \boldsymbol{\Xi}_{22}\end{pmatrix}
$$

for scalar t_{11}, c_1 and ξ_{11}. Note that t_{11}, $\boldsymbol{t}_{21}$ and $\boldsymbol{T}_{22}$ are mutually independent and $t_{11}^2 \sim \chi_m^2$ and $\boldsymbol{t}_{21} \sim \mathcal{N}_{p-1}(\boldsymbol{0}, \boldsymbol{I}_{p-1})$. Then

$$\begin{aligned}
&E_I\left[\operatorname{tr}\widehat{\boldsymbol{\Sigma}}_2^{-1}\boldsymbol{W}\widehat{\boldsymbol{\Sigma}}_2^{-1}\boldsymbol{\Xi}\right] \\
&= E_I\left[\operatorname{tr}\boldsymbol{T}'^{-1}\boldsymbol{C}^2\boldsymbol{T}^{-1}\boldsymbol{\Xi}\right] \\
&= E_I\left[\operatorname{tr}\begin{pmatrix} t_{11}^{-1} & -t_{11}^{-1}\boldsymbol{t}_{21}'\boldsymbol{T}_{22}'^{-1} \\ \boldsymbol{0} & \boldsymbol{T}_{22}'^{-1}\end{pmatrix}\begin{pmatrix} c_1 & \boldsymbol{0}' \\ \boldsymbol{0} & \boldsymbol{C}_2\end{pmatrix}\begin{pmatrix} t_{11}^{-1} & \boldsymbol{0}' \\ -t_{11}^{-1}\boldsymbol{T}_{22}^{-1}\boldsymbol{t}_{21} & \boldsymbol{T}_{22}^{-1}\end{pmatrix}\begin{pmatrix} \xi_{11} & \boldsymbol{\Xi}_{12} \\ \boldsymbol{\Xi}_{21} & \boldsymbol{\Xi}_{22}\end{pmatrix}\right] \\
&= E_I\left[\xi_{11}t_{11}^{-2}\left(c_1^2+\boldsymbol{t}_{21}'\boldsymbol{T}_{22}'^{-1}\boldsymbol{C}_2^2\boldsymbol{T}_{22}^{-1}\boldsymbol{t}_{21}\right)+\operatorname{tr}\boldsymbol{T}_{22}'^{-1}\boldsymbol{C}_2^2\boldsymbol{T}_{22}^{-1}\boldsymbol{\Xi}_{22}\right] \\
&= E_I\left[\frac{\xi_{11}}{m-2}\left(c_1^2+\boldsymbol{t}_{21}'\boldsymbol{T}_{22}'^{-1}\boldsymbol{C}_2^2\boldsymbol{T}_{22}^{-1}\boldsymbol{t}_{21}\right)+\operatorname{tr}\boldsymbol{T}_{22}'^{-1}\boldsymbol{C}_2^2\boldsymbol{T}_{22}^{-1}\boldsymbol{\Xi}_{22}\right].
\end{aligned}$$

This argument can be repeated to have that

$$\begin{aligned}
&E_I\left[\operatorname{tr}\boldsymbol{T}'^{-1}\boldsymbol{C}^2\boldsymbol{T}^{-1}\boldsymbol{\Xi}\right] \\
&= E_I\left[\frac{\xi_{11}}{m-2}\left\{c_1^2+\frac{1}{m-3}\left(c_2^2+\operatorname{tr}\boldsymbol{T}_{33}'^{-1}\boldsymbol{C}_3^2\boldsymbol{T}_{33}^{-1}\right)\right\}\right. \\
&\qquad \left.+\frac{\xi_{22}}{m-3}\left(c_2^2+\operatorname{tr}\boldsymbol{T}_{33}'^{-1}\boldsymbol{C}_3^2\boldsymbol{T}_{33}^{-1}\right)+\operatorname{tr}\boldsymbol{T}_{33}'^{-1}\boldsymbol{C}_3^2\boldsymbol{T}_{33}^{-1}\boldsymbol{\Xi}_{33}\right] \\
&= \cdots \quad \cdots \\
&= \sum_{i=1}^p \xi_{ii}\left\{\prod_{j=i}^p \frac{c_j^2}{\prod_{k=i}^j (m-k-1)}\right\},
\end{aligned} \tag{3.16}$$

where $\boldsymbol{T}_{33}$, $\boldsymbol{C}_3$ and $\boldsymbol{\Xi}_{33}$ denote $(p-2)\times(p-2)$ right lower corners of $\boldsymbol{T}$, $\boldsymbol{C}$ and $\boldsymbol{\Lambda}$. Similarly, we observe that

$$\begin{aligned}
&E_I\left[\operatorname{tr}\boldsymbol{T}\boldsymbol{C}\boldsymbol{T}^{-1}\boldsymbol{\Xi}\right] \\
&= E_I\left[\operatorname{tr}\begin{pmatrix} t_{11} & \boldsymbol{0}' \\ \boldsymbol{t}_{21} & \boldsymbol{T}_{22}\end{pmatrix}\begin{pmatrix} c_1 & \boldsymbol{0}' \\ \boldsymbol{0} & \boldsymbol{C}_2\end{pmatrix}\begin{pmatrix} t_{11}^{-1} & \boldsymbol{0}' \\ -t_{11}^{-1}\boldsymbol{T}_{22}^{-1}\boldsymbol{t}_{21} & \boldsymbol{T}_{22}^{-1}\end{pmatrix}\begin{pmatrix} \xi_{11} & \boldsymbol{\Xi}_{12} \\ \boldsymbol{\Xi}_{21} & \boldsymbol{\Xi}_{22}\end{pmatrix}\right] \\
&= E_I\left[c_1\xi_{11}+t_{11}^{-1}\left(c_1\boldsymbol{\Xi}_{12}\boldsymbol{t}_{21}-\boldsymbol{\Xi}_{12}\boldsymbol{T}_{22}\boldsymbol{C}_2\boldsymbol{T}_{22}^{-1}\boldsymbol{t}_{21}\right)+\operatorname{tr}\boldsymbol{T}_{22}\boldsymbol{C}_2\boldsymbol{T}_{22}^{-1}\boldsymbol{\Xi}_{22}\right] \\
&= c_1\xi_{11}+E_I\left[\operatorname{tr}\boldsymbol{T}_{22}\boldsymbol{C}_2\boldsymbol{T}_{22}^{-1}\boldsymbol{\Lambda}_{22}\right] \\
&= \sum_{i=1}^p c_i\xi_{ii}.
\end{aligned} \tag{3.17}$$

Combining (3.16) and (3.17) gives the risk expression (3.5) of Proposition 1.

The expression (3.6) can be easily derived by letting $\tau_i = (\tau_{i-1}+\xi_{ii})/(m-i-1)$. For the unbiased estimator (3.7), note that

$$E_\Xi[\boldsymbol{T}^{-1}\boldsymbol{S}\boldsymbol{T}'^{-1}] = nE_I[\boldsymbol{T}^{-1}\boldsymbol{\Xi}\boldsymbol{T}'^{-1}]$$

$$= n \begin{pmatrix} \frac{\xi_{11}}{m-2} & * \\ * & \frac{\xi_{11}}{m-2}\boldsymbol{T}_{22}^{-1}\boldsymbol{T}_{22}^{\prime -1} + \boldsymbol{T}_{22}^{-1}\boldsymbol{\Xi}_{22}\boldsymbol{T}_{22}^{\prime -1} \end{pmatrix}$$

$$= n \begin{pmatrix} \frac{\xi_{11}}{m-2} & & * \\ & \left(\frac{\xi_{11}}{m-2} + \xi_{22}\right)\frac{1}{m-3} & \\ * & & \left(\frac{\xi_{11}}{m-2} + \xi_{22}\right)\frac{1}{m-3}\boldsymbol{T}_{33}^{-1}\boldsymbol{T}_{33}^{\prime -1} + \boldsymbol{T}_{33}^{-1}\boldsymbol{\Xi}_{33}\boldsymbol{T}_{33}^{\prime -1} \end{pmatrix}.$$

Repeating these arguments, we see that for $F_{ii} = (\boldsymbol{T}^{-1}\boldsymbol{S}\boldsymbol{T}^{\prime -1})_{ii}$,

$$E[F_{ii}] = n\left\{E[F_{i-1,i-1}] + \xi_{ii}\right\}/(m-i-1) = n\tau_i,$$

which gives the unbiased estimator (3.7), and the proof is complete. □□

Proof of Proposition 2. Noting that $\widehat{\boldsymbol{\Sigma}}_2^{-1} = (m-p-1)\boldsymbol{W}^{-1}$, from (3.14), we see that

$$\begin{aligned} R^*(\widehat{\boldsymbol{\Delta}}_2, \omega) &= (m-p-1)E_\Sigma\left[\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}_2^{-1} - 2\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}_2^{-1}\right] \\ &= (m-p-1)E_{\Sigma=I}\left[\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}^2\boldsymbol{\Lambda} - 2\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Lambda}\right]. \end{aligned} \tag{3.18}$$

Let

$$\boldsymbol{U} = \begin{pmatrix} u_{11} & \boldsymbol{0}' \\ \boldsymbol{u}_{21} & \boldsymbol{U}_{22} \end{pmatrix}$$

for scalar u_{11}. The same arguments as in Proposition 1 gives that

$$\begin{aligned} E_I\left[\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}^2\boldsymbol{\Lambda}\right] &= E_I\left[\mathrm{tr}\,\boldsymbol{UDU}'\boldsymbol{UDU}'\boldsymbol{\Lambda}\right] \\ &= E_I\Big[\lambda_{11}\left(d_1^2u_{11}^4 + d_1^2u_{11}^2\boldsymbol{u}_{21}'\boldsymbol{u}_{21}\right) \\ &\quad +\mathrm{tr}\left\{d_1^2u_{11}^2\boldsymbol{u}_{21}\boldsymbol{u}_{21}' + (d_1\boldsymbol{u}_{21}\boldsymbol{u}_{21}' + \boldsymbol{U}_{22}\boldsymbol{D}_2\boldsymbol{U}_{22}')^2\right\}\Big] \\ &= E_I\Big[\lambda_{11}(n+p+1)nd_1^2 + (n+p+1)d_1^2\mathrm{tr}\,\boldsymbol{\Lambda}_{22} \\ &\quad +2d_1\mathrm{tr}\,\boldsymbol{U}_{22}\boldsymbol{D}_2\boldsymbol{U}_{22}'\boldsymbol{\Lambda}_{22} + \mathrm{tr}\,(\boldsymbol{U}_{22}\boldsymbol{D}_2\boldsymbol{U}_{22}')^2\boldsymbol{\Lambda}_{22}\Big], \end{aligned} \tag{3.19}$$

where $\boldsymbol{U}_{22}$, $\boldsymbol{D}_{22}$ and $\boldsymbol{\Lambda}_{22}$ denote $(p-1)\times(p-1)$ right lower corners of $\boldsymbol{U}$, $\boldsymbol{D}$ and $\boldsymbol{\Lambda}$. On the other hand, we have that

$$\begin{aligned} E_I\left[\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Lambda}\right] &= E_I\left[\mathrm{tr}\,\boldsymbol{UDU}'\boldsymbol{\Lambda}\right] \\ &= E_I\left[d_1u_{11}^2\lambda_{11} + \mathrm{tr}\,(d_1\boldsymbol{u}_{21}\boldsymbol{u}_{21}' + \boldsymbol{U}_{22}\boldsymbol{D}_2\boldsymbol{U}_{22}')\boldsymbol{\Lambda}_{22}\right] \\ &= d_1n\lambda_{11} + d_1\mathrm{tr}\,\boldsymbol{\Lambda}_{22} + E_I\left[\mathrm{tr}\,\boldsymbol{U}_{22}\boldsymbol{D}_2\boldsymbol{U}_{22}'\boldsymbol{\Lambda}_{22}\right]. \end{aligned} \tag{3.20}$$

Combining (3.18), (3.19) and (3.20), we get that

$$\begin{aligned} R^*(\widehat{\boldsymbol{\Delta}}_2, \omega) &= (m-p-1)\Big\{(\lambda_{11}n + \mathrm{tr}\,\boldsymbol{\Lambda}_{22})\left[(n+p+1)d_1^2 - 2d_1\right] \\ &\quad +E_I\left[\mathrm{tr}\,(\boldsymbol{U}_{22}\boldsymbol{D}_2\boldsymbol{U}_{22}')^2\boldsymbol{\Lambda}_{22} - 2(1-d_1)\mathrm{tr}\,\boldsymbol{U}_{22}\boldsymbol{D}_2\boldsymbol{U}_{22}'\boldsymbol{\Lambda}_{22}\right]\Big\}. \end{aligned} \tag{3.21}$$

Next, the same argument can be applied to the second term in the bracket $\{\cdot\}$ of (3.21), and we have

$$
\begin{aligned}
&\text{(the second term)} \\
= \;& \lambda_{22}(n+p-1)(n-1)d_2^2 + d_2^2(n+p-1)\mathrm{tr}\,\boldsymbol{\Lambda}_{33} \\
&+E_I\left[2d_2\mathrm{tr}\,\boldsymbol{U}_{33}\boldsymbol{D}_3\boldsymbol{U}_{33}'\boldsymbol{\Lambda}_{33} + \mathrm{tr}\,\left(\boldsymbol{U}_{33}\boldsymbol{D}_3\boldsymbol{U}_{33}'\right)^2\boldsymbol{\Lambda}_{33}\right] \\
&-2(1-d_1)\left\{(n-1)d_2\lambda_{22} + d_2\mathrm{tr}\,\boldsymbol{\Lambda}_{33} + E_I\left[\mathrm{tr}\,\boldsymbol{U}_{33}\boldsymbol{D}_3\boldsymbol{U}_{33}'\boldsymbol{\Lambda}_{33}\right]\right\} \\
= \;& (\lambda_{22}(n-1) + \mathrm{tr}\,\boldsymbol{\Lambda}_{33})\left\{(n+p-1)d_2^2 - 2(1-d_1)d_2\right\} \\
&+E_I\left[\mathrm{tr}\,\left(\boldsymbol{U}_{33}\boldsymbol{D}_3\boldsymbol{U}_{33}'\right)^2\boldsymbol{\Lambda}_{33} - 2(1-d_1-d_2)\mathrm{tr}\,\boldsymbol{U}_{33}\boldsymbol{D}_3\boldsymbol{U}_{33}'\boldsymbol{\Lambda}_{33}\right],
\end{aligned}
$$

where $\boldsymbol{U}_{33}$, $\boldsymbol{D}_{33}$ and $\boldsymbol{\Lambda}_{33}$ denote $(p-2)\times(p-2)$ right lower corners of $\boldsymbol{U}$, $\boldsymbol{D}$ and $\boldsymbol{\Lambda}$. Repeating these arguments, we can see that the risk function of $\widehat{\boldsymbol{\Delta}}_2$ is expressed by (3.10) of Proposition 2.

For the unbiased estimator (3.11), note that

$$
\begin{aligned}
&E_\Lambda[\boldsymbol{U}'\boldsymbol{W}^{-1}\boldsymbol{U}] = (m-p-1)^{-1}E_I[\boldsymbol{U}'\boldsymbol{\Lambda}\boldsymbol{U}] \\
=\;& \frac{1}{m-p-1}\begin{pmatrix} \lambda_{11}n + \mathrm{tr}\,\boldsymbol{\Lambda}_{22} & * \\ * & \boldsymbol{U}_2'\boldsymbol{\Lambda}_{22}\boldsymbol{U}_{22} \end{pmatrix} \\
=\;& \frac{1}{m-p-1}\begin{pmatrix} \lambda_{11}n + \mathrm{tr}\,\boldsymbol{\Lambda}_{22} & & * \\ & \lambda_{22}(n-1) + \mathrm{tr}\,\boldsymbol{\Lambda}_{33} & \\ * & & \boldsymbol{U}_{33}'\boldsymbol{\Lambda}_{33}\boldsymbol{U}_{33} \end{pmatrix}.
\end{aligned}
$$

Hence, we see that for $G_{ii} = (\boldsymbol{U}'\boldsymbol{W}^{-1}\boldsymbol{U})_{ii}$,

$$
E[G_{ii}] = (m-p-1)^{-1}\left\{\lambda_{ii}(n-i+1) + \mathrm{tr}\,\boldsymbol{\Lambda}_{i+1,i+1}\right\} = (m-p-1)^{-1}\tau_i^*,
$$

which gives the unbiased estimator (3.11), and the proof is complete. □□

4 Stein Type Orthogonally Equivariant Estimators

The James-Stein type estimators $\widehat{\boldsymbol{\Delta}}_1^{JS}$ and $\widehat{\boldsymbol{\Delta}}_2^{JS}$ treated in Section 3 has a shortcoming in that they depend on the coordinate systems. To overcome this undesirable property, we need to consider orthogonally equivariant estimators.

In contrast with the results in estimation of covariance matrix, we could not provide any orthogonally equivariant estimators superior to the James-Stein type estimators in terms of risk. The reason is that the risks of $\widehat{\boldsymbol{\Delta}}_1^{JS}$ and $\widehat{\boldsymbol{\Delta}}_2^{JS}$ are not constants, but depend on the coordinate systems as seen from Propositions 1 and 2. Instead of this, we here show that the Stein type orthogonally equivariant estimators associated with $\widehat{\boldsymbol{\Delta}}_1^{JS}$ and $\widehat{\boldsymbol{\Delta}}_2^{JS}$ have uniformly smaller risks than $\widehat{\boldsymbol{\Delta}}_0$.

4.1 Stein type orthogonally equivariant estimator associated with $\widehat{\Delta}_1^{JS}$

Let $O(p)$ be the group of $p \times p$ orthogonal matrices. Let $x_1, \ldots, x_p$, $x_1 \geq \ldots \geq x_p$, be eigenvalues of $\boldsymbol{W}$ such that $\boldsymbol{W} = \boldsymbol{HXH}'$ for $\boldsymbol{H} \in O(p)$ and $\boldsymbol{X} = \mathrm{diag}\,(x_1, \ldots, x_p)$. The Stein type estimator associated with $\widehat{\boldsymbol{\Delta}}_1^{JS}$ is given by

$$\widehat{\boldsymbol{\Delta}}_1^{S} = \frac{1}{n+p+1} \boldsymbol{SH} \mathrm{diag}\left(\frac{c_1}{x_1}, \ldots, \frac{c_p}{x_p}\right) \boldsymbol{H}', \tag{4.1}$$

for $c_i = m - i - 1$, and our purpose is to establish the dominance of $\widehat{\boldsymbol{\Delta}}_1^{S}$ over $\widehat{\boldsymbol{\Delta}}_0$.

More generally, we consider estimators of the form

$$\begin{aligned} \widehat{\boldsymbol{\Delta}}_1(\boldsymbol{\Psi}) &= (n+p+1)^{-1} \boldsymbol{SH\Psi}(\boldsymbol{x})\boldsymbol{H}', \\ \boldsymbol{\Psi}(\boldsymbol{x}) &= \mathrm{diag}\,(\psi_1(\boldsymbol{x}), \ldots, \psi(\boldsymbol{x})) \end{aligned} \tag{4.2}$$

for positive functions $\psi_i(\boldsymbol{x})$'s of $\boldsymbol{x} = (x_1, \ldots, x_p)$.

Proposition 3. *The risk function of $\widehat{\boldsymbol{\Delta}}_1(\boldsymbol{\Psi})$ is given by*

$$R(\widehat{\boldsymbol{\Delta}}_1(\boldsymbol{\Psi}), \omega) = (n+p+1)^{-1} E\,[\mathrm{tr}\, \boldsymbol{SH\Psi}^*(\boldsymbol{x})\boldsymbol{H}'] + m\mathrm{tr}\, \boldsymbol{\Lambda}, \tag{4.3}$$

where $\boldsymbol{\Psi}^*(\boldsymbol{x}) = \mathrm{diag}\,(\psi_1^*(\boldsymbol{x}), \ldots, \psi_p^*(\boldsymbol{x}))$ for

$$\psi_i^*(\boldsymbol{x}) = x_i\psi_i^2 - 2(m-p-1)\psi_i - 2\sum_{j \neq i} \frac{x_i\psi_i - x_j\psi_j}{x_i - x_j} - 4\frac{\partial(x_i\psi)}{\partial x_i}. \tag{4.4}$$

From Proposition 3, the $\psi_i^*(\boldsymbol{x})$ for the Stein type estimator $\widehat{\boldsymbol{\Delta}}_1^{S}$ is

$$\begin{aligned} & \frac{c_i^2}{x_i} - 2(m-p-1)\frac{c_i}{x_i} - 2\sum_{j \neq i} \frac{c_i - c_j}{x_i - x_j} \\ = & -\frac{(m-i-1)(m+i-2p-1)}{x_i} - 2\sum_{j \neq i} \frac{j-i}{x_i - x_j} \end{aligned}$$

while for the estimator $\widehat{\boldsymbol{\Delta}}_0$, the corresponding quantity is $-(m-p-1)^2/x_i$. Hence the risk difference can be written as

$$\begin{aligned} & \left\{ R(\widehat{\boldsymbol{\Delta}}_0, \omega) - R(\widehat{\boldsymbol{\Delta}}_1^{S}, \omega) \right\} (n+p+1) \\ = & \; E\left[\sum_{i=1}^{p} a_{ii} \left\{ -\frac{(p-i)^2}{x_i} + 2\sum_{j \neq i} \frac{j-i}{x_i - x_j} \right\} \right], \end{aligned} \tag{4.5}$$

where $a_{ii} = (\boldsymbol{H}'\boldsymbol{S}\boldsymbol{H})_{ii}$. Here note that

$$
\begin{aligned}
\sum_{j\neq i} \frac{j-i}{x_i - x_j} &= \sum_{j=1}^{i-1} \frac{i-j}{x_j - x_i} + \sum_{j=i+1}^{p} \frac{j-i}{x_i - x_j} \\
&\geq \frac{1}{x_i - x_{i+1}} + \frac{2}{x_i - x_{i+2}} + \cdots + \frac{p-i}{x_i - x_p} \\
&\geq \frac{\sum_{j=1}^{p-i} j}{x_i - x_p} = \frac{(p-i)(p-i+1)}{2(x_i - x_p)},
\end{aligned}
$$

which is used to show that

$$
\begin{aligned}
-\frac{(p-i)^2}{x_i} + 2\sum_{j\neq i} \frac{j-i}{x_i - x_j} &\geq -\frac{(p-i)^2}{x_i} + \frac{(p-i)(p-i+1)}{x_i - x_p} \\
&= \frac{p-i}{x_i(x_i - x_p)} \left\{x_i + (p-i)x_p\right\},
\end{aligned}
$$

which is nonnegative, and we get

Corollary 3. *The Stein type orthogonally equivariant estimator $\widehat{\boldsymbol{\Delta}}_1^S$ dominates $\widehat{\boldsymbol{\Delta}}_0$ relative to the risk (2.6).*

4.2 Stein type orthogonally equivariant estimator associated with $\widehat{\Delta}_2^{JS}$

Let $\ell_1, \ldots, \ell_p$, $\ell_1 \geq \cdots \geq \ell_p$, be eigenvalues of $\boldsymbol{S}$ such that $\boldsymbol{S} = \boldsymbol{K}\boldsymbol{L}\boldsymbol{K}'$ for $\boldsymbol{K} \in O(p)$ and $\boldsymbol{L} = \mathrm{diag}\,(\ell_1, \ldots, \ell_p)$. The Stein type estimator associated with $\widehat{\boldsymbol{\Delta}}_2^{JS}$ is given by

$$
\widehat{\boldsymbol{\Delta}}_2^S = (m-p-1)\boldsymbol{K}\mathrm{diag}\,(d_1\ell_1, \ldots, d_p\ell_p)\,\boldsymbol{K}'\boldsymbol{W}^{-1}, \tag{4.6}
$$

for d_i's defined by (3.4), and we shall verify the superiority of $\widehat{\boldsymbol{\Delta}}_2^S$ over $\widehat{\boldsymbol{\Delta}}_0$.

We first provide the risk function for estimators of the general form

$$
\begin{aligned}
\widehat{\boldsymbol{\Delta}}_2(\boldsymbol{\Psi}) &= (m-p-1)\boldsymbol{K}\boldsymbol{\Psi}(\boldsymbol{\ell})\boldsymbol{K}'\boldsymbol{W}^{-1}, \\
\boldsymbol{\Psi}(\boldsymbol{\ell}) &= \mathrm{diag}\,(\psi_1(\boldsymbol{\ell}), \ldots, \psi(\boldsymbol{\ell})),
\end{aligned} \tag{4.7}
$$

for positive functions $\psi_i(\boldsymbol{\ell})$'s of $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_p)$.

Proposition 4. *The risk function of $\widehat{\boldsymbol{\Delta}}_2(\boldsymbol{\Psi})$ is given by*

$$
R(\widehat{\boldsymbol{\Delta}}_2(\boldsymbol{\Psi}), \omega) = (m-p-1)^2 E_\omega \left[\mathrm{tr}\,\boldsymbol{K}\boldsymbol{\Psi}^*(\boldsymbol{\ell})\boldsymbol{K}'\boldsymbol{W}^{-1}\right] + m\mathrm{tr}\,\boldsymbol{\Lambda}, \tag{4.8}
$$

where $\boldsymbol{\Psi}^*(\boldsymbol{\ell}) = \mathrm{diag}\,(\psi_1^*(\boldsymbol{\ell}), \ldots, \psi_p^*(\boldsymbol{\ell}))$ for

$$
\psi_i^*(\boldsymbol{\ell}) = (n-p-1)\frac{\psi_i^2}{\ell_i} - 2\psi_i + 2\sum_{j\neq i} \frac{\psi_i(\psi_i - \psi_j)}{\ell_i - \ell_j} + 4\psi_i \frac{\partial \psi_i}{\partial \ell_i}. \tag{4.9}
$$

From Proposition 4, the $\psi_i^*(\boldsymbol{\ell})$ for the Stein type estimator $\widehat{\boldsymbol{\Delta}}_2^S$ is given by

$$(n-p-1)d_i^2\ell_i - 2d_i\ell_i + 2\sum_{j\neq i}\frac{d_i\ell_i(d_i\ell_i - d_j\ell_j)}{\ell_i - \ell_j} + 4d_i^2\ell_i,$$

while the corresponding quantity for $\widehat{\boldsymbol{\Delta}}_0$ is $-(n+p+1)^{-1}\ell_i$. Here it is noted that

$$\begin{aligned}
\sum_{j\neq i}\frac{d_i\ell_i(d_i\ell_i - d_j\ell_j)}{\ell_i - \ell_j} &= \left\{\sum_{j=1}^{i-1}\frac{d_i\ell_i(d_j\ell_j - d_i\ell_i)}{\ell_j - \ell_i}\right\} + \left\{\sum_{j=i+1}^{p}\frac{d_i\ell_i(d_i\ell_i - d_j\ell_j)}{\ell_i - \ell_j}\right\} \\
&= \left\{(i-1)d_i^2\ell_i + \sum_{j=1}^{i-1}\frac{d_i(d_j - d_i)\ell_i\ell_j}{\ell_j - \ell_i}\right\} \\
&\quad + \left\{(p-i)d_i^2\ell_i + \sum_{j=i+1}^{p}\frac{d_i(d_i - d_j)\ell_i\ell_j}{\ell_i - \ell_j}\right\} \\
&\leq (p-1)d_i^2\ell_i + \sum_{j=1}^{i-1} d_i(d_j - d_i)\ell_i \qquad (4.10) \\
&= (p-i)d_i^2\ell_i + (\sum_{j=1}^{i-1} d_j)d_i\ell_i,
\end{aligned}$$

where the extreme inequality in (4.10) follows from the inequalities that $\ell_j/(\ell_j - \ell_i) \geq 0$ and

$$\sum_{j=i+1}^{p} d_i(d_i - d_j)\ell_i\ell_j/(\ell_i - \ell_j) < 0,$$

since $d_1 \leq d_2 \leq \cdots \leq d_p$. Hence

$$\begin{aligned}
&(n-p-1)d_i^2\ell_i - 2d_i\ell_i + 2\sum_{j\neq i}\frac{d_i\ell_i(d_i\ell_i - d_j\ell_j)}{\ell_i - \ell_j} + 4d_i^2\ell_i \\
&\leq (n-p-1+2(p-i)+4)d_i^2\ell_i - 2\left(1 - \sum_{j=1}^{i-1} d_j\right) d_i\ell_i \\
&= \left\{(n+p+3-2i)d_i^2 - 2\left(1 - \sum_{j=1}^{i-1} d_j\right) d_i\right\}\ell_i \\
&= -\frac{(1-\sum_{j=1}^{i-1} d_j)^2}{n+p+3-2i}\ell_i
\end{aligned}$$

from the definition of d_i. From the inequality (3.12), it follows that

$$-\frac{(1-\sum_{j=1}^{i-1} d_j)^2}{n+p+3-2i}\ell_i \leq -\frac{\ell_i}{n+p+1},$$

which proves the dominance property of $\widehat{\boldsymbol{\Delta}}_2^S$ over $\widehat{\boldsymbol{\Delta}}_0$.

Corollary 4. *The Stein type orthogonally equivariant estimator* $\widehat{\boldsymbol{\Delta}}_2^S$ *dominates* $\widehat{\boldsymbol{\Delta}}_0$ *relative to the risk (2.6).*

4.3 Stein type scale-equivariant estimators

Let $\boldsymbol{A}$ be a $p \times p$ nonsingular matrix such that

$$\boldsymbol{S} = \boldsymbol{A}\boldsymbol{A}' \quad \text{and} \quad \boldsymbol{W} = \boldsymbol{A}\boldsymbol{F}\boldsymbol{A}' \quad \text{for} \quad \boldsymbol{F} = \text{diag}\,(f_1, \dots, f_p), \quad f_1 \geq f_2 \geq \dots \geq f_p.$$

Then we consider estimators of the form

$$\begin{aligned} \widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi}) &= \boldsymbol{A}\boldsymbol{\Psi}(\boldsymbol{f})\boldsymbol{A}^{-1}, \\ \boldsymbol{\Psi}(\boldsymbol{f}) &= \text{diag}\ (\psi_1(\boldsymbol{f}), \dots, \psi_p(\boldsymbol{f})) \end{aligned} \tag{4.11}$$

for positive functions $\psi_i(\boldsymbol{f})$'s of $\boldsymbol{f} = (f_1, \dots, f_p)$. This estimator is equivariant under the group of scale transformations

$$(\boldsymbol{S}, \boldsymbol{W}) \to (\boldsymbol{B}\boldsymbol{S}\boldsymbol{B}', \boldsymbol{B}\boldsymbol{W}\boldsymbol{B}')$$

for $p \times p$ nonsingular matrix $\boldsymbol{B}$.

Proposition 5. *The risk function of* $\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi})$ *is given by*

$$R(\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi}), \omega) = E_\omega\left[r_3(\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi}))\right] + m\text{tr}\,\boldsymbol{A},$$

where

$$\begin{aligned} r_3(\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi})) = \sum_{i=1}^{p} \Bigg\{ &(n + p - 3) f_i \psi_i^2 - 4 f_i^2 \psi_i \frac{\partial \psi_i}{\partial f_i} - 2 \sum_{j>i} \frac{f_i^2 \psi_i^2 - f_j^2 \psi_j^2}{f_i - f_j} \\ &- 2(m - p + 1)\psi_i - 4 f_i \frac{\partial \psi_i}{\partial f_i} - 4 \sum_{j>i} \frac{f_i \psi_i - f_j \psi_j}{f_i - f_j} \Bigg\}. \end{aligned} \tag{4.12}$$

Let us consider a scale-equivariant Stein type estimator of the form

$$\widehat{\boldsymbol{\Delta}}_3^S = \boldsymbol{A}\text{diag}\left(\frac{b_1}{f_1}, \dots, \frac{b_p}{f_p}\right)\boldsymbol{A}^{-1}, \tag{4.13}$$

for $b_i = (m + p - 2i - 1)/(n - p + 2i + 1)$. These constants b_i's were suggested by Konno (1992a) in estimation of a matrix mean, and have the order relation that $b_1 > b_2 > \dots > b_p$. Using Proposition 5, we shall show that the estimator $\widehat{\boldsymbol{\Delta}}_3^S$ dominates $\widehat{\boldsymbol{\Delta}}_0$ relative to the risk (2.6).

From (4.12), the risk function of the estimator $\widehat{\boldsymbol{\Delta}}_3^S$ is given by

$$
\begin{aligned}
&R(\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi}),\omega) - m\operatorname{tr}\boldsymbol{\Lambda} \qquad (4.14)\\
&= \sum_{i=1}^{p} E_\omega\left[(n+p+1)\frac{b_i^2}{f_i} - 2\sum_{j>i}\frac{b_i^2-b_j^2}{f_i-f_j} - 2(m-p-1)\frac{b_i}{f_i} - 4\sum_{j>i}\frac{b_i-b_j}{f_i-f_j}\right].
\end{aligned}
$$

Konno (1992) has demonstrated that for $k=1,2$ and $b_i > b_j$ for $i<j$,

$$
\begin{aligned}
\sum_{i=1}^{p}\sum_{j=i+1}^{p}\frac{b_i^k-b_j^k}{f_i-f_j} &= \sum_{i=1}^{p}\frac{1}{f_i}\sum_{j=i+1}^{p}\frac{f_i}{f_i-f_j}\left(b_i^k-b_j^k\right)\\
&\geq \sum_{i=1}^{p}\frac{1}{f_i}\sum_{j=i+1}^{p}\left(b_i^k-b_j^k\right) \qquad (4.15)\\
&= \sum_{i=1}^{p}\frac{1}{f_i}\left\{(p-i)d_i^k - \sum_{j=i+1}^{p} b_j^k\right\}.
\end{aligned}
$$

On the other hand, the risk of the estimator $\widehat{\boldsymbol{\Delta}}_0$ is derived by putting $b_1 = \cdots = b_p = (m-p-1)/(n+p+1)$ in (4.14), and we observe that

$$
R(\widehat{\boldsymbol{\Delta}}_0,\omega) - m\operatorname{tr}\boldsymbol{\Lambda} = -\frac{(m-p-1)^2}{n+p+1}\sum_{i+1}^{p}\frac{1}{f_i}. \qquad (4.16)
$$

Combining (4.14), (4.15) and (4.16), we see that the risk difference of $\widehat{\boldsymbol{\Delta}}_3^S$ and $\widehat{\boldsymbol{\Delta}}_0$ is evaluated as

$$
R(\widehat{\boldsymbol{\Delta}}_3^S,\omega) - R(\widehat{\boldsymbol{\Delta}}_0,\omega) = \sum_{i=1}^{p} E_\omega\left[f_i^{-1}h(i)\right],
$$

where

$$
h(i) = (n-p+2i+1)b_i^2 - 2(m+p-2i-1)b_i + 2\sum_{j=i+1}^{p} b_j(b_j+2) + \frac{(m-p-1)^2}{n+p+1},
$$

so that it is sufficient to show that

$$
h(i) \leq 0 \qquad (4.17)
$$

for $i = 1,\ldots,p-1$.

To verify that (4.17), we note that $h(i-1)$ is rewritten as

$$
\begin{aligned}
h(i-1) &= -\frac{(m+p-2i+1)^2}{n-p+2i-1} + 2\sum_{j=i}^{p} b_j(b_j+2) + \frac{(m-p-1)^2}{n+p+1}\\
&= h(i) - \frac{(a_i'+2)^2}{a_i-2} + 2\frac{a_i'^2}{a_i^2} + 4\frac{a_i'}{a_i} + \frac{a_i'^2}{a_i}\\
&= \sum_{k=i}^{p}\left\{-\frac{(a_k'+2)^2}{a_k-2} + 2\frac{a_k'^2}{a_k^2} + 4\frac{a_k'}{a_k} + \frac{a_k'^2}{a_k}\right\}, \qquad (4.18)
\end{aligned}
$$

where $a_i' = m + p - 2i$ and $a_i = n - p + 2i + 1$. It can be easily checked that for each k,

$$-\frac{(a_k'+2)^2}{a_k-2} + 2\frac{a_k'^2}{a_k^2} + 4\frac{a_k'}{a_k} + \frac{a_k'^2}{a_k} \le 0,$$

which shows the inequality (4.17).

Corollary 5. *The scale-equivariant Stein type estimator* $\widehat{\boldsymbol{\Delta}}_3^S$ *dominates* $\widehat{\boldsymbol{\Delta}}_0$ *relative to the risk (2.6).*

4.4 Proofs

We here derive the exact expressions of the risk functions given in Propositions 3, 4 and 5. An essential tool for their derivations is the Wishart identity obtained by Stein (1977) and Haff (1979).

Let $\boldsymbol{G}(\boldsymbol{S})$ be a $p \times p$ matrix such that the (i,j) element $g_{ij}(\boldsymbol{S})$ is a function of $\boldsymbol{S} = (s_{ij})$ and define the differential operator $\boldsymbol{D}_S$ by

$$\{\boldsymbol{D}_S \boldsymbol{G}(\boldsymbol{S})\}_{ij} = \sum_{a=1}^{p} d_{ia} g_{aj}(\boldsymbol{S}),$$

where

$$d_{ia} = \frac{1}{2}(1+\delta_{ia})\frac{\partial}{\partial s_{ia}}$$

with $\delta_{ia} = 1$ for $i = a$ and $\delta_{ia} = 0$ for $i \ne a$. When $\boldsymbol{S}$ is distributed as $\mathcal{W}_p(\boldsymbol{\Sigma}, n)$, the following identity holds:

$$E_\Sigma\left[\mathrm{tr}\left\{\boldsymbol{G}(\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\right\}\right] = E_\Sigma\left[(n-p-1)\mathrm{tr}\left\{\boldsymbol{G}(\boldsymbol{S})\boldsymbol{S}^{-1}\right\} + 2\mathrm{tr}\left\{\boldsymbol{D}_S\boldsymbol{G}(\boldsymbol{S})\right\}\right].$$

It is noted that this identity can be extended to elliptically contoured distributions as shown by Kubokawa and Srivastava (1999). Also note that a similar identity is given for $\boldsymbol{W}$ having $\mathcal{W}(\boldsymbol{\Sigma}_2, m)$.

Applying the Stein-Haff identity, we observe that

$$\begin{aligned} R^*(\widehat{\boldsymbol{\Delta}}, \omega) &= R(\widehat{\boldsymbol{\Delta}}, \omega) - m\mathrm{tr}\,\boldsymbol{\Lambda} \\ &= E_\omega\left[\mathrm{tr}\left(\widehat{\boldsymbol{\Delta}}\boldsymbol{W}\widehat{\boldsymbol{\Delta}}'\boldsymbol{\Sigma}^{-1}\right) - 2\mathrm{tr}\left(\widehat{\boldsymbol{\Delta}}\boldsymbol{W}\boldsymbol{\Sigma}_2^{-1}\right)\right] \qquad (4.19) \\ &= E_\omega\Big[(n-p-1)\mathrm{tr}\,\widehat{\boldsymbol{\Delta}}\boldsymbol{W}\widehat{\boldsymbol{\Delta}}'\boldsymbol{S}^{-1} + 2\mathrm{tr}\,\boldsymbol{D}_S(\widehat{\boldsymbol{\Delta}}\boldsymbol{W}\widehat{\boldsymbol{\Delta}}') \\ &\quad -2(m-p-1)\mathrm{tr}\,\widehat{\boldsymbol{\Delta}} - 4\mathrm{tr}\,\boldsymbol{D}_W(\widehat{\boldsymbol{\Delta}}\boldsymbol{W})\Big], \end{aligned}$$

which is the basis of the next derivation.

Proof of Proposition 3. Since $\widehat{\boldsymbol{\Delta}}_1(\boldsymbol{\Psi}) = a\boldsymbol{S}\boldsymbol{H}\boldsymbol{\Psi}(\boldsymbol{x})\boldsymbol{H}'$ for $a = (n+p+1)^{-1}$, we have that

$$
\begin{aligned}
&R^*(\widehat{\boldsymbol{\Delta}}_1(\boldsymbol{\Psi}), \omega) \\
&= E_\omega \Big[(n-p-1)a^2 \mathrm{tr}\, \boldsymbol{S}\boldsymbol{H}\{\boldsymbol{\Psi}(\boldsymbol{x})\}^2\boldsymbol{X}\boldsymbol{H}' + 2a^2 \mathrm{tr}\, \boldsymbol{D}_S\left\{\boldsymbol{S}\boldsymbol{H}\{\boldsymbol{\Psi}(\boldsymbol{x})\}^2\boldsymbol{X}\boldsymbol{H}'\boldsymbol{S}\right\} \\
&\quad -2(m-p-1)a \mathrm{tr}\, \boldsymbol{S}\boldsymbol{H}\boldsymbol{\Psi}(\boldsymbol{x})\boldsymbol{H}' - 4\mathrm{tr}\, \boldsymbol{D}_W\left\{\boldsymbol{H}\boldsymbol{\Psi}(\boldsymbol{x})\boldsymbol{X}\boldsymbol{H}'\boldsymbol{S}\right\}\Big]. \qquad (4.20)
\end{aligned}
$$

Here for $\boldsymbol{B} = \boldsymbol{H}\{\boldsymbol{\Psi}(\boldsymbol{x})\}^2\boldsymbol{X}\boldsymbol{H}'$, from Haff (1981) and Konno (1991), it is observed that

$$
\mathrm{tr}\, \boldsymbol{D}_S(\boldsymbol{S}\boldsymbol{B}\boldsymbol{S}) = \mathrm{tr}\,(\boldsymbol{D}_S\boldsymbol{S})\boldsymbol{B}\boldsymbol{S} + \mathrm{tr}\,(\boldsymbol{S}\boldsymbol{D}_S)'\boldsymbol{B}\boldsymbol{S}. \qquad (4.21)
$$

Since $(\boldsymbol{D}_S\boldsymbol{S})_{ij} = \sum_a d_{ia}s_{aj} = 2^{-1}(p+1)\delta_{ij}$, the first term is evaluated as

$$
\mathrm{tr}\,(\boldsymbol{D}_S\boldsymbol{S})\boldsymbol{B}\boldsymbol{S} = \frac{p+1}{2}\mathrm{tr}\, \boldsymbol{B}\boldsymbol{S}.
$$

Similarly, the second term is

$$
\mathrm{tr}\,(\boldsymbol{S}\boldsymbol{D}_S)'\boldsymbol{B}\boldsymbol{S} = \frac{p+1}{2}\mathrm{tr}\, \boldsymbol{B}\boldsymbol{S}.
$$

Hence the first two terms in (4.20) is equal to

$$
(n+p+1)a^2 \mathrm{tr}\, \boldsymbol{S}\boldsymbol{H}\{\boldsymbol{\Psi}(\boldsymbol{x})\}^2\boldsymbol{X}\boldsymbol{H}'. \qquad (4.22)
$$

Concerning the fourth term in (4.20), Stein (1977) showed that for $\boldsymbol{\Phi}(\boldsymbol{x}) = \mathrm{diag}\,(\phi_1(\boldsymbol{x}), \ldots, \phi_p(\boldsymbol{x}))$,

$$
\boldsymbol{D}_W(\boldsymbol{H}\boldsymbol{\Phi}(\boldsymbol{x})\boldsymbol{H}') = \boldsymbol{H}\boldsymbol{\Phi}^{(1)}(\boldsymbol{x})\boldsymbol{H}', \qquad (4.23)
$$

where $\boldsymbol{\Phi}^{(1)}(\boldsymbol{x}) = \mathrm{diag}\,(\phi_1^{(1)}(\boldsymbol{x}), \ldots, \phi_p^{(1)}(\boldsymbol{x}))$ for

$$
\phi_i^{(1)}(\boldsymbol{x}) = \frac{1}{2}\sum_{j\neq i}\frac{\phi_i(\boldsymbol{x}) - \phi_j(\boldsymbol{x})}{x_i - x_j} + \frac{\partial\phi_i(\boldsymbol{x})}{\partial x_i}.
$$

This calculation is used to get that

$$
\begin{aligned}
\mathrm{tr}\, \boldsymbol{D}_W(\boldsymbol{H}\boldsymbol{\Psi}(\boldsymbol{x})\boldsymbol{X}\boldsymbol{H}'\boldsymbol{S}) &= \mathrm{tr}\, \boldsymbol{D}_W(\boldsymbol{H}\boldsymbol{\Psi}(\boldsymbol{x})\boldsymbol{X}\boldsymbol{H}')\boldsymbol{S} \\
&= \mathrm{tr}\, \boldsymbol{H}\boldsymbol{\Phi}^*(\boldsymbol{x})\boldsymbol{H}'\boldsymbol{S}, \qquad (4.24)
\end{aligned}
$$

where $\boldsymbol{\Phi}^*(\boldsymbol{x}) = \mathrm{diag}\,(\phi_1^*(\boldsymbol{x}), \ldots, \phi_p^*(\boldsymbol{x}))$ for

$$
\phi_i^*(\boldsymbol{x}) = \frac{1}{2}\sum_{j\neq i}\frac{x_i\psi_i - x_j\psi_j}{x_i - x_j} + \psi_i + x_i\frac{\partial\psi_i}{\partial x_i}.
$$

Combining (4.20), (4.22) and (4.24) provides the expression of the risk (4.3) in Proposition 3. □□

Proof of Proposition 4. Since $\widehat{\boldsymbol{\Delta}}_2(\boldsymbol{\Psi}) = a\boldsymbol{K}\boldsymbol{\Psi}(\boldsymbol{\ell})\boldsymbol{K}'\boldsymbol{W}^{-1}$ for $a = m - p - 1$, $R^*(\widehat{\boldsymbol{\Delta}}_2(\boldsymbol{\Psi}), \omega)$ is rewritten by

$$\begin{aligned} R^*(\widehat{\boldsymbol{\Delta}}_2(\boldsymbol{\Psi}), \omega) &= a^2 E_\omega \Big[(n-p-1)\mathrm{tr}\, \boldsymbol{K}\{\boldsymbol{\Psi}(\boldsymbol{\ell})\}^2 \boldsymbol{L}^{-1}\boldsymbol{K}'\boldsymbol{W}^{-1} \\ &\quad + 2\mathrm{tr}\, \boldsymbol{D}_S(\boldsymbol{K}\boldsymbol{\Psi}(\boldsymbol{\ell})\boldsymbol{K}'\boldsymbol{W}^{-1}\boldsymbol{K}\boldsymbol{\Psi}(\boldsymbol{\ell})\boldsymbol{K}') - 2\mathrm{tr}\, \boldsymbol{K}\boldsymbol{\Psi}(\boldsymbol{\ell})\boldsymbol{K}'\boldsymbol{W}^{-1}\Big]. \end{aligned} \tag{4.25}$$

Using the calculations (4.21) and (4.23), we see that

$$\begin{aligned} \mathrm{tr}\, \boldsymbol{D}_S(\boldsymbol{B}\boldsymbol{W}^{-1}\boldsymbol{B}) &= \mathrm{tr}\, (\boldsymbol{D}_S\boldsymbol{B})\boldsymbol{W}^{-1}\boldsymbol{B} + \mathrm{tr}\, (\boldsymbol{B}\boldsymbol{D}_S)'\boldsymbol{W}^{-1}\boldsymbol{B} \\ &= 2\mathrm{tr}\, \boldsymbol{K}\boldsymbol{\Psi}^{**}(\boldsymbol{\ell})\boldsymbol{K}'\boldsymbol{W}^{-1}\boldsymbol{B}, \end{aligned} \tag{4.26}$$

where $\boldsymbol{B} = \boldsymbol{K}\boldsymbol{\Psi}(\boldsymbol{\ell})\boldsymbol{K}'$ and $\boldsymbol{\Psi}^{**} = \mathrm{diag}\,(\psi_1^{**}, \dots, \psi_p^{**})$ for

$$\psi_i^{**}(\boldsymbol{\ell}) = \frac{1}{2}\sum_{j \neq i} \frac{\psi_i - \psi_j}{\ell_i - \ell_j} + \frac{\partial \psi_i}{\partial \ell_i}.$$

Therefore the expression (4.8) in Proposition 4 is obtained by combining (4.25) and (4.26). □□

Proof of Proposition 5. Since $\widehat{\boldsymbol{\Delta}}_3 = \boldsymbol{A}\boldsymbol{\Psi}(\boldsymbol{f})\boldsymbol{A}^{-1}$ with $\boldsymbol{S} = \boldsymbol{A}\boldsymbol{A}'$ and $\boldsymbol{W} = \boldsymbol{A}\boldsymbol{F}\boldsymbol{A}'$, $\boldsymbol{R}^*(\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi}), \omega)$ is rewritten by

$$\begin{aligned} \boldsymbol{R}^*(\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi}), \omega) &= E_\omega\left[(n-p-1)\mathrm{tr}\, \boldsymbol{\Psi}\boldsymbol{F}\boldsymbol{\Psi} + 2\mathrm{tr}\, \boldsymbol{D}_S(\boldsymbol{A}\boldsymbol{\Psi}\boldsymbol{F}\boldsymbol{\Psi}\boldsymbol{A}') \right. \\ &\quad \left. -2(m-p-1)\mathrm{tr}\, \boldsymbol{\Psi} - 4\mathrm{tr}\, \boldsymbol{D}_W(\boldsymbol{A}\boldsymbol{\Psi}\boldsymbol{F}\boldsymbol{A}')\right]. \end{aligned} \tag{4.27}$$

Here the following calculations due to Loh (1988) and Konno (1992a) are useful:

$$\mathrm{tr}\, \boldsymbol{D}_S(\boldsymbol{A}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{A}') = \sum_{i=1}^p \left\{ p\phi_i - f_i\frac{\partial \phi_i}{\partial f_i} - \sum_{j>i} \frac{f_i\phi_i - f_j\phi_j}{f_i - f_j} \right\}, \tag{4.28}$$

$$\mathrm{tr}\, \boldsymbol{D}_W(\boldsymbol{A}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{A}') = \sum_{i=1}^p \left\{ \frac{\partial \phi_i}{\partial f_i} + \sum_{j>i} \frac{\phi_i - \phi_j}{f_i - f_j} \right\}, \tag{4.29}$$

where $\boldsymbol{\Phi}(\boldsymbol{\ell}) = \mathrm{diag}\,(\phi_1(\boldsymbol{\ell}), \dots, \phi_p(\boldsymbol{\ell}))$. Combining (4.27), (4.28) and (4.29) provides that

$$\begin{aligned} R^*(\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi}), \omega) &= \sum_{i=1}^p E_\omega \left[(n-p-1)f_i\psi_i^2 + 2pf_i\psi_i^2 - 2f_i\left(\psi_i^2 + 2f_i\psi_i\frac{\partial \psi_i}{\partial f_i}\right) \right. \\ &\quad -2\sum_{j>i}\frac{f_i^2\psi_i^2 - f_j^2\psi_j^2}{f_i - f_j} - 2(m-p-1)\psi_i - 4\left(\psi_i + f_i\frac{\partial \psi_i}{\partial f_i}\right) \\ &\quad \left. -4\sum_{j>i}\frac{f_i\psi_i - f_j\psi_j}{f_i - f_j}\right], \end{aligned}$$

which leads to the expression (4.12) of Proposition 5. □□

5 Other Improved Procedures

5.1 Efron-Morris type improved estimators

As another type of orthogonally equivariant estimators, we here consider the Efron-Morris type estimators and show their superiority to $\widehat{\boldsymbol{\Delta}}_0$.

Corresponding to three kinds of estimators in Section 3, we treat the following orthogonally equivariant estimators

$$
\begin{aligned}
\widehat{\boldsymbol{\Delta}}_1^{EM} &= \frac{1}{n+p+1}\boldsymbol{S}\left\{\alpha_1 \boldsymbol{W}^{-1} + \beta_1 \frac{1}{\mathrm{tr}\,\boldsymbol{W}}\right\}, &(5.1)\\
\widehat{\boldsymbol{\Delta}}_2^{EM} &= (m-p-1)\left\{\alpha_2 \boldsymbol{S} + \beta_2 \frac{1}{\mathrm{tr}\,\boldsymbol{S}^{-1}}\right\}\boldsymbol{W}^{-1}, &(5.2)\\
\widehat{\boldsymbol{\Delta}}_3^{EM} &= \alpha_3 \boldsymbol{S}\boldsymbol{W}^{-1} + \beta_3 \frac{1}{\mathrm{tr}\,\boldsymbol{S}^{-1}\boldsymbol{W}}\boldsymbol{I}_p, &(5.3)
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha_1 &= m-p-1, & \beta_1 &= p-1,\\
\alpha_2 &= \frac{1}{n+p+1}, & \beta_2 &= \frac{p-1}{(n+p+1)(n-p+3)},\\
\alpha_3 &= \frac{m-p-1}{n+p+1}, & \beta_3 &= \frac{(p-1)(p+2)(n+m)}{(n+p+1)(n-p+3)}.
\end{aligned}
$$

The risk function of $\widehat{\boldsymbol{\Delta}}_1^{EM}$ can be provided by Proposition 3 where $\psi_i^*(\boldsymbol{x})$ in (4.4) is calculated as

$$
\begin{aligned}
\psi_i^{*EM}(\boldsymbol{x}) &= \frac{\alpha_1^2 - 2(m-p-1)\alpha_1}{x_i} - 2\frac{(m-p-1-\alpha_1)\beta_1 + (p+1)\beta_1}{\sum_j x_j} + \frac{(\beta_1^2 + 4\beta_1)x_i}{(\sum_j x_j)^2}\\
&\le \frac{\alpha_1^2 - 2(m-p-1)\alpha_1}{x_i} + \frac{\beta_1^2 - 2(p-1)\beta_1 - 2(m-p-1-\alpha_1)\beta_1}{\sum_j x_j}\\
&= -\frac{(m-p-1)^2}{x_i} - \frac{(p-1)^2}{\sum_j x_j}. \qquad (5.4)
\end{aligned}
$$

Since the corresponding quantity in the risk of $\widehat{\boldsymbol{\Delta}}_0$ is $-(m-p-1)^2/x_i$, the inequality (5.4) demonstrates that $\widehat{\boldsymbol{\Delta}}_1^{EM}$ dominates $\widehat{\boldsymbol{\Delta}}_0$.

For the risk of $\widehat{\boldsymbol{\Delta}}_2^{EM}$, the $\psi_i^*(\boldsymbol{\ell})$ in (4.9) of Proposition 4 is evaluated as

$$
\begin{aligned}
\psi_i^{*EM}(\boldsymbol{\ell}) &= \left\{(n+p+1)\alpha_2^2 - 2\alpha_2\right\}\ell_i + \frac{2(n\alpha_2 - 1)\beta_2}{\sum_j \ell_j^{-1}}\\
&\quad + \frac{1}{\ell_i(\sum_j \ell_j^{-1})^2}\left\{(n-p-1)\beta_2^2 + 4\alpha_2\beta_2 + \frac{4\beta_2^2}{\ell_i\sum_j \ell_j^{-1}}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\le \left\{(n+p+1)\alpha_2^2 - 2\alpha_2\right\}\ell_i + \frac{(n-p+3)\beta_2^2 - 2\left(1-(n+2)\alpha_2\right)\beta_2}{\ell_i(\sum_j \ell_j^{-1})^2} \\
&= -\frac{\ell_i}{n+p+1} - \frac{(p-1)^2}{(n+p+1)^2(n-p+3)}\frac{1}{\ell_i(\sum_j \ell_j^{-1})^2},
\end{aligned}
$$

which is smaller than $-\ell_i/(n+p+1)$, the corresponding quantity in the risk of $\widehat{\boldsymbol{\Delta}}_0$. This shows the dominance property of $\widehat{\boldsymbol{\Delta}}_2^{EM}$ over $\widehat{\boldsymbol{\Delta}}_0$.

Finally for the risk of $\widehat{\boldsymbol{\Delta}}_3^{EM}$, it follows from Proposition 5 that

$$
\begin{aligned}
r_3(\widehat{\boldsymbol{\Delta}}_3^{EM}) &= \sum_i f_i^{-1}\left\{(n+p+1)\alpha_3^2 - 2(m-p-1)\alpha_3\right\} + 4\beta_3^2\frac{\sum_i f_i^2}{(\sum_i f_i)^3} \\
&\quad + \frac{1}{\sum_i f_i}\left\{(n-p-1)\beta_3^2 - 2\left[(p-1)(p+2) + p(m-p-1) - (np+2)\alpha_3\right]\right\} \\
&\le \sum_i f_i^{-1}\left\{(n+p+1)\alpha_3^2 - 2(m-p-1)\alpha_3\right\} \\
&\quad + \frac{1}{\sum_i f_i}\left\{(n-p+3)\beta_3^2 - 2\left[(p-1)(p+2) + p(m-p-1) - (np+2)\alpha_3\right]\right\} \\
&= -\sum_i f_i^{-1}\frac{(m-p-1)^2}{n+p+1} + \frac{1}{\sum_i f_i}\left\{(n-p+3)\beta_3^2 - 2(p-1)(p+2)(1+\alpha_3)\right\},
\end{aligned}
$$

where the last equality is derived by substituting $\alpha_3 = (m-p-1)/(n+p+1)$. Since $\beta_3 = (p-1)(p+2)(1+\alpha_3)/(n-p+3)$, we get that

$$
r_3(\widehat{\boldsymbol{\Delta}}_3^{EM}) \le -\sum_i f_i^{-1}\frac{(m-p-1)^2}{n+p+1} - \frac{1}{\sum_i f_i}\frac{1}{n-p+3}\left\{\frac{(p-1)(p+2)(n+m)}{n+p+1}\right\}^2,
$$

which is smaller than $r_3(\widehat{\boldsymbol{\Delta}}_0) = -\sum_i f_i^{-1}(m-p-1)^2/(n+p+1)$.

In the sequel, the above three dominance results are summarized by the following corollary.

Corollary 6. *The Efron-Morris type estimators* $\widehat{\boldsymbol{\Delta}}_1^{EM}$, $\widehat{\boldsymbol{\Delta}}_2^{EM}$ *and* $\widehat{\boldsymbol{\Delta}}_3^{EM}$ *dominate* $\widehat{\boldsymbol{\Delta}}_0$ *relative to the risk (2.6).*

5.2 Improvements by use of order restriction

Recalling that $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma} + r\boldsymbol{\Sigma}_A$, we notice that there is the order restriction $\boldsymbol{\Sigma}_2 > \boldsymbol{\Sigma}$ between $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_2$. The estimators treated in the previous sections can be shown to be further improved upon by using this knowledge. The resulting estimators of $\boldsymbol{\Theta}$ correspond to the positive-part Stein estimator for $m = 1$.

Let us denote any estimator of $\boldsymbol{\Delta} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1}$ by $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_2^{-1}$. For the $\widehat{\boldsymbol{\Sigma}}$, let $\widehat{\boldsymbol{A}}$ be a $p\times p$ nonsingular matrix such that $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}'$. Let $\boldsymbol{\Gamma}$ be a $p\times p$ orthogonal matrix and

$\boldsymbol{G} = \mathrm{diag}\,(g_1, \ldots, g_p)$, $g_1 \geq \ldots \geq g_p$, such that

$$\widehat{\boldsymbol{A}}'\widehat{\boldsymbol{\Sigma}}_2^{-1}\widehat{\boldsymbol{A}} = \boldsymbol{\Gamma G \Gamma}'. \tag{5.5}$$

We thus employ the notation

$$\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Delta}}(\boldsymbol{G}) = \widehat{\boldsymbol{A}}\boldsymbol{\Gamma G \Gamma}'\widehat{\boldsymbol{A}}^{-1}.$$

For improving upon $\widehat{\boldsymbol{\Delta}}(\boldsymbol{G})$, we consider truncating $\boldsymbol{G}$ as

$$\boldsymbol{G}^{TR} = \begin{cases} \boldsymbol{I}_p & \text{if} \quad \boldsymbol{G} \geq \boldsymbol{I}_p \\ \boldsymbol{G} & \text{otherwise,} \end{cases} \tag{5.6}$$

yielding the estimator

$$\widehat{\boldsymbol{\Delta}}^{TR} = \widehat{\boldsymbol{\Delta}}(\boldsymbol{G}^{TR}) = \begin{cases} \boldsymbol{I}_p & \text{if} \quad \boldsymbol{G} \geq \boldsymbol{I}_p \\ \widehat{\boldsymbol{\Delta}}(\boldsymbol{G}) & \text{otherwise.} \end{cases} \tag{5.7}$$

For verifying this dominance property, let $\widetilde{\boldsymbol{W}} = \widehat{\boldsymbol{A}}^{-1}\boldsymbol{W}\widehat{\boldsymbol{A}}'^{-1}$, $\widetilde{\boldsymbol{\Sigma}}^{-1} = \widehat{\boldsymbol{A}}'\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{\Sigma}}_2^{-1} = \widehat{\boldsymbol{A}}'\boldsymbol{\Sigma}_2^{-1}\widehat{\boldsymbol{A}}$. The risk difference of $\widehat{\boldsymbol{\Delta}}$ and $\widehat{\boldsymbol{\Delta}}^{TR}$ is

$$\begin{aligned}
& R(\widehat{\boldsymbol{\Delta}}, \omega) - R(\widehat{\boldsymbol{\Delta}}^{TR}, \omega) \\
&= E_\omega\left[\mathrm{tr}\,\boldsymbol{\Gamma G \Gamma}'\widetilde{\boldsymbol{W}}\boldsymbol{\Gamma G \Gamma}'\widetilde{\boldsymbol{\Sigma}}^{-1} - 2\mathrm{tr}\,\boldsymbol{\Gamma G \Gamma}'\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{\Sigma}}_2^{-1}\right] \\
&\quad -E_\omega\left[\mathrm{tr}\,\boldsymbol{\Gamma G}^{TR}\boldsymbol{\Gamma}'\widetilde{\boldsymbol{W}}\boldsymbol{\Gamma G}^{TR}\boldsymbol{\Gamma}'\widetilde{\boldsymbol{\Sigma}}^{-1} - 2\mathrm{tr}\,\boldsymbol{\Gamma G}^{TR}\boldsymbol{\Gamma}'\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{\Sigma}}_2^{-1}\right] \\
&= E_\omega\left[\mathrm{tr}\,\boldsymbol{\Gamma}\left(\boldsymbol{G} - \boldsymbol{G}^{TR}\right)\boldsymbol{\Gamma}'\widetilde{\boldsymbol{W}}\boldsymbol{\Gamma}\left(\boldsymbol{G} + \boldsymbol{G}^{TR}\right)\boldsymbol{\Gamma}'\widetilde{\boldsymbol{\Sigma}}^{-1} - 2\mathrm{tr}\,\boldsymbol{\Gamma}\left(\boldsymbol{G} - \boldsymbol{G}^{TR}\right)\boldsymbol{\Gamma}'\widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{\Sigma}}_2^{-1}\right].
\end{aligned}$$

Noting that $\boldsymbol{G} - \boldsymbol{G}^{TR} \geq \boldsymbol{0}$ and that $\widetilde{\boldsymbol{\Sigma}}^{-1} \geq \widetilde{\boldsymbol{\Sigma}}_2^{-1}$, we evaluate the risk difference as

$$\begin{aligned}
& R(\widehat{\boldsymbol{\Delta}}, \omega) - R(\widehat{\boldsymbol{\Delta}}^{TR}, \omega) \\
&\geq E_\omega\left[\mathrm{tr}\,\boldsymbol{\Gamma}\left(\boldsymbol{G} - \boldsymbol{G}^{TR}\right)\boldsymbol{\Gamma}'\widetilde{\boldsymbol{W}}\boldsymbol{\Gamma}\left(\boldsymbol{G} + \boldsymbol{G}^{TR} - 2\boldsymbol{I}_p\right)\boldsymbol{\Gamma}'\widetilde{\boldsymbol{\Sigma}}^{-1}\right] \\
&= E_\omega\left[\mathrm{tr}\,\boldsymbol{\Gamma}\left(\boldsymbol{G} - \boldsymbol{I}_p\right)\boldsymbol{\Gamma}'\widetilde{\boldsymbol{W}}\boldsymbol{\Gamma}\left(\boldsymbol{G} - \boldsymbol{I}_p\right)\boldsymbol{\Gamma}'\widetilde{\boldsymbol{\Sigma}}^{-1}I(\boldsymbol{G} \geq \boldsymbol{I}_p)\right] \\
&\geq 0,
\end{aligned} \tag{5.8}$$

where $I(\cdot)$ is the indicator function. Hence we get

Proposition 6. *The truncated estimator* $\widehat{\boldsymbol{\Delta}}^{TR} = \widehat{\boldsymbol{\Delta}}(\boldsymbol{G}^{TR})$ *given by (5.7) dominates* $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Delta}}(\boldsymbol{G})$ *relative to the risk (2.6).*

The truncated rule $\boldsymbol{G}^{TR}$ given by (5.6) takes the truncated value $\boldsymbol{I}_p$ when $g_p \geq 1$ since $g_1 \geq \ldots \geq g_p$. Instead of this, it seems desirable to consider a rule that truncates componentwise. We thus deal with the truncation rule

$$\boldsymbol{G}^* = \text{diag}\,(\min(g_1, 1), \ldots, \min(g_p, 1))\,.$$

Then from (5.8), the risk difference is evaluated as

$$\begin{aligned}
&R(\widehat{\boldsymbol{\Delta}}, \omega) - R(\widehat{\boldsymbol{\Delta}}(\boldsymbol{G}^*), \omega) \\
&\geq \sum_{r=1}^{p} E_\omega \Bigg[\text{tr} \begin{pmatrix} \boldsymbol{G}_r - \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \boldsymbol{\Gamma}' \widetilde{\boldsymbol{W}} \boldsymbol{\Gamma} \\
&\quad \times \begin{pmatrix} \boldsymbol{G}_r - \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & 2(\overline{\boldsymbol{G}}_{p-r} - \boldsymbol{I}_{p-r}) \end{pmatrix} \boldsymbol{\Gamma}' \widehat{\boldsymbol{\Sigma}}_2^{-1} \boldsymbol{\Gamma} I(g_r > 1 \geq g_{r+1}) \Bigg]
\end{aligned}$$

where $\boldsymbol{G}_r = \text{diag}\,(g_1, \ldots, g_r)$ and $\overline{\boldsymbol{G}}_{p-r} = \text{diag}\,(g_{r+1}, \ldots, g_p)$. Noting that $\overline{\boldsymbol{G}}_{p-r} - \boldsymbol{I}_{p-r} < \boldsymbol{0}$, we see that $R(\widehat{\boldsymbol{\Delta}}, \omega) \geq R(\widehat{\boldsymbol{\Delta}}(\boldsymbol{G}^*), \omega)$ if $\boldsymbol{\Gamma}' \widetilde{\boldsymbol{W}} \boldsymbol{\Gamma}$ is a diagonal matrix.

Proposition 7. *If* $\widetilde{\boldsymbol{W}}^{-1} = \widehat{\boldsymbol{A}}' \boldsymbol{W}^{-1} \widehat{\boldsymbol{A}}$ *and* $\widehat{\boldsymbol{A}}' \widehat{\boldsymbol{\Sigma}}_2^{-1} \widehat{\boldsymbol{A}}$ *are simultaneously diagonalizable, then the componentwise truncated estimator* $\widehat{\boldsymbol{\Delta}}(\boldsymbol{G}^*)$ *dominates* $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Delta}}(\boldsymbol{G})$ *relative to the risk (2.6).*

Since $\widehat{\boldsymbol{\Sigma}}_2^{JS}$ and $\widehat{\boldsymbol{\Sigma}}_2^{S}$ take $\widehat{\boldsymbol{\Sigma}}_2^{-1} = (m - p - 1)\boldsymbol{W}^{-1}$, the condition of Proposition 7 is satisfied, and they are improved upon by the corresponding componentwise truncated estimators.

The same arguments can be applied to the scale-equivariant estimators of the form

$$\begin{aligned}
\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi}) &= \boldsymbol{A} \boldsymbol{\Psi}(\boldsymbol{f}) \boldsymbol{A}^{-1}, \\
\boldsymbol{\Psi}(\boldsymbol{f}) &= \text{diag}\,(\psi_1(\boldsymbol{f}), \ldots, \psi_p(\boldsymbol{f}))\,.
\end{aligned}$$

As a componentwise truncated estimator, we consider the form of the form

$$\begin{aligned}
\widehat{\boldsymbol{\Delta}}_3^{TR} &= \widehat{\delta}_3(\boldsymbol{\Psi}^{TR}) = \boldsymbol{A} \boldsymbol{\Psi}^{TR}(\boldsymbol{f}) \boldsymbol{A}^{-1}, \\
\boldsymbol{\Psi}^{TR}(\boldsymbol{f}) &= \text{diag}\,(\min\{\psi_1(\boldsymbol{f}), 1\}, \ldots, \min\{\psi_p(\boldsymbol{f}), 1\})\,,
\end{aligned}$$

and we can get

Corollary 7. *The componentwise truncated estimator* $\widehat{\boldsymbol{\Delta}}_3^{TR} = \widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi}^{TR})$ *dominates* $\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi})$ *relative to the risk (2.6).*

6 Estimation of a Matrix Mean in a Fixed Effects Model

In the previous sections, several types of estimators dominating $\widehat{\boldsymbol{\Delta}}_0$ have been proposed, and from (2.3) and (2.5), it is seen that the resulting estimators of $\boldsymbol{\Theta}$ are better than the estimator

$$\widehat{\boldsymbol{\Theta}}_0 = (\overline{\boldsymbol{y}}_{1\cdot}, \dots, \overline{\boldsymbol{y}}_{k\cdot}) - \widehat{\boldsymbol{\Delta}}_0 \widetilde{\boldsymbol{Y}},$$

where

$$\begin{aligned} \widetilde{\boldsymbol{Y}} &= \left(\overline{\boldsymbol{y}}_{1\cdot} - \widehat{\boldsymbol{\beta}}\boldsymbol{b}_1, \dots, \overline{\boldsymbol{y}}_{k\cdot} - \widehat{\boldsymbol{\beta}}\boldsymbol{b}_k\right) \\ \widehat{\boldsymbol{\Delta}}_0 &= \frac{m-p-1}{n+p+1}\boldsymbol{S}\boldsymbol{W}^{-1}. \end{aligned}$$

In this section, we demonstrates that these dominance results hold still in the fixed effects models.

Consider the fixed effects model (2.1) where $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_k$ are $p \times 1$ unknown fixed effects such that $\sum_{i=1}^k \boldsymbol{\alpha}_i \boldsymbol{b}_i' = \boldsymbol{0}$. Let $\boldsymbol{B} = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_k)$ and assume that $\mathrm{rank}\,(\boldsymbol{B}) = q_1 \le q < k$. Letting $\boldsymbol{P} = \boldsymbol{B}'(\boldsymbol{B}\boldsymbol{B}')^{-}\boldsymbol{B}$, we observe that

$$\begin{aligned} (\overline{\boldsymbol{y}}_{1\cdot} - \boldsymbol{\theta}_1, \dots, \overline{\boldsymbol{y}}_{k\cdot} - \boldsymbol{\theta}_k)\,\boldsymbol{P} &= \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\boldsymbol{B}, \\ (\overline{\boldsymbol{y}}_{1\cdot} - \boldsymbol{\theta}_1, \dots, \overline{\boldsymbol{y}}_{k\cdot} - \boldsymbol{\theta}_k)\,(\boldsymbol{I}_p - \boldsymbol{P}) &= \widetilde{\boldsymbol{Y}} - \tilde{\boldsymbol{\alpha}}, \end{aligned}$$

where $\tilde{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_k)$. When we look into estimators of the general form

$$\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}) = (\overline{\boldsymbol{y}}_{1\cdot}, \dots, \overline{\boldsymbol{y}}_{k\cdot}) - \widehat{\boldsymbol{\Delta}}\widetilde{\boldsymbol{Y}} \tag{6.1}$$

for $p \times p$ matrix $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Delta}}(\boldsymbol{S}, \boldsymbol{W})$, the difference $\boldsymbol{\Theta}(\widehat{\boldsymbol{\Delta}}) - \boldsymbol{\Theta}$ is written by

$$\begin{aligned} \widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}) - \boldsymbol{\Theta} &= (\overline{\boldsymbol{y}}_{1\cdot} - \boldsymbol{\theta}_1, \dots, \overline{\boldsymbol{y}}_{k\cdot} - \boldsymbol{\theta}_k)\,(\boldsymbol{P} + \boldsymbol{I}_p - \boldsymbol{P}) - \widehat{\boldsymbol{\Delta}}\widetilde{\boldsymbol{Y}} \\ &= (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{B} + \left\{\widetilde{\boldsymbol{Y}} - \tilde{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\Delta}}\widetilde{\boldsymbol{Y}}\right\}. \end{aligned}$$

Note that *bbeh*, $\boldsymbol{S}$ and $\widetilde{\boldsymbol{Y}}$ (or $\boldsymbol{W}$) are mutually independent and that

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &\sim \mathcal{N}_{p\times q}\left(\boldsymbol{\beta}, r^{-1}\boldsymbol{\Sigma}, (\boldsymbol{B}\boldsymbol{B}')^{-}\right), \\ \widetilde{\boldsymbol{Y}} &\sim \mathcal{N}_{p\times k}\left(\tilde{\boldsymbol{\alpha}}, r^{-1}\boldsymbol{\Sigma}, \boldsymbol{I}_p - \boldsymbol{P}\right), \\ \boldsymbol{S} &\sim \mathcal{W}_p(\boldsymbol{\Sigma}, n), \end{aligned}$$

where $\mathrm{rank}\,\boldsymbol{B}\boldsymbol{B}' = q_1 \le q$ and $\mathrm{rank}\,(\boldsymbol{I}_p - \boldsymbol{P}) = m = k - q_1$. Then the risk function of $\widehat{\boldsymbol{\Theta}}$ in terms of (2.2) is

$$\begin{aligned} R_m(\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}), \omega) &= E_\omega\left[\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{B}\boldsymbol{B}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\right] \\ &\quad + E_\omega\left[\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{Y}} - \tilde{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\Delta}}\widetilde{\boldsymbol{Y}})(\widetilde{\boldsymbol{Y}} - \tilde{\boldsymbol{\alpha}} - \widehat{\boldsymbol{\Delta}}\widetilde{\boldsymbol{Y}})'\right]. \end{aligned}$$

Since $\boldsymbol{I}_p - \boldsymbol{P}$ is idempotent, there exists a $k \times m$ matrix $\boldsymbol{Q}_1$ such that $\boldsymbol{Q}_1'\boldsymbol{Q}_1 = \boldsymbol{I}_m$ and $\boldsymbol{I}_p - \boldsymbol{P} = \boldsymbol{Q}_1\boldsymbol{Q}_1'$. Define $p \times m$ matrices $\boldsymbol{Z}$ and $\boldsymbol{\mu}$ by

$$\begin{aligned}\boldsymbol{Z} &= \sqrt{r}\widetilde{\boldsymbol{Y}}\boldsymbol{Q}_1,\\ \boldsymbol{\mu} &= \sqrt{r}\tilde{\boldsymbol{\alpha}}\boldsymbol{Q}_1.\end{aligned}$$

Then $\boldsymbol{Z} \sim \mathcal{N}_{p\times m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{I}_m)$ and $\boldsymbol{W} = \boldsymbol{Z}\boldsymbol{Z}'$, so that $R_m(\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}), \omega)$ is expressed as

$$\begin{aligned}R_m(\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}), \omega) = &\frac{pq_1}{r}\\ &+\frac{1}{r}E_\omega\left[\mathrm{tr}\left(\boldsymbol{Z} - \boldsymbol{\mu} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{S}, \boldsymbol{Z}\boldsymbol{Z}')\boldsymbol{Z}\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Z} - \boldsymbol{\mu} - \widehat{\boldsymbol{\Delta}}(\boldsymbol{S}, \boldsymbol{Z}\boldsymbol{Z}')\boldsymbol{Z}\right)\right].\end{aligned}$$

When the prior distribution of $\boldsymbol{\mu}$ is supposed as $\pi : \boldsymbol{\mu} \sim \mathcal{N}_{p\times m}(\boldsymbol{0}, r\boldsymbol{\Sigma}_A)$, the Bayes risk of $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}})$ is

$$\begin{aligned}E^\pi\left[R_m(\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}), \omega)\right] &= \frac{pq_1 + pm}{r} - \frac{m}{r}\mathrm{tr}\,\boldsymbol{\Sigma}(\boldsymbol{\Sigma} + r\boldsymbol{\Sigma}_A)^{-1}\\ &\quad+\frac{1}{r}E_\omega\left[\mathrm{tr}\left(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}\right)'\boldsymbol{\Sigma}^{-1}\left(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}\right)\boldsymbol{W}\right]\\ &= \frac{pk}{r} + \frac{1}{r}E^\pi\left[E_\omega\left[\mathrm{tr}\,\widehat{\boldsymbol{\Delta}}'\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Delta}}\boldsymbol{W} - 2\mathrm{tr}\,\widehat{\boldsymbol{\Delta}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}\boldsymbol{W}\right]\right].\end{aligned}$$

If there exists an unbiased estimator $\widehat{R}^*(\boldsymbol{S}, \boldsymbol{Z}\boldsymbol{Z}')$ such that

$$E^\pi\left[E_\omega\left[\widehat{R}^*(\boldsymbol{S}, \boldsymbol{Z}\boldsymbol{Z}')\right]\right] = E^\pi\left[E_\omega\left[\mathrm{tr}\,\widehat{\boldsymbol{\Delta}}'\boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Delta}}\boldsymbol{W} - 2\mathrm{tr}\,\widehat{\boldsymbol{\Delta}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}\boldsymbol{W}\right]\right],$$

the Bayes risk can be represented by

$$E^\pi\left[R_m(\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}), \omega)\right] = E^\pi\left[E_{\Sigma,\mu\mu'}\left[\frac{pk}{r} + \frac{1}{r}\widehat{R}^*(\boldsymbol{S}, \boldsymbol{Z}\boldsymbol{Z}')\right]\right].$$

It is here noted that $E_{\Sigma,\mu\mu'}[\widehat{R}^*(\boldsymbol{S}, \boldsymbol{Z}\boldsymbol{Z}')]$ is a function of $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}\boldsymbol{\mu}'$, and that $\boldsymbol{\mu}\boldsymbol{\mu}'$ has $\mathcal{W}_p(r\boldsymbol{\Sigma}_A, m)$. Since the Wishart distribution is complete, the same arguments as used in Efron and Morris (1996) shows that

$$R_m(\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}), \omega) = E_{\Sigma,\mu\mu'}\left[\frac{pk}{r} + \frac{1}{r}\widehat{R}^*(\boldsymbol{S}, \boldsymbol{Z}\boldsymbol{Z}')\right].$$

Hence the risk function of $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}})$ in the fixed effects model can be derived automatically from the risk of $\widehat{\boldsymbol{\Delta}}$ in the mixed linear model.

Proposition 8. *In the fixed effects model, consider the problem of estimating the unknown matrix of parameters* $\boldsymbol{\Theta} = \boldsymbol{\beta}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k) + (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k)$ *by the estimator* $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}})$ *given by (6.1) relative to the risk (2.2).*

(1) *For the estimator* $\widehat{\boldsymbol{\Delta}}_1(\boldsymbol{\Psi})$ *given by (4.2), the unbiased estimator of the risk of* $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}_1(\boldsymbol{\Psi}))$ *is*

$$\frac{pk}{r} + \frac{1}{r(n+p+1)}\mathrm{tr}\, \boldsymbol{S}\boldsymbol{H}\boldsymbol{\Psi}^*(\boldsymbol{x})\boldsymbol{H}',$$

where $\boldsymbol{\Psi}^*(\boldsymbol{x})$ *is given by (4.4).*

(2) *For the estimator* $\widehat{\boldsymbol{\Delta}}_2(\boldsymbol{\Psi})$ *given by (4.7), the unbiased estimator of the risk of* $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}_2(\boldsymbol{\Psi}))$ *is*

$$\frac{pk}{r} + \frac{(m-p-1)^2}{r}\mathrm{tr}\, \boldsymbol{K}\boldsymbol{\Psi}^*(\boldsymbol{\ell})\boldsymbol{K}'\boldsymbol{W}^{-1},$$

where $\boldsymbol{\Psi}^*(\boldsymbol{\ell})$ *is given by (4.9).*

(3) *For the scale-equivariant estimator* $\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi})$ *given by (4.11), the unbiased estimator of the risk of* $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi}))$ *is*

$$\frac{pk}{r} + \frac{1}{r}r_3(\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi})),$$

where $r_3(\widehat{\boldsymbol{\Delta}}_3(\boldsymbol{\Psi}))$ *is given by (4.12).*

Corollary 8. *In the fixed effects model, the estimators* $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}_i^S)$ *and* $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}_i^{EM})$ *for* $i = 1, 2, 3$ *dominate the estimator* $\widehat{\boldsymbol{\Theta}}_0 = \widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}_0)$ *for the risk (2.2).*

For th ecase of $\boldsymbol{b}_1 = \cdots = \boldsymbol{b}_k = \boldsymbol{0}$, results (1) and (3) of Proposition 8 were given by Konno(1990a,b). However, by using the arguments of Efron and Morris' (1976) approach, we obtain simpler proofs even in the general case.

For the James-Stein type estimators $\widehat{\boldsymbol{\Delta}}_1^{JS}$ and $\widehat{\boldsymbol{\Delta}}_2^{JS}$, unbiased estimators of the risk functions of $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}_1^{JS})$ and $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}_2^{JS})$ can be provided by combining (2.5) and Propositions 1 and 2, so that it can be shown that $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}_1^{JS})$ and $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}_2^{JS})$ dominate $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}_0)$ in the fixed effects model.

7 An Extension of the Model and Remarks

We here consider an extension of the model (2.1) and investigate whether the series of dominance results in the previous sections hold in the extended model.

A simple extension of the model is given by

$$\boldsymbol{y}_{ij} = \boldsymbol{\beta}\boldsymbol{b}_{ij} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, r, \tag{7.1}$$

where $\boldsymbol{b}_{ij}$'s are $q \times 1$ known vectors and the other parameters and constants are the same as defined in (2.1). Then the exponent in the joint distribution of $\boldsymbol{y}_{ij}$'s is written by

$$\sum_{i,j}\left(\boldsymbol{y}_{ij} - \boldsymbol{\beta}\boldsymbol{b}_{ij} - \boldsymbol{\alpha}_i\right)' \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}_{ij} - \boldsymbol{\beta}\boldsymbol{b}_{ij} - \boldsymbol{\alpha}_i\right) + \sum_i \boldsymbol{\alpha}_i'\boldsymbol{\Sigma}_A^{-1}\boldsymbol{\alpha}_i$$

$$
\begin{aligned}
= & \sum_{i,j}\left(\boldsymbol{y}_{ij}-\overline{\boldsymbol{y}}_{i\cdot}-\boldsymbol{\beta}(\boldsymbol{b}_{ij}-\overline{\boldsymbol{b}}_{i\cdot})\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}_{ij}-\overline{\boldsymbol{y}}_{i\cdot}-\boldsymbol{\beta}(\boldsymbol{b}_{ij}-\overline{\boldsymbol{b}}_{i\cdot})\right) \\
& +\sum_{i}\left(\boldsymbol{\theta}_i-\widehat{\boldsymbol{\theta}}_i^B\right)'\left(r\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Sigma}_A^{-1}\right)\left(\boldsymbol{\theta}_i-\widehat{\boldsymbol{\theta}}_i^B\right) \\
& +r\sum_{i}\left(\overline{\boldsymbol{y}}_{i\cdot}-\boldsymbol{\beta}\overline{\boldsymbol{b}}_{i\cdot}\right)'\boldsymbol{\Sigma}_2^{-1}\left(\overline{\boldsymbol{y}}_{i\cdot}-\boldsymbol{\beta}\overline{\boldsymbol{b}}_{i\cdot}\right),
\end{aligned}
$$

where $\boldsymbol{\theta}_i=\boldsymbol{\beta}\overline{\boldsymbol{b}}_{i\cdot}+\boldsymbol{\alpha}_i$ for $\overline{\boldsymbol{b}}_{i\cdot}=r^{-1}\sum_j \boldsymbol{b}_{ij}$ and

$$
\widehat{\boldsymbol{\theta}}_i^B=\overline{\boldsymbol{y}}_{i\cdot}-\boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1}\left(\overline{\boldsymbol{y}}_{i\cdot}-\boldsymbol{\beta}\overline{\boldsymbol{b}}_{i\cdot}\right). \tag{7.2}
$$

Let $\boldsymbol{U}=(\boldsymbol{u}_{11},\ldots,\boldsymbol{u}_{1r};\cdots;\boldsymbol{u}_{k1},\ldots,\boldsymbol{u}_{kr})$ and $\boldsymbol{C}=(\boldsymbol{c}_{11},\ldots,\boldsymbol{c}_{1r};\cdots;\boldsymbol{c}_{k1},\ldots,\boldsymbol{c}_{kr})$ for $\boldsymbol{u}_{ij}=\boldsymbol{y}_{ij}-\overline{\boldsymbol{y}}_{i\cdot}$ and $\boldsymbol{c}_{ij}=\boldsymbol{b}_{ij}-\overline{\boldsymbol{b}}_{i\cdot}$. Then

$$
\begin{aligned}
& \sum_{i,j}\left(\boldsymbol{y}_{ij}-\overline{\boldsymbol{y}}_{i\cdot}-\boldsymbol{\beta}(\boldsymbol{b}_{ij}-\overline{\boldsymbol{b}}_{i\cdot})\right)'\boldsymbol{\Sigma}^{-1}\sum_{i,j}\left(\boldsymbol{y}_{ij}-\overline{\boldsymbol{y}}_{i\cdot}-\boldsymbol{\beta}(\boldsymbol{b}_{ij}-\overline{\boldsymbol{b}}_{i\cdot})\right) \\
= & \operatorname{tr}\boldsymbol{\Sigma}^{-1}(\boldsymbol{U}-\boldsymbol{\beta}\boldsymbol{C})(\boldsymbol{U}-\boldsymbol{\beta}\boldsymbol{C})' \\
= & \operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}+\operatorname{tr}\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\beta}}_1-\boldsymbol{\beta})\boldsymbol{C}\boldsymbol{C}'(\widehat{\boldsymbol{\beta}}_1-\boldsymbol{\beta})',
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_1 &= \boldsymbol{U}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^-, \\
\boldsymbol{S} &= (\boldsymbol{U}-\widehat{\boldsymbol{\beta}}_1\boldsymbol{C})(\boldsymbol{U}-\widehat{\boldsymbol{\beta}}_1\boldsymbol{C})'.
\end{aligned}
$$

Also letting $\overline{\boldsymbol{Y}}=(\overline{\boldsymbol{y}}_{1\cdot},\ldots,\overline{\boldsymbol{y}}_{k\cdot})$ and $\overline{\boldsymbol{B}}=(\overline{\boldsymbol{b}}_{1\cdot},\ldots,\overline{\boldsymbol{b}}_{k\cdot})$, we see that

$$
\begin{aligned}
& r\sum_{i}(\overline{\boldsymbol{y}}_{i\cdot}-\boldsymbol{\beta}\overline{\boldsymbol{b}}_{i\cdot})'\boldsymbol{\Sigma}_2^{-1}(\overline{\boldsymbol{y}}_{i\cdot}-\boldsymbol{\beta}\overline{\boldsymbol{b}}_{i\cdot}) \\
= & r\operatorname{tr}\boldsymbol{\Sigma}_2^{-1}(\overline{\boldsymbol{Y}}-\boldsymbol{\beta}\overline{\boldsymbol{B}})(\overline{\boldsymbol{Y}}-\boldsymbol{\beta}\overline{\boldsymbol{B}})' \\
= & \operatorname{tr}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{W}+r\operatorname{tr}\boldsymbol{\Sigma}_2^{-1}(\widehat{\boldsymbol{\beta}}_2-\boldsymbol{\beta})\overline{\boldsymbol{B}}\,\overline{\boldsymbol{B}}'(\widehat{\boldsymbol{\beta}}_2-\boldsymbol{\beta})',
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_2 &= \overline{\boldsymbol{Y}}\,\overline{\boldsymbol{B}}'(\overline{\boldsymbol{B}}\,\overline{\boldsymbol{B}}')^-, \\
\boldsymbol{W} &= r(\overline{\boldsymbol{Y}}-\widehat{\boldsymbol{\beta}}_2\overline{\boldsymbol{B}})(\overline{\boldsymbol{Y}}-\widehat{\boldsymbol{\beta}}_2\overline{\boldsymbol{B}})'.
\end{aligned}
$$

Assuming that $\operatorname{rank}(\boldsymbol{C}\boldsymbol{C}')=q_1\le q$ and $\operatorname{rank}(\overline{\boldsymbol{B}}\,\overline{\boldsymbol{B}}')=q_2\le q$, we observe that $\boldsymbol{S}$, $\boldsymbol{W}$, $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ are mutually independent and that

$$
\begin{aligned}
\boldsymbol{S} &\sim \mathcal{W}_p(\boldsymbol{\Sigma},n), \quad n=k(r-1)-q_1, \\
\boldsymbol{W} &\sim \mathcal{W}_p(\boldsymbol{\Sigma}_2,m), \quad m=k-q_2, \\
\widehat{\boldsymbol{\beta}}_1 &\sim \mathcal{N}_{p\times q}(\boldsymbol{\beta},\boldsymbol{\Sigma},(\boldsymbol{C}\boldsymbol{C}')^-), \\
\widehat{\boldsymbol{\beta}}_2 &\sim \mathcal{N}_{p\times q}(\boldsymbol{\beta},r^{-1}\boldsymbol{\Sigma}_2,(\overline{\boldsymbol{B}}\,\overline{\boldsymbol{B}}')^-),
\end{aligned}
$$

where $\widehat{\boldsymbol{\beta}}_i$ has the degenerated normal distribution in the case that $q_i < q$ for $i = 1, 2$.

As an empirical Bayes procedure suggested from (7.2), we consider the estimator

$$\widehat{\boldsymbol{\theta}}_i^{EB}(\widehat{\boldsymbol{\beta}}_2) = \overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\Delta}}(\overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\beta}}_2\overline{\boldsymbol{b}}_{i\cdot}), \tag{7.3}$$

where the estimator $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}_2^{-1}$ based on $\boldsymbol{S}$ and $\boldsymbol{W}$ is constructed. The risk of the estimator $\widehat{\boldsymbol{\Theta}}^{EB}(\widehat{\boldsymbol{\beta}}_2) = (\widehat{\boldsymbol{\theta}}_1^{EB}(\widehat{\boldsymbol{\beta}}_2), \ldots, \widehat{\boldsymbol{\theta}}_1^{EB}(\widehat{\boldsymbol{\beta}}_2))$ is given by

$$R_m(\widehat{\boldsymbol{\Theta}}^{EB}(\widehat{\boldsymbol{\beta}}_2), \omega) = r^{-1}E_\omega\left[\mathrm{tr}\,(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})\boldsymbol{W}\right] + r^{-1}(pk - m\mathrm{tr}\,\boldsymbol{\Delta}),$$

so that all the dominance results in the previous sections can be applied to the model (7.1).

It is noted that the regression coefficients $\boldsymbol{\beta}$ has two independent estimators $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ with different covariance matrices. Hence it is natural to consider random weighted combined estimator $\widehat{\boldsymbol{\beta}}$ of the form

$$\begin{aligned}\mathbf{vec}\,(\widehat{\boldsymbol{\beta}}) &= \left[\{(\boldsymbol{C}\boldsymbol{C}')^- \otimes \widehat{\boldsymbol{\Sigma}}\}^- + \{(\overline{\boldsymbol{B}}\,\overline{\boldsymbol{B}}')^- \otimes r^{-1}\widehat{\boldsymbol{\Sigma}}_2\}^-\right]^- \\ &\quad \times \left[\{(\boldsymbol{C}\boldsymbol{C}')^- \otimes \widehat{\boldsymbol{\Sigma}}\}^-\mathbf{vec}\,(\widehat{\boldsymbol{\beta}}_1) + \{(\overline{\boldsymbol{B}}\,\overline{\boldsymbol{B}}')^- \otimes r^{-1}\widehat{\boldsymbol{\Sigma}}_2\}^-\mathbf{vec}\,(\widehat{\boldsymbol{\beta}}_2)\right],\end{aligned}$$

where $\mathbf{vec}\,(\boldsymbol{U}) = (\boldsymbol{u}_1', \ldots, \boldsymbol{u}_q')'$ for $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_q)$ and $\otimes$ denotes the Kronecker product. However it is difficult to study any exact dominance property for the combined estimator $\widehat{\boldsymbol{\beta}}$.

A practically appealing model may be the case with unequal replications:

$$\boldsymbol{y}_{ij} = \boldsymbol{\beta}\boldsymbol{b}_{ij} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i, \tag{7.4}$$

which was discussed by Fuller and Harter (1987) for estimation of small area. It seems, however, intractable to establish exact dominance results in the model (7.4). The works of deriving efficient estimators by using approximation (asymptotic) theories rests in the future.

Acknowledgments. The research of the first author was supported in part by a grant from the Center for International Research on the Japanese Economy, the University of Tokyo, and by the Ministry of Education, Japan, Grant No. 11680320. The research of the second author was supported in part by Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Amemiya, Y. (1994). On multivariate mixed model analysis. In: *Multivariate Analysis and Its Applications*, IMS Lecture Notes - Monograph Series **24**.

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.*, **83**, 28-36.

Berger, J.O.(1985). *Statistical Decision Theory and Bayesian Analysis.* 2nd Ed., Springer-Verlag, New York.

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

Bilodeau, M. and Srivastava, M.S. (1992). Estimation of the eigenvalues of $\Sigma_1\Sigma_2^{-1}$. *J. Multivariate Anal.*, **41**, 1-13.

Brandwein, A.C. and Strawderman, W.E. (1990). Stein estimation: The spherically symmetric case. *Statist. Science*, **5**, 356-369.

Casella, G. (1985). An introduction to empirical Bayes data analysis. *Amer. Statist.*, **39**, 83-87.

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

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. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.*, **70**, 311-319.

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

Fay, R.E. and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269-277.

Fuller and Harter (1987). The multivariate components of variance model for small area estimation. In: *Small Area Statistics* (R. Platek, J.N.K. Rao, C.E. Sarndal and M.P. Singh, eds.), John Wiley and Sons, New York.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statist. Science*, **9**, 55-93.

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

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

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

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, **1**, 361-379. University of California Press, Berkeley.

Kariya, T., Konno, Y. and Strawderman, W.E. (1996). Double shrinkage estimators in the GMANOVA model. *J. Multivariate Anal.*, **56**, 245-258.

Kariya, T., Konno, Y. and Strawderman, W.E. (1999). Construction of shrinkage estimators for the regression coefficient matrix in the GMANOVA model. *Commun. Statist. - Theory Methods*, **28**, 597-611.

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. (1992a). *Improved estimation of matrix of normal mean and eigenvalues in the multivariate F - distribution.* Doctoral Dissertation, Institute of Mathematics, University of Tsukuba.

Konno, Y. (1992b). On estimating eigenvalues of the scale matrix of the multivariate F distribution. *Sankhya* (Ser. A), **54**, 241-251.

Kubokawa, T. (1998). The Stein phenomenon in simultaneous estimation: A review. *Applied Statistical Science III* (eds. S.E. Ahmed, M. Ahsanullah and B.K. Sinha), 143-173, NOVA Science Publishers, Inc., New York.

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

Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation.* 2nd Ed., Springer-Verlag, New York.

Loh, W.-L. (1988). *Estimating covariance matrices.* Ph.D. theses. Stanford University.

Nickerson, D.M. and Basawa, I.V. (1992). Shrinkage estimation for linear models with correlated errors. *Sankhya* (Ser. A), **54**, 411-424.

Peixoto, J.L. and Harville, D.A.(1986). Comparisons of alternative predictors under the balanced one-way random model. *J. Amer. Statist. Assoc.*, **81**, 431-436.

Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.*, **85**, 163-171.

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

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proc. Third Berkeley Symp. Math. Statist. Probab.*, **1**, 197-206. University of California University, Berkeley.

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

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

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

Zheng, Z. (1986). On estimation of matrix of normal mean. *J. Multivariate Anal.*, **18**, 70-82.

Zidek, J. (1978). Deriving unbiased risk estimators of multinormal mean and regression coefficient estimators using zonal polynomials. *Ann. Statist.*, **6**, 769-782.