CIRJE-F-180

Prediction in Multivariate Mixed Linear Models

Tatsuya Kubokawa
University of Tokyo

M. S. Srivastava
University of Toronto

October 2002

Prediction in Multivariate Mixed Linear Models

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

October 30, 2002

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. We propose scale invariant Stein type shrinkage estimators for the ratio of the two covariance matrices. 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.

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, posted land price data.

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

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 applications and theoretical studies of these models have been given by Fay and Herriot (1979), Battese, Harter and Fuller (1988), Prasad and Rao (1990), Ghosh and Rao (1994) and references therein.

In contrast to these activities in the univariate mixed linear models, the multivariate mixed linear models have received little attention except in the multiple regression model with replicates considered by Rao (1975) in which the vector of p response variables $\boldsymbol{y}_i$ follow the model

$$\boldsymbol{y}_i = \boldsymbol{X}\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_i, \tag{1.1}$$

$i = 1, \ldots, k$ where $\boldsymbol{X}$ is a $p \times m$ matrix of known constants, $\boldsymbol{\gamma}$ is an m-vector of unknown parameters and, $\boldsymbol{\eta}_i$ and $\boldsymbol{\epsilon}_i$ are independently distributed with $\boldsymbol{\eta}_i$ i.i.d. as $\mathcal{N}_m(\mathbf{0}, \boldsymbol{F})$ and $\boldsymbol{\epsilon}_i$ i.i.d. as $\mathcal{N}_p(\mathbf{0}, \sigma^2\boldsymbol{I})$. Reinsel (1985) extended this model in which

$$\boldsymbol{\gamma}_i = \boldsymbol{\beta}\boldsymbol{b}_i + \boldsymbol{\eta}_i \tag{1.2}$$

where $\boldsymbol{\beta}$ is an $m \times q$ matrix of unknown parameters and $\boldsymbol{b}_i$ are vectors of known constants. Battese *et al*. (1988)'s model is a variation of the model given in (1.1).

In both models (1.1) and (1.2), $\boldsymbol{y}_i$ are independently distributed random vectors with covariance matrix given by

$$\mathbf{Cov}\,(\boldsymbol{y}_i) = \boldsymbol{X}\boldsymbol{F}\boldsymbol{X}' + \sigma^2\boldsymbol{I}_p.$$

Thus, when $m < p$, both σ^2 and $\boldsymbol{F}$ can be unbiasedly estimated. However, when

$$\mathbf{Cov}\,(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma},$$

a completely unknown matrix, no optimality results are available in predicting, say, $\boldsymbol{\beta}\boldsymbol{b}_i + \boldsymbol{\eta}_i$ when $\boldsymbol{\eta}_i$ are random or when $\boldsymbol{\eta}_i$ are fixed (with some constraints). Thus, in this paper, we consider the following multivariate mixed linear models

$$\boldsymbol{y}_{ij} = \boldsymbol{\beta}\boldsymbol{b}_{ij} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}, \tag{1.3}$$

$i = 1, \ldots, k$, $j = 1, \ldots, r$, where $\boldsymbol{\beta}$ is a $p \times q$ matrix of unknown parameters and $\boldsymbol{b}_{ij}$ are $q \times 1$ known vectors. It is assumed that $\boldsymbol{\alpha}_i$ and $\boldsymbol{\epsilon}_{ij}$ are independently distributed where $\boldsymbol{\alpha}_i$ are i.i.d. $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_A)$ and $\boldsymbol{\epsilon}_{ij}$ are i.i.d. $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$. Let

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

and

$$\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k).$$

Our objective is to predict $\boldsymbol{\Theta}$ under the squared loss function where $\widehat{\boldsymbol{\Theta}}$ is chosen to minimize the risk

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

where $\omega = (\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_A)$ denotes the set of unknown parameters.

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 Section 2, we consider the case when $\boldsymbol{b}_{ij} = \boldsymbol{b}_i$, $j = 1, \dots, r$. It is shown that the problem of predicting small area means is reduced to that of estimating the ratio $\boldsymbol{\Delta}$ of covariance matrices. 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.

In the estimation of a covariance matrix and a ratio of two 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. We therefore consider scale invariant Stein-type shrinkage estimator for $\boldsymbol{\Delta}$ in Section 3. By taking into account the order restriction between the two covariance matrices, another improved predictor is obtained in Section 3. It may be mentioned that an estimator of $\boldsymbol{\Delta}$ can also be obtained by considering multivariate beta distribution as in Bilodeau and Srivastava (1992) and Konno (1992b).

In Section 4, numerical comparisons of several improved predictors are given based on Monte Carlo simulation. The Monte Carlo studies show that the Efron-Morris type truncated predictor proposed in Section 3 has very nice risk-performances with much smaller risks than the sample mean vector. An example treating data derived by Monte Carlo simulation is given there.

The problem of estimating a matrix mean in a fixed effect linear regression model is addressed in Section 5. In Section 6, we consider the general model (1.3) and give an example analyzing the posted land price data in Kanagawa prefecture in Japan.

2 A Mixed Linear Model

In this section, 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

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

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

The multivariate mixed linear model is also interpreted as an empirical Bayes model: $\boldsymbol{y}_{ij} \sim \mathcal{N}_p(\boldsymbol{\theta}_i, \boldsymbol{\Sigma})$ 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

$$\widehat{\boldsymbol{\theta}}_i^B = \widehat{\boldsymbol{\theta}}_i^B(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_A) = \overline{\boldsymbol{y}}_{i\cdot} - \boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1}(\overline{\boldsymbol{y}}_{i\cdot} - \boldsymbol{\beta}\boldsymbol{b}_i),$$

for $\overline{\boldsymbol{y}}_{i\cdot} = r^{-1}\sum_{j=1}^r \boldsymbol{y}_{ij}$ and $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma} + r\boldsymbol{\Sigma}_A$. 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 distributions. Let

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

where $\boldsymbol{A}^-$ denotes the generalized inverse of the 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

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

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 respectively estimated by $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Sigma}}_A$ (or $\widehat{\boldsymbol{\Sigma}}_2$) which are functions of $\widehat{\boldsymbol{\beta}}$, $\boldsymbol{S}$ and $\boldsymbol{W}$. This results in an empirical Bayes estimator (or feasible BLUP) of $\boldsymbol{\theta}_i$ of the form

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

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 $\boldsymbol{0} \le \widehat{\boldsymbol{\Sigma}}^{1/2}\widehat{\boldsymbol{\Sigma}}_2^{-1}\widehat{\boldsymbol{\Sigma}}^{1/2} \le \boldsymbol{I}_p$, where $\boldsymbol{A}^{1/2}$ denotes the positive definite factorization 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 $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{0}$ 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 the existing estimators including $\overline{\boldsymbol{y}}_{i\cdot}$.

Let $\widehat{\mathbf{\Theta}}^{EB} = (\widehat{\boldsymbol{\theta}}_1^{EB}, \ldots, \widehat{\boldsymbol{\theta}}_k^{EB})$. Noting that $E_\omega[\mathrm{tr}\,(\widehat{\mathbf{\Theta}}^B - \mathbf{\Theta})'\mathbf{\Sigma}^{-1}(\widehat{\mathbf{\Theta}}^B - \mathbf{\Theta})] = pk/r - (k/r)\mathrm{tr}\,\mathbf{\Sigma}\mathbf{\Sigma}_2^{-1}$, we see that the risk function of $\widehat{\mathbf{\Theta}}^{EB}$ is written by

$$
\begin{aligned}
&R_m(\widehat{\mathbf{\Theta}}^{EB}, \omega) \qquad (2.5)\\
&\quad = E_\omega\left[\mathrm{tr}\,(\widehat{\mathbf{\Theta}}^{EB} - \widehat{\mathbf{\Theta}}^B)'\mathbf{\Sigma}^{-1}(\widehat{\mathbf{\Theta}}^{EB} - \widehat{\mathbf{\Theta}}^B)\right] + E_\omega\left[\mathrm{tr}\,(\widehat{\mathbf{\Theta}}^B - \mathbf{\Theta})'\mathbf{\Sigma}^{-1}(\widehat{\mathbf{\Theta}}^B - \mathbf{\Theta})\right]\\
&\quad = r^{-1}E_\omega\left[\mathrm{tr}\,\left\{(\widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Sigma}}_2^{-1} - \mathbf{\Sigma}\mathbf{\Sigma}_2^{-1})'\mathbf{\Sigma}^{-1}(\widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Sigma}}_2^{-1} - \mathbf{\Sigma}\mathbf{\Sigma}_2^{-1})\boldsymbol{W}\right\}\right] + kr^{-1}\left(p - \mathrm{tr}\,\mathbf{\Sigma}\mathbf{\Sigma}_2^{-1}\right),
\end{aligned}
$$

so that the risk of $\widehat{\mathbf{\Theta}}^{EB}$ depends on the risk of the estimators of the ratio of covariance matrices defined by

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

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

The usual unbiased estimator of $\mathbf{\Delta}$ is $\widehat{\mathbf{\Delta}}^{UB} = n^{-1}(m-p-1)\boldsymbol{S}\boldsymbol{W}^{-1}$ with risk

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

Since the crude estimator $(\overline{\boldsymbol{y}}_1., \ldots, \overline{\boldsymbol{y}}_k.)$ of $\mathbf{\Theta}$ corresponds to the the estimator $\widehat{\mathbf{\Delta}} = \mathbf{0}$, its risk has $R(\mathbf{0}, \omega) = m\mathrm{tr}\,\mathbf{\Delta}$ and this shows that $\widehat{\mathbf{\Delta}}^{UB}$ is better than the crude estimator $\widehat{\mathbf{\Delta}} = \mathbf{0}$ for $m > p+1$ and $n > p+1$. More generally the estimator $\widehat{\mathbf{\Delta}}(a) = a\boldsymbol{S}\boldsymbol{W}^{-1}$, a multiple of $\boldsymbol{S}\boldsymbol{W}^{-1}$, has the risk $R(\widehat{\mathbf{\Delta}}(a), \omega) = \{n(n+p+1)(m-p-1)^{-1}a^2 - 2na + m\}\,\mathrm{tr}\,\mathbf{\Delta}$, which can be minimized at $a = (m-p-1)/(n+p+1)$ with risk

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

where

$$
\widehat{\mathbf{\Delta}}_0 = (n+p+1)^{-1}(m-p-1)\boldsymbol{S}\boldsymbol{W}^{-1}. \qquad (2.7)
$$

3 Improved Estimators

In this section, we shall provide several types of estimators of $\mathbf{\Delta}$ improving upon $\widehat{\mathbf{\Delta}}_0$ relative to the risk $R(\widehat{\mathbf{\Delta}}, \omega)$.

We first consider two types of James-Stein type estimators based on Bartlett's decomposition. 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 $\boldsymbol{S} = \boldsymbol{T}\boldsymbol{T}'$ and $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{U}'$. Then we can consider two types of James-Stein estimators:

$$
\widehat{\mathbf{\Delta}}_1^{JS} = (n+p+1)^{-1}\boldsymbol{S}\boldsymbol{U}'^{-1}\boldsymbol{C}\boldsymbol{U}^{-1} \quad \text{and} \quad \widehat{\mathbf{\Delta}}_2^{JS} = (m-p-1)\boldsymbol{T}\boldsymbol{D}\boldsymbol{T}'\boldsymbol{W}^{-1},
$$

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

$$c_i = m - i - 1, \qquad i = 1, \ldots, p, \tag{3.1}$$

$$d_i = \frac{1}{n+p+3-2i} \prod_{j=1}^{i-1} \left(1 - \frac{1}{n+p+3-2j}\right), \quad i = 2, \ldots, p, \tag{3.2}$$

and $d_1 = (n+p+1)^{-1}$. It has been shown in Kubokawa and Srivastava (1999) that these estimators have smaller risk than the estimator $\widehat{\boldsymbol{\Delta}}_0$. However, these estimators depend on the coordinate systems and therefore no further consideration will be given to these estimators in this paper. Instead, we shall consider scale equivariant estimators as described below.

Let $\boldsymbol{A}$ be a $p \times p$ nonsingular matrix such that $\boldsymbol{S} = \boldsymbol{A}\boldsymbol{A}'$ and $\boldsymbol{W} = \boldsymbol{A}\boldsymbol{F}\boldsymbol{A}'$ for $\boldsymbol{F} = \mathrm{diag}\,(f_1, \ldots, f_p)$, $f_1 \geq f_2 \geq \ldots \geq f_p$. Then we consider estimators of the form

$$\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}) = \boldsymbol{A}\boldsymbol{\Psi}(\boldsymbol{f})\boldsymbol{A}^{-1}, \quad \boldsymbol{\Psi}(\boldsymbol{f}) = \mathrm{diag}\ (\psi_1(\boldsymbol{f}), \ldots, \psi_p(\boldsymbol{f})) \tag{3.3}$$

for positive functions $\psi_i(\boldsymbol{f})$'s of $\boldsymbol{f} = (f_1, \ldots, f_p)$. 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 $\boldsymbol{B}$. Using the same arguments as in Konno (1991, 92a), we can write the risk function of $\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi})$.

Proposition 1. *The risk function of* $\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi})$ *is given by*

$$R(\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}), \omega) = E_\omega \left[r(\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi})) \right] + m \mathrm{tr}\, \boldsymbol{\Lambda}, \tag{3.4}$$

where

$$\begin{aligned} r(\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi})) = & \sum_{i=1}^{p} \Bigg\{ (n+p-3) f_i \psi_i^2 - 4 f_i^2 \psi_i \frac{\partial \psi_i}{\partial f_i} - 2 \sum_{j>i} \frac{f_i^2 \psi_i^2 - f_j^2 \psi_j^2}{f_i - f_j} \\ & -2(m-p+1)\psi_i - 4 f_i \frac{\partial \psi_i}{\partial f_i} - 4 \sum_{j>i} \frac{f_i \psi_i - f_j \psi_j}{f_i - f_j} \Bigg\}. \end{aligned} \tag{3.5}$$

Proof. The method of proof is similar to Konno (1992a), using 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 $d_{ia} = 2^{-1}(1 + \delta_{ia}) \partial / \partial s_{ia}$ with $\delta_{ia} = 1$ for $i = a$ and $\delta_{ia} = 0$ for $i \neq a$. When $\boldsymbol{S}$ is distributed as $\mathcal{W}_p(\boldsymbol{\Sigma}, n)$, the Stein-Haff identity is expressed by $E_\Sigma[\mathrm{tr}\,\{\boldsymbol{G}(\boldsymbol{S})\boldsymbol{\Sigma}^{-1}\}] = E_\Sigma[(n-p-1)\mathrm{tr}\,\{\boldsymbol{G}(\boldsymbol{S})\boldsymbol{S}^{-1}\} + 2\mathrm{tr}\,\{\boldsymbol{D}_S \boldsymbol{G}(\boldsymbol{S})\}]$. 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

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

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

$$\begin{aligned}\boldsymbol{R}^*(\widehat{\mathbf{\Delta}}(\mathbf{\Psi}), \omega) =& E_\omega \left[(n-p-1)\mathrm{tr}\, \mathbf{\Psi}\boldsymbol{F}\mathbf{\Psi} + 2\mathrm{tr}\, \boldsymbol{D}_S(\boldsymbol{A}\mathbf{\Psi}\boldsymbol{F}\mathbf{\Psi}\boldsymbol{A}')\right. \\ & \left. -2(m-p-1)\mathrm{tr}\, \mathbf{\Psi} - 4\mathrm{tr}\, \boldsymbol{D}_W(\boldsymbol{A}\mathbf{\Psi}\boldsymbol{F}\boldsymbol{A}')\right]. \end{aligned} \tag{3.6}$$

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

$$\mathrm{tr}\, \boldsymbol{D}_S(\boldsymbol{A}\mathbf{\Phi}(\boldsymbol{F})\boldsymbol{A}') = \sum_{i=1}^{p} \left\{ p\phi_i - f_i \frac{\partial \phi_i}{\partial f_i} - \sum_{j>i} \frac{f_i\phi_i - f_j\phi_j}{f_i - f_j} \right\}, \tag{3.7}$$

$$\mathrm{tr}\, \boldsymbol{D}_W(\boldsymbol{A}\mathbf{\Phi}(\boldsymbol{F})\boldsymbol{A}') = \sum_{i=1}^{p} \left\{ \frac{\partial \phi_i}{\partial f_i} + \sum_{j>i} \frac{\phi_i - \phi_j}{f_i - f_j} \right\}, \tag{3.8}$$

where $\mathbf{\Phi}(\boldsymbol{\ell}) = \mathrm{diag}\,(\phi_1(\boldsymbol{\ell}), \ldots, \phi_p(\boldsymbol{\ell}))$. Combining (3.6), (3.7) and (3.8) provides that

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

which leads to the expression (3.5) of Proposition 1. ■

In the next two subsections, we consider two scale invariant estimators which we call Stein type and Efron-Morris type of estimators.

3.1 Stein type scale-equivariant estimator

In this subsection, we consider the Stein type scale-equivariant estimator of the form

$$\widehat{\mathbf{\Delta}}^{ST} = \boldsymbol{A}\mathrm{diag}\,(b_1/f_1, \ldots, b_p/f_p)\,\boldsymbol{A}^{-1}, \tag{3.9}$$

for $b_i = (m+p-2i-1)/(n-p+2i+1)$. These constants b_i's were given by Konno (1991) in estimation of a matrix mean, and have the order relation that $b_1 > b_2 > \ldots > b_p$. It has been shown by Srivastava and Solankey (2002) that this estimator has a very good risk performance as compared to many of its competitors including the ones proposed by Breiman and Friedman (1997). Using (3.4) and (3.5), we shall show that the estimator $\widehat{\mathbf{\Delta}}^{ST}$ dominates $\widehat{\mathbf{\Delta}}_0$ relative to the risk (2.6).

From (3.4) and (3.5), the risk function of the estimator $\widehat{\mathbf{\Delta}}^{ST}$ is given by

$$\begin{aligned} & R(\widehat{\mathbf{\Delta}}^{ST}, \omega) - m\mathrm{tr}\,\mathbf{\Lambda} \\ & = \sum_{i=1}^{p} E_\omega\left[(n+p+1)\frac{b_i^2}{f_i} - 2\sum_{j>i}\frac{b_i^2 - b_j^2}{f_i - f_j} - 2(m-p-1)\frac{b_i}{f_i} - 4\sum_{j>i}\frac{b_i - b_j}{f_i - f_j}\right]. \end{aligned} \tag{3.10}$$

Following Konno (1991), we have that for $k = 1, 2$ and $b_i > b_j$ for $i < j$,

$$\sum_{i=1}^{p}\sum_{j=i+1}^{p}\frac{b_i^k - b_j^k}{f_i - f_j} = \sum_{i=1}^{p}\frac{1}{f_i}\sum_{j=i+1}^{p}\frac{f_i}{f_i - f_j}\left(b_i^k - b_j^k\right) \tag{3.11}$$
$$\geq \sum_{i=1}^{p}\frac{1}{f_i}\sum_{j=i+1}^{p}\left(b_i^k - b_j^k\right) = \sum_{i=1}^{p}\frac{1}{f_i}\left\{(p-i)b_i^k - \sum_{j=i+1}^{p} b_j^k\right\},$$

since $f_i/(f_i - f_j) > 1$. On the other hand, the risk of the estimator $\widehat{\mathbf{\Delta}}_0$ is derived by putting $b_1 = \cdots = b_p = (m-p-1)/(n+p+1)$ in (3.10), and we observe that

$$R(\widehat{\mathbf{\Delta}}_0, \omega) - m\mathrm{tr}\,\mathbf{\Lambda} = -\frac{(m-p-1)^2}{n+p+1}\sum_{i=1}^{p} E_\omega\left[\frac{1}{f_i}\right]. \tag{3.12}$$

Combining (3.10), (3.11) and (3.12), we see that the risk difference of $\widehat{\mathbf{\Delta}}^{ST}$ and $\widehat{\mathbf{\Delta}}_0$ is evaluated as $R(\widehat{\mathbf{\Delta}}^{ST}, \omega) - R(\widehat{\mathbf{\Delta}}_0, \omega) = \sum_{i=1}^{p} E_\omega\left[f_i^{-1} h(i)\right]$, 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

$$h(i) \leq 0 \tag{3.13}$$

for $i = 1, \ldots, p-1$, since it is easy to verify that $h(p) \leq 0$. Although this proof was given by Konno (1991), a simple proof is given below. Note that $h(i-1)$ is rewritten as

$$\begin{aligned} h(i-1) &= -\frac{(m+p-2i+1)^2}{n-p+2i-1} + 2\sum_{j=i}^{p} b_j(b_j+2) + \frac{(m-p-1)^2}{n+p+1} \\ &= h(i) - \frac{(a_i'+2)^2}{a_i - 2} + 2\frac{a_i'^2}{a_i^2} + 4\frac{a_i'}{a_i} + \frac{a_i'^2}{a_i} \\ &= \sum_{k=i}^{p}\left\{-\frac{(a_k'+2)^2}{a_k - 2} + 2\frac{a_k'^2}{a_k^2} + 4\frac{a_k'}{a_k} + \frac{a_k'^2}{a_k}\right\}, \end{aligned} \tag{3.14}$$

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

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

which shows the inequality (3.13).

Proposition 2. *The scale-equivariant Stein type estimator* $\widehat{\mathbf{\Delta}}^{ST}$ *dominates* $\widehat{\mathbf{\Delta}}_0$ *relative to the risk (2.6).*

3.2 Efron-Morris type scale-equivariant estimator

Another scale-equivariant estimator we call Efron-Morris type is given by

$$\widehat{\boldsymbol{\Delta}}^{EM} = \alpha \boldsymbol{S}\boldsymbol{W}^{-1} + \beta \frac{1}{\mathrm{tr}\, \boldsymbol{S}^{-1}\boldsymbol{W}} \boldsymbol{I}_p, \tag{3.15}$$

where

$$\alpha = \frac{m-p-1}{n+p+1}, \qquad \beta = \frac{(p-1)(p+2)(1+\alpha)}{n-p+3} = \frac{(p-1)(p+2)(m+n)}{(n+p+1)(n-p+3)}.$$

That is, $\widehat{\boldsymbol{\Delta}}^{EM} = \boldsymbol{A}\boldsymbol{\Psi}(\boldsymbol{f})\boldsymbol{A}^{-1}$ where $\boldsymbol{\Psi}(\boldsymbol{f}) = \mathrm{diag}\,(\psi_1, \dots, \psi_p)$ and $\psi_i = \alpha/f_i + \beta/\sum_{j=1}^p f_j$. Thus, from (3.5), it follows that the function $r(\cdot)$ for $\widehat{\boldsymbol{\Delta}}^{EM}$ is given by

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

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

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

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

Proposition 3. *The Efron-Morris type scale equivariant estimator* $\widehat{\boldsymbol{\Delta}}^{EM}$ *dominates* $\widehat{\boldsymbol{\Delta}}_0$ *relative to the risk (2.6).*

Konno (1992a) proposed the constant $\beta^* = (p-1)(p+2)/(n-p+3)$ instead of β and we denote the Efron-Morris type estimator for the constant β^* by $\widehat{\boldsymbol{\Delta}}(\beta^*)$. Noting that $\beta^* < \beta$ and that $\sum_i f_i^2 \le (\mathrm{tr}\, \boldsymbol{F})^2$, we can see that $\widehat{\boldsymbol{\Delta}}^{EM}$ given by (3.15) improves on $\boldsymbol{\Delta}^{EM}(\beta^*)$, which is better than $\widehat{\boldsymbol{\Delta}}_0$.

3.3 Improvement 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 $m = 1$.

For the scale-equivariant estimator $\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}) = \boldsymbol{A}\boldsymbol{\Psi}(\boldsymbol{f})\boldsymbol{A}^{-1}$ with

$$\boldsymbol{\Psi}(\boldsymbol{f}) = \text{diag}\,(\psi_1(\boldsymbol{f}), \dots, \psi_p(\boldsymbol{f})),$$

given by (3.3), consider the truncated estimator of the form

$$\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}^*) = \boldsymbol{A}\boldsymbol{\Psi}^*(\boldsymbol{f})\boldsymbol{A}^{-1}, \quad \boldsymbol{\Psi}^*(\boldsymbol{f}) = \text{diag}\,(\min\{\psi_1(\boldsymbol{f}), 1\}, \dots, \min\{\psi_p(\boldsymbol{f}), 1\})\,.$$

Then we can get

Proposition 4. *The truncated estimator $\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}^*)$ dominates $\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi})$ relative to the risk (2.6).*

Proof. The risk difference of the estimators $\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi})$ and $\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}^*)$ with $\boldsymbol{\Psi} = \text{diag}\,(\psi_1, \dots, \psi_p)$ and $\boldsymbol{\Psi}^* = \text{diag}\,(\psi_1^*, \dots, \psi_p^*)$ is written by

$$\begin{aligned} & R(\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}), \omega) - R(\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}^*), \omega) \\ & \quad = E\left[\text{tr}\,\left(\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}) - \widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}^*)\right)' \boldsymbol{\Sigma}^{-1}\left(\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}) + \widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}^*) - 2\boldsymbol{\Delta}\right)\boldsymbol{W}\right] \\ & \quad = E\left[\text{tr}\,\left(\boldsymbol{\Psi} + \boldsymbol{\Psi}^* - 2\boldsymbol{A}^{-1}\boldsymbol{\Delta}\boldsymbol{A}\right)\boldsymbol{F}\left(\boldsymbol{\Psi} - \boldsymbol{\Psi}^*\right)\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}'^{-1}\right]. \end{aligned} \tag{3.16}$$

Since $\boldsymbol{\Delta} \leq \boldsymbol{I}$ and $\boldsymbol{\Psi} \geq \boldsymbol{\Psi}^*$, it can be seen that the r.h.s. of the last equation in (3.16) is greater than or equal to $E[2\text{tr}\,(\boldsymbol{\Psi}^* - \boldsymbol{I})\boldsymbol{F}(\boldsymbol{\Psi} - \boldsymbol{\Psi}^*)\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}]$, which is equal to zero since $\psi_i^* = \min(\psi, 1)$. This proves Proposition 4. ■

Applying the truncation rule to the Stein and Efron-Morris type scale-equivariant estimators $\widehat{\boldsymbol{\Delta}}^{ST}$ and $\widehat{\boldsymbol{\Delta}}^{EM}$, we obtain the truncated estimators

$$\widehat{\boldsymbol{\Delta}}^{ST*} = \widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}^{ST*}), \quad \boldsymbol{\Psi}^{ST*} = \text{diag}\,\left(\min\left\{\frac{b_i}{f_i}, 1\right\}, i = 1, \dots, p\right), \tag{3.17}$$

$$\widehat{\boldsymbol{\Delta}}^{EM*} = \widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}^{EM*}), \quad \boldsymbol{\Psi}^{EM*} = \text{diag}\,\left(\min\left\{\frac{\alpha}{f_i} + \frac{\beta}{\sum_{j=1}^p 1/f_j}, 1\right\}, i = 1, \dots, p\right), \tag{3.18}$$

which improve on $\widehat{\boldsymbol{\Delta}}^{ST}$ and $\widehat{\boldsymbol{\Delta}}^{EM}$.

4 Numerical Risk Compariasons and Empirical Studies

We shall investigate the risk-performances of estimators of $\boldsymbol{\Theta}$ numerically. From (2.3) and (2.5), the risk function of the estimator $\widehat{\boldsymbol{\Theta}}^{EB}$ given by (2.3) is expressed by the risk given by (2.6), that is, the problem is reduced to that of estimating the ratio of the covariance matrices $\boldsymbol{\Delta} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1}$. Since the estimator $(\overline{\boldsymbol{y}}_{1\cdot}, \dots, \overline{\boldsymbol{y}}_{k\cdot})$ of $\boldsymbol{\Theta}$ is minimax in terms of the risk (2.2) and it corresponds to the estimator $\widehat{\boldsymbol{\Delta}} = \boldsymbol{0}$ in the reduced problem with the risk $R(\boldsymbol{0}, \omega) = m\mathrm{tr}\,\boldsymbol{\Delta}$, we use the relative risk efficiency

$$Eff(\widehat{\boldsymbol{\Delta}}, \omega) = E_\omega \left[\mathrm{tr}\,(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})'\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})\boldsymbol{W} \right] / (m\mathrm{tr}\,\boldsymbol{\Delta}) \tag{4.1}$$

to investigate the performances of the estimators estimators $\widehat{\boldsymbol{\Delta}}$ of $\boldsymbol{\Delta}$. When the efficiency $Eff(\widehat{\boldsymbol{\Delta}}, \omega)$ is less than or equal to 1, it means that the estimator $\widehat{\boldsymbol{\theta}}_i^{EB} = \overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\Delta}}(\overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\beta}}\boldsymbol{b}_i)$, $i = 1, \dots, k$, with the $\widehat{\boldsymbol{\Delta}}$ is minimax, namely, it improves on $(\overline{\boldsymbol{y}}_{1\cdot}, \dots, \overline{\boldsymbol{y}}_{k\cdot})$ in terms of the risk (2.2).

The estimators we want to compare are the unbiased estimator $\widehat{\boldsymbol{\Delta}}^{UB} = n^{-1}(m - p - 1)\boldsymbol{S}\boldsymbol{W}^{-1}$, the James-Stein type estimators $\widehat{\boldsymbol{\Delta}}_1^{JS} = (n + p + 1)^{-1}\boldsymbol{S}\boldsymbol{U}'^{-1}\boldsymbol{C}\boldsymbol{U}^{-1}$ and $\widehat{\boldsymbol{\Delta}}_2^{JS} = (m - p - 1)\boldsymbol{T}\boldsymbol{D}\boldsymbol{T}'\boldsymbol{W}^{-1}$ given in the begining of Section 3, the best scale multiple estimator $\widehat{\boldsymbol{\Delta}}_0 = (n + p + 1)^{-1}(m - p - 1)\boldsymbol{S}\boldsymbol{W}^{-1}$, the Stein type estimator

$$\widehat{\boldsymbol{\Delta}}^{ST} = \boldsymbol{A}\mathrm{diag}\,(b_1/f_1, \dots, b_p/f_p)\,\boldsymbol{A}^{-1}$$

given by (3.9), the Efron-Morris type estimator

$$\widehat{\boldsymbol{\Delta}}^{EM} = \alpha\boldsymbol{S}\boldsymbol{W}^{-1} + \beta\frac{1}{\mathrm{tr}\,\boldsymbol{S}^{-1}\boldsymbol{W}}\boldsymbol{I}_p$$

given by (3.15), the other Efron-Morris type estimator

$$\widehat{\boldsymbol{\Delta}}^{EMK} = \alpha\boldsymbol{S}\boldsymbol{W}^{-1} + \beta^*\frac{1}{\mathrm{tr}\,\boldsymbol{S}^{-1}\boldsymbol{W}}\boldsymbol{I}_p$$

for the constant β^* given by Konno (1992a), the truncated estimator $\widehat{\boldsymbol{\Delta}}^{BE*} = \widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}^{BE*})$ for

$$\boldsymbol{\Psi}^{BE*} = \mathrm{diag}\,\left(\min\left\{(n + p + 1)^{-1}(m - p - 1)/f_i, 1\right\}, i = 1, \dots, p\right),$$

the Stein type truncated estimator $\widehat{\boldsymbol{\Delta}}^{ST*} = \widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}^{ST*})$ for

$$\boldsymbol{\Psi}^{ST*} = \mathrm{diag}\,\left(\min\left\{b_i/f_i, 1\right\}, i = 1, \dots, p\right)$$

given by (3.17) and the Efron-Morris type truncated estimator $\widehat{\boldsymbol{\Delta}}^{EM*} = \widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}^{EM*})$ for

$$\boldsymbol{\Psi}^{EM*} = \mathrm{diag}\,\left(\min\left\{\frac{\alpha}{f_i} + \frac{\beta}{\sum_{j=1}^p 1/f_j}, 1\right\}, i = 1, \dots, p\right)$$

given by (3.18). These estimators are, respectively, abbreviated by the notations

$$UB, JS_1, JS_2, BE, ST, EM, EMK, BE^*, ST^*, EM^*.$$

Table 1: Relative Risk Efficiencies of the Estimators in the Case that $p = 3$, $q = 2$, $k = 20$, $r = 2$, $n = 20$, $m = 18$, $\boldsymbol{\Sigma} = \boldsymbol{I}_3$ and $\boldsymbol{\Sigma}_A = \boldsymbol{H}\mathrm{diag}\,(\gamma/10,\ 2\gamma/10,\ 3\gamma/10)\boldsymbol{H}'$ for $\gamma = 0, \ldots, 9$

γ	UB	JS_1	JS_2	BE	ST	EM	EMK	BE^*	ST^*	EM^*
0	0.377	0.348	0.352	0.351	0.247	0.267	0.282	0.268	0.159	0.117
1	0.377	0.348	0.352	0.351	0.251	0.268	0.284	0.274	0.170	0.131
2	0.377	0.348	0.352	0.351	0.259	0.272	0.286	0.282	0.187	0.153
3	0.377	0.349	0.352	0.351	0.267	0.275	0.289	0.289	0.203	0.172
4	0.377	0.349	0.352	0.351	0.275	0.279	0.292	0.294	0.216	0.187
5	0.377	0.349	0.352	0.351	0.282	0.282	0.294	0.297	0.227	0.200
6	0.377	0.349	0.352	0.351	0.288	0.285	0.297	0.299	0.236	0.209
7	0.377	0.349	0.352	0.351	0.293	0.288	0.299	0.301	0.243	0.217
8	0.377	0.349	0.352	0.351	0.298	0.291	0.301	0.302	0.249	0.224
9	0.377	0.349	0.352	0.351	0.302	0.293	0.303	0.303	0.254	0.229

Table 2: Relative Risk Efficiencies of the Estimators in the case that $p = 10$, $q = 2$, $k = 20$, $r = 2$, $n = 20$, $m = 18$, $\boldsymbol{\Sigma} = \boldsymbol{I}_{10}$ and $\boldsymbol{\Sigma}_A = \boldsymbol{H}\mathrm{diag}\,(i \times \gamma/10,\ i = 1, \ldots, 10)\boldsymbol{H}'$ for $\gamma = 0, \ldots, 9$

γ	UB	JS_1	JS_2	BE	ST	EM	EMK	BE^*	ST^*	EM^*
0	0.824	0.732	0.749	0.747	0.345	0.255	0.299	0.676	0.239	0.122
1	0.824	0.735	0.750	0.747	0.382	0.303	0.343	0.689	0.308	0.208
2	0.824	0.736	0.750	0.747	0.421	0.352	0.387	0.700	0.367	0.283
3	0.824	0.737	0.751	0.747	0.452	0.389	0.421	0.708	0.408	0.334
4	0.824	0.738	0.751	0.747	0.476	0.418	0.447	0.713	0.439	0.372
5	0.824	0.739	0.751	0.747	0.496	0.441	0.468	0.716	0.464	0.402
6	0.824	0.739	0.751	0.747	0.514	0.460	0.486	0.719	0.484	0.425
7	0.824	0.739	0.751	0.747	0.528	0.477	0.501	0.720	0.501	0.445
8	0.824	0.740	0.751	0.747	0.541	0.492	0.514	0.721	0.515	0.461
9	0.824	0.740	0.751	0.747	0.553	0.504	0.526	0.722	0.527	0.476

Table 3: Relative Risk Efficiencies of the Estimators in the case that $p = 10$, $q = 8$, $k = 20$, $r = 3$, $n = 40$, $m = 12$, $\boldsymbol{\Sigma} = \boldsymbol{I}_{10}$ and $\boldsymbol{\Sigma}_A = \boldsymbol{H}\mathrm{diag}\,(i \times \gamma/10,\ i = 1, \ldots, 10)\boldsymbol{H}'$ for $\gamma = 1, \ldots, 10$

γ	UB	JS_1	JS_2	BE	ST	EM	EMK	BE^*	ST^*	EM^*
0	0.941	0.856	0.935	0.935	0.287	0.098	0.101	0.895	0.173	0.038
1	0.941	0.872	0.936	0.935	0.361	0.224	0.227	0.898	0.289	0.179
2	0.941	0.880	0.937	0.936	0.426	0.326	0.328	0.900	0.371	0.285
3	0.942	0.886	0.937	0.936	0.474	0.396	0.398	0.902	0.427	0.358
4	0.942	0.890	0.937	0.936	0.512	0.448	0.450	0.903	0.469	0.412
5	0.942	0.894	0.938	0.936	0.543	0.489	0.491	0.904	0.503	0.455
6	0.943	0.896	0.938	0.937	0.570	0.523	0.525	0.905	0.532	0.490
7	0.943	0.899	0.938	0.937	0.593	0.552	0.553	0.905	0.556	0.519
8	0.943	0.901	0.938	0.937	0.613	0.576	0.577	0.906	0.578	0.544
9	0.943	0.903	0.938	0.937	0.631	0.597	0.598	0.906	0.596	0.565

The risk functions or the relative risk efficiencies of the above estimators are computed through simulation experiments based on 50,000 replications. The simulation experiments are done in the following three cases:

Case 1: $p = 3$, $q = 2$, $k = 20$, $r = 2$, $n = 20$, $m = 18$, $\boldsymbol{\Sigma} = \boldsymbol{I}_3$ and

$$\boldsymbol{\Sigma}_A = \boldsymbol{H}\mathrm{diag}\,(\gamma/10,\ 2\gamma/10,\ 3\gamma/10)\boldsymbol{H}', \quad \boldsymbol{\Sigma}_2 = \boldsymbol{I}_3 + r\boldsymbol{\Sigma}_A$$

for $\gamma = 0, \ldots, 9$ and an orthogonal matrix $\boldsymbol{H}$,

Case 2: $p = 10$, $q = 2$, $k = 20$, $r = 2$, $n = 20$, $m = 18$, $\boldsymbol{\Sigma} = \boldsymbol{I}_{10}$ and

$$\boldsymbol{\Sigma}_A = \boldsymbol{H}\mathrm{diag}\,(i \times \gamma/10,\ i = 1, \ldots, 10)\boldsymbol{H}', \quad \boldsymbol{\Sigma}_2 = \boldsymbol{I}_{10} + r\boldsymbol{\Sigma}_A$$

for $\gamma = 0, \ldots, 9$.

Case 3: $p = 10$, $q = 8$, $k = 20$, $r = 3$, $n = 40$, $m = 12$, $\boldsymbol{\Sigma} = \boldsymbol{I}_{10}$ and

$$\boldsymbol{\Sigma}_A = \boldsymbol{H}\mathrm{diag}\,(i \times \gamma/10,\ i = 1, \ldots, 10)\boldsymbol{H}', \quad \boldsymbol{\Sigma}_2 = \boldsymbol{I}_{10} + r\boldsymbol{\Sigma}_A$$

for $\gamma = 0, \ldots, 9$.

The relative risk efficiencies $Eff(\widehat{\boldsymbol{\Delta}}, \omega)$ of the above estimators for the three cases are given in Tables 1, 2 and 3, respectively. Form these tables, the following conclusions can be drawn.

(1) The Efron-Morris type truncated estimator EM^* has very nice risk behaviours and it is the best among all the estimators so that it is highly recommended.

(2) The comparison between the Stein-type and Efron-Morris type estimators ST and EM depend on the cases. Table 1 reveals that ST is better than EM while Tables 2 and 3 show EM is superior. Anyway, EM^* is better than both of them.

(3) The estimator EM is better than EMK, but the differences in the risks are quite small.

(4) The estimators UB, JS_1, JS_2, BE and BE^* are minimax, but much worse than EM^* through the three tables.

We now give an example for data derived by Monte Carlo simulation.

Example 1. (*Simulated Data*) We first treat a data set $\boldsymbol{y}_{ij}$, $i = 1, \dots, k$, $j = 1, \dots, r$, derived by Monte Carlo simulation under the model (2.1) in the case that $k = 20$, $r = 3$, $q = 2$, $p = 5$, $n = k(r-1)$, $m = k - q$, $\boldsymbol{\Sigma} = \boldsymbol{I}_5$, $\boldsymbol{\Sigma}_A = \boldsymbol{H}\mathrm{diag}\,(i/10, i = 1, \dots, 5)\boldsymbol{H}'$ and $\beta_{ij} = (\boldsymbol{\beta})_{ij} = i + j - 2$. The covariates $\boldsymbol{b}_1, \dots, \boldsymbol{b}_{20}$ are derived as independent random variables from a 2-variate normal distribution with

$$E[\boldsymbol{b}_i] = \frac{i-1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{Cov}\,(\boldsymbol{b}_i) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

for $\rho = 0.4$. Then $(\overline{\boldsymbol{y}}_{1\cdot}, \dots, \overline{\boldsymbol{y}}_{20\cdot})$, $\widehat{\boldsymbol{\beta}}$, $\boldsymbol{S}$ and $\boldsymbol{W}$ are computed as the followings:

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} -0.0343 & 1.0667 \\ 0.8944 & 2.0810 \\ 1.9547 & 3.0670 \\ 2.9659 & 4.0867 \\ 3.9309 & 5.0445 \end{pmatrix},$$

$$(\overline{\boldsymbol{y}}_{1\cdot}, \dots, \overline{\boldsymbol{y}}_{20\cdot})' = \begin{pmatrix} 0.251 & 0.117 & -1.170 & -1.532 & -1.471 \\ -1.289 & -0.259 & -1.098 & 0.903 & -0.639 \\ 2.022 & 3.771 & 4.620 & 7.731 & 7.409 \\ 2.280 & 9.107 & 13.457 & 17.784 & 24.354 \\ 1.465 & 6.646 & 10.798 & 14.410 & 19.976 \\ 1.982 & 6.993 & 11.499 & 16.937 & 21.567 \\ 5.674 & 13.799 & 21.814 & 30.496 & 39.932 \\ 2.055 & 7.280 & 14.994 & 19.060 & 22.942 \\ 5.436 & 10.756 & 20.050 & 30.312 & 35.811 \\ 4.825 & 13.262 & 21.981 & 29.531 & 38.479 \\ 6.975 & 19.493 & 30.814 & 42.024 & 53.244 \\ 5.159 & 15.304 & 27.075 & 37.801 & 47.510 \\ 3.978 & 13.889 & 23.322 & 31.381 & 42.676 \\ 5.160 & 17.225 & 29.040 & 42.796 & 53.938 \\ 6.515 & 21.040 & 33.741 & 49.448 & 62.569 \\ 11.707 & 29.466 & 49.589 & 68.192 & 87.503 \\ 6.078 & 22.272 & 37.262 & 52.430 & 65.752 \\ 10.216 & 26.370 & 46.310 & 64.047 & 80.195 \\ 9.751 & 29.300 & 46.661 & 66.522 & 84.010 \\ 10.239 & 24.600 & 43.285 & 60.447 & 77.882 \end{pmatrix},$$

$$
\boldsymbol{S} = \begin{pmatrix} 41.921 & 3.293 & 0.951 & 0.720 & 3.357 \\ 3.293 & 45.926 & -9.940 & 11.328 & 7.114 \\ 0.951 & -9.940 & 46.946 & 10.198 & -0.629 \\ 0.720 & 11.328 & 10.198 & 46.175 & -8.015 \\ 3.357 & 7.114 & -0.629 & -8.015 & 58.362 \end{pmatrix},
$$

$$
\boldsymbol{W} = \begin{pmatrix} 43.722 & -5.520 & 7.103 & 1.167 & -3.000 \\ -5.520 & 29.700 & -2.892 & -10.936 & -2.106 \\ 7.103 & -2.892 & 34.376 & 6.710 & -0.493 \\ 1.167 & -10.936 & 6.710 & 58.105 & -3.625 \\ -3.000 & -2.106 & -0.493 & -3.625 & 13.327 \end{pmatrix}.
$$

Based on these statistics, $\boldsymbol{\theta}_i = \boldsymbol{\beta}\boldsymbol{b}_i + \boldsymbol{\alpha}_i$ can be predicted by

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

for an estimator $\widehat{\boldsymbol{\Delta}}$ of $\boldsymbol{\Delta} = \boldsymbol{\Sigma}(\boldsymbol{\Sigma}+r\boldsymbol{\Sigma}_A)^{-1}$. We denote the predictors $\widehat{\boldsymbol{\theta}}_i$ for the estimators $\widehat{\boldsymbol{\Delta}}^{UB}$, $\widehat{\boldsymbol{\Delta}}^{ST}$, $\widehat{\boldsymbol{\Delta}}^{EM}$, $\widehat{\boldsymbol{\Delta}}^{EM*}$ by

$$
pUB,\ pST,\ pEM,\ pEM^*, \quad \text{respectively.}
$$

The sample mean vector $(\overline{\boldsymbol{y}}_{1\cdot}, \ldots, \overline{\boldsymbol{y}}_{20\cdot})$ corresponding to $\widehat{\boldsymbol{\Delta}} = \boldsymbol{0}$ is also treated as a predictor.

The predicted values by pST, pEM and pEM^* for $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_{11}$ and $\boldsymbol{\theta}_{20}$ are reported in Table 4 where $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \theta_{i3}, \theta_{i4}, \theta_{i5})'$. For each i, the values of $\boldsymbol{\theta}_i$, $\overline{\boldsymbol{y}}_i$ and $\widehat{\boldsymbol{\beta}}\boldsymbol{b}_i$ are also given there. It is revealed in Table 4 that the predicted values for pST, pEM and pEM^* are between $\overline{\boldsymbol{y}}_i$ and $\widehat{\boldsymbol{\beta}}\boldsymbol{b}_i$ because of shrinking $\overline{\boldsymbol{y}}_i$ towards $\widehat{\boldsymbol{\beta}}\boldsymbol{b}_i$. It is also seen that the predicted values by pEM^* are, in most cases, closer to the true values of $\boldsymbol{\theta}_i$ than those by $\overline{\boldsymbol{y}}_i$. Since we know the true values of $\boldsymbol{\theta}_i$ in this example, we can compute the prediction errors such as $\sum_{i=1}^k (\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)'(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)$, which are given by $\{(\overline{\boldsymbol{y}}_{1\cdot}, \ldots, \overline{\boldsymbol{y}}_{20\cdot}), 35.984\}$, $\{pUB, 27.232\}$, $\{pST, 23.032\}$, $\{pEM, 23.973\}$ and $\{pEM^*, 20.364\}$. The predictor pEM^* with the Efron-Morris type truncated estimator $\widehat{\boldsymbol{\Delta}}^{EM*}$ has the smallest prediction error. ■

5 Estimation of Mean Matrix in a Fixed Effects Model

In this section we consider the model (2.1)

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

in which $\boldsymbol{\alpha}_i$ are unknown parameters representing the fixed effects of the ith group. We shall assume that

$$
\sum_{i=1}^k \boldsymbol{\alpha}_i \boldsymbol{b}_i' = \boldsymbol{0},
$$

and $\boldsymbol{\epsilon}_{ij}$ are i.i.d. $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. Thus, $\mathbf{Cov}\,(\boldsymbol{y}_{ij}) = \boldsymbol{\Sigma}$. Our aim is to estimate $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k)$, where

$$
\boldsymbol{\theta}_i = \boldsymbol{\beta}\boldsymbol{b}_i + \boldsymbol{\alpha}_i, \quad i = 1, \ldots, k,
$$

Table 4: Predicted values of the Predictors for $i = 2, 11, 20$

		θ_{i1}	θ_{i2}	θ_{i3}	θ_{i4}	θ_{i5}
	$\boldsymbol{\theta}_i$	-1.3065	-1.1717	-0.4064	-0.1669	-1.1618
	$\overline{\boldsymbol{y}}_i$	-1.2899	-0.2590	-1.0989	0.9033	-0.6391
	$\widehat{\boldsymbol{\beta}\boldsymbol{b}_i}$	-0.2721	-0.4260	-0.5585	-0.7047	-0.8405
$i = 2$	UB	-1.1024	-0.6585	-0.9092	0.4891	-0.9594
	pST	-0.9390	-0.6180	-0.8582	0.3247	-0.9145
	pEM	-0.9166	-0.6409	-0.8223	0.2111	-0.9592
	pEM^*	-0.8906	-0.5870	-0.8319	0.2125	-0.8174
	$\boldsymbol{\theta}_i$	6.389	19.253	29.775	41.484	53.350
	$\overline{\boldsymbol{y}}_i$	6.975	19.493	30.814	42.024	53.244
	$\widehat{\boldsymbol{\beta}\boldsymbol{b}_i}$	6.805	18.282	30.256	42.196	53.491
$i = 11$	UB	6.862	18.946	30.679	41.762	53.310
	pST	6.855	18.865	30.592	41.812	53.343
	pEM	6.841	18.767	30.581	41.832	53.352
	pEM^*	6.849	18.783	30.578	41.832	53.394
	$\boldsymbol{\theta}_i$	9.291	26.246	43.522	59.932	77.322
	$\overline{\boldsymbol{y}}_i$	10.239	24.600	43.285	60.447	77.882
	$\widehat{\boldsymbol{\beta}\boldsymbol{b}_i}$	9.046	25.732	43.274	60.700	77.196
$i = 20$	UB	9.874	24.921	43.297	60.792	77.023
	pST	9.729	25.054	43.301	60.790	77.109
	pEM	9.675	25.113	43.293	60.799	76.993
	pEM^*	9.719	25.204	43.277	60.801	77.232

under the loss function

$$\operatorname{tr}(\widehat{\mathbf{\Theta}}-\mathbf{\Theta})'\mathbf{\Sigma}^{-1}(\widehat{\mathbf{\Theta}}-\mathbf{\Theta}),$$

where $\widehat{\mathbf{\Theta}}$ is an estimator of $\mathbf{\Theta}$. In previous sections, several estimators dominating $\widehat{\mathbf{\Delta}}_0$ have been proposed, and from (2.3) and (2.5), it is seen that the resulting estimators of $\mathbf{\Theta}$ in the random effects model are better than the estimator

$$\widehat{\mathbf{\Theta}}_0 = (\overline{\boldsymbol{y}}_1.,\dots,\overline{\boldsymbol{y}}_k.) - \widehat{\mathbf{\Delta}}_0\widetilde{\boldsymbol{Y}},$$

where $\widetilde{\boldsymbol{Y}} = \left(\overline{\boldsymbol{y}}_1. - \widehat{\boldsymbol{\beta}}\boldsymbol{b}_1,\dots,\overline{\boldsymbol{y}}_k. - \widehat{\boldsymbol{\beta}}\boldsymbol{b}_k\right)$ and $\widehat{\mathbf{\Delta}}_0 = (m-p-1)(n+p+1)^{-1}\boldsymbol{S}\boldsymbol{W}^{-1}$. In this section, we demonstrates that these dominance results still hold 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 $\sum_{i=1}^k \boldsymbol{\alpha}_i\boldsymbol{b}_i' = \mathbf{0}$. Let $\boldsymbol{B} = (\boldsymbol{b}_1,\dots,\boldsymbol{b}_k)$ and assume that $\operatorname{rank}(\boldsymbol{B}) = q_1 \le q < k$. Letting $\boldsymbol{P} = \boldsymbol{B}'(\boldsymbol{B}\boldsymbol{B}')^-\boldsymbol{B}$, we observe that

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

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

$$\widehat{\mathbf{\Theta}}(\widehat{\mathbf{\Delta}}) = (\overline{\boldsymbol{y}}_1.,\dots,\overline{\boldsymbol{y}}_k.) - \widehat{\mathbf{\Delta}}\widetilde{\boldsymbol{Y}} \tag{5.1}$$

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

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

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

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

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

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

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 $\boldsymbol{Z} = \sqrt{r}\widetilde{\boldsymbol{Y}}\boldsymbol{Q}_1$ and $\boldsymbol{\mu} = \sqrt{r}\widetilde{\boldsymbol{\alpha}}\boldsymbol{Q}_1$. Then $\boldsymbol{Z} \sim \mathcal{N}_{p\times m}(\boldsymbol{\mu},\mathbf{\Sigma},\boldsymbol{I}_m)$ and $\boldsymbol{W} = \boldsymbol{Z}\boldsymbol{Z}'$, so that $R_m(\widehat{\mathbf{\Theta}}(\widehat{\mathbf{\Delta}}),\omega)$ is expressed as

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

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

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

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

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

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

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

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 5. *In the fixed effects model, consider the problem of estimating the unknown matrix of parameters* $\boldsymbol{\Theta} = \boldsymbol{\beta}(\boldsymbol{b}_1, \dots, \boldsymbol{b}_k) + (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_k)$ *by the estimator* $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}})$ *given by (5.1) relative to the risk (2.2). For the scale-equivariant estimator* $\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi})$ *given by (3.3), the unbiased estimator of the risk of* $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}))$ *is* $pk/r + (1/r)r(\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}))$*, where* $r(\widehat{\boldsymbol{\Delta}}(\boldsymbol{\Psi}))$ *is given by (3.5).*

Corollary 1. *In the fixed effects model, the estimators* $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}^{ST})$ *and* $\widehat{\boldsymbol{\Theta}}(\widehat{\boldsymbol{\Delta}}^{EM})$ *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 the case of $\boldsymbol{b}_1 = \cdots = \boldsymbol{b}_k = \mathbf{0}$, Proposition 5 was given by Konno(1991). However, by using the arguments of Efron and Morris (1976), we obtain simpler proofs even in the general case.

6 An Extension of the Model and Remarks

We here consider the model given in (1.3) which is 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

$$\boldsymbol{y}_{ij} = \boldsymbol{\beta}\boldsymbol{b}_{ij} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, r, \tag{6.1}$$

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

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

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

$$\widehat{\boldsymbol{\theta}}_i^B = \overline{\boldsymbol{y}}_{i\cdot} - \boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1}\left(\overline{\boldsymbol{y}}_{i\cdot} - \boldsymbol{\beta}\overline{\boldsymbol{b}}_{i\cdot}\right). \tag{6.2}$$

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 $\boldsymbol{c}_{ij} = \boldsymbol{b}_{ij} - \overline{\boldsymbol{b}}_{i\cdot}$. Also, let $\boldsymbol{P}_2 = \overline{\boldsymbol{B}}'(\overline{\boldsymbol{B}}\,\overline{\boldsymbol{B}}')^-\overline{\boldsymbol{B}}$ and $\boldsymbol{P}_1 = \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^-\boldsymbol{C}$ for $\overline{\boldsymbol{B}} = (\overline{\boldsymbol{b}}_{1\cdot}, \ldots, \overline{\boldsymbol{b}}_{k\cdot})$. Then,

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

where $\widehat{\boldsymbol{\beta}}_1 = \boldsymbol{U}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^-$ and $\boldsymbol{S} = (\boldsymbol{U} - \widehat{\boldsymbol{\beta}}_1\boldsymbol{C})(\boldsymbol{U} - \widehat{\boldsymbol{\beta}}_1\boldsymbol{C})'$. Letting $\overline{\boldsymbol{Y}} = (\overline{\boldsymbol{y}}_{1\cdot}, \ldots, \overline{\boldsymbol{y}}_{k\cdot})$, we see that

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

where $\widehat{\boldsymbol{\beta}}_2 = \overline{\boldsymbol{Y}}\,\overline{\boldsymbol{B}}'(\overline{\boldsymbol{B}}\,\overline{\boldsymbol{B}}')^-$ and $\boldsymbol{W} = r(\overline{\boldsymbol{Y}} - \widehat{\boldsymbol{\beta}}_2\overline{\boldsymbol{B}})(\overline{\boldsymbol{Y}} - \widehat{\boldsymbol{\beta}}_2\overline{\boldsymbol{B}})'$. Assuming that rank $(\boldsymbol{C}\boldsymbol{C}') = q_1 \le q$ and 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

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

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 (6.2), we consider the estimator

$$\widehat{\boldsymbol{\theta}}_i^{EB}(\widehat{\boldsymbol{\beta}}_2) = \overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\Delta}}(\overline{\boldsymbol{y}}_{i\cdot} - \widehat{\boldsymbol{\beta}}_2\overline{\boldsymbol{b}}_{i\cdot}), \tag{6.3}$$

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 (6.1).

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

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

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:

$$\boldsymbol{y}_{ij} = \boldsymbol{\beta}\boldsymbol{b}_{ij} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i, \tag{6.4}$$

which was discussed by Fuller and Harter (1987) for estimation of small area. It seems, however, intractable to establish any exact dominance results in the model (6.4). The work of deriving efficient estimators by using approximation (asymptotic) theories rests in the future.

Example 2. (*Posted Land Price Data*) We here treat the posted land price data in Japan which are disclosed as the Digital National Information and also obtained from the web page of the Ministry of Land, Infrastructure and Transport. 45 small areas are chosen from Kanagawa prefecture next to Tokyo, and 5 spots are taken from each area. For the j-th spot in the i-th small area, the posted land prices (Yen) per m^2 of the spot in 2001 and 1996, denoted by y_{ij1} and y_{ij2} respectively, are observed with two covariates b_{ij2} and b_{ij3} where b_{ij2} is the distance from the spot to the nearby railway station and b_{ij3} is the nearby distance from the station to the Tokyo station. All the data are transformed by logarithm before carrying out the analysis. Many people living in Kanagawa use railways to commute to Tokyo, so that the distances to the nearby station and to Tokyo affect the land prices. In fact, the multiple correlation coefficient between y_{ij1}'s and (b_{ij2}, b_{ij3})'s is 0.7932.

Since the land price depends on the region such as city and town, we need to consider regional effect in the model building. We thus employ the mixed linear model (6.1) to analyze the data, namely,

$$\boldsymbol{y}_{ij} = \boldsymbol{\beta}\boldsymbol{b}_{ij} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}, \quad i = 1, \dots, 45, \quad j = 1, \dots, 5,$$

where $p = 2$, $q = 3$, $k = 45$, $r = 5$, $\boldsymbol{y}_{ij} = (y_{ij1}, y_{ij2})^t$ and $\boldsymbol{b}_{ij} = (b_{ij1}, b_{ij2} b_{ij3})^t$ for $b_{ij1} = 1$. Then, $n = k(r-1) - q + 1 = 178$ and $m = k - q = 42$. Also $\widehat{\boldsymbol{\beta}}_2$, $\boldsymbol{S}$ and $\boldsymbol{W}$ are given by

$$\widehat{\boldsymbol{\beta}}_2 = \begin{pmatrix} 15.204 & -0.390 & -0.196 \\ 15.569 & -0.400 & -0.212 \end{pmatrix},$$

$$\boldsymbol{S} = \begin{pmatrix} 3.488 & 4.885 \\ 4.885 & 7.560 \end{pmatrix}, \quad \boldsymbol{W} = \begin{pmatrix} 3.018 & 3.510 \\ 3.510 & 4.685 \end{pmatrix},$$

which yeild the unbiased estimates of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_A$, given by

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 0.01959 & 0.02744 \\ 0.02744 & 0.04247 \end{pmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_A = \begin{pmatrix} 0.01045 & 0.01122 \\ 0.01122 & 0.01381 \end{pmatrix}.$$

Calculating the modified likelihood ratio test for the hypothesis $H_0 : \boldsymbol{\Sigma}_A = \boldsymbol{0}$ (see page 490 in Srivastava (2002)), we see that the null hypothesis is rejected, which implies that the regional effects exist in the data.

Based on these statistics, the average land price $\boldsymbol{\theta}_i = \boldsymbol{\beta}\overline{\boldsymbol{b}}_{i\cdot} + \boldsymbol{\alpha}_i$ in the i-th area can be estimated by (6.3)

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

for an estimator $\widehat{\boldsymbol{\Delta}}$ of $\boldsymbol{\Delta} = \boldsymbol{\Sigma}(\boldsymbol{\Sigma} + r\boldsymbol{\Sigma}_A)^{-1}$. Using the same notations as in Example 1, we denote the estimators $\widehat{\boldsymbol{\theta}}_i$ for the estimators $\widehat{\boldsymbol{\Delta}}^{UB}$, $\widehat{\boldsymbol{\Delta}}^{ST}$ and $\widehat{\boldsymbol{\Delta}}^{EM}$ by pUB, pST and pEM. The sample mean vector $\overline{\boldsymbol{y}}_{i\cdot}$, the regression synthetic estimate $\widehat{\boldsymbol{\beta}}_2\overline{\boldsymbol{b}}_{i\cdot}$ and the estimates by pUB, pST and pEM are reported in Table 5 for the nine choosed small areas, $i = 3, 8, 14, 17, 20, 25, 35, 36, 45$, where the estimates given in Table 5 are transformed by exponential, namely, they give estimates of average price (Yen) per m^2. It is revealed in Table 5 that the estimates pUB, pST and pEM are between $\overline{\boldsymbol{y}}_{i\cdot}$ and $\widehat{\boldsymbol{\beta}}_2\overline{\boldsymbol{b}}_{i\cdot}$ because of shrinking $\overline{\boldsymbol{y}}_{i\cdot}$ towards $\widehat{\boldsymbol{\beta}}_2\overline{\boldsymbol{b}}_{i\cdot}$. However, the shrinkage factors $\widehat{\boldsymbol{\Delta}}^{UB}$, $\widehat{\boldsymbol{\Delta}}^{ST}$ and $\widehat{\boldsymbol{\Delta}}^{EM}$ are not so large as their values are given by

$$\widehat{\boldsymbol{\Delta}}^{UB} = \begin{pmatrix} -0.09717 & 0.30125 \\ -0.43998 & 0.68319 \end{pmatrix}, \quad \widehat{\boldsymbol{\Delta}}^{ST} = \begin{pmatrix} -0.07767 & 0.28610 \\ -0.41786 & 0.66345 \end{pmatrix},$$

$$\widehat{\boldsymbol{\Delta}}^{EM} = \begin{pmatrix} -0.08156 & 0.29625 \\ -0.43269 & 0.68586 \end{pmatrix}.$$

It is also noted that the truncated rules (3.17) and (3.18) do not change the estimates for the data treated in this example, namely, $\widehat{\boldsymbol{\Delta}}^{ST*} = \widehat{\boldsymbol{\Delta}}^{ST}$ and $\widehat{\boldsymbol{\Delta}}^{EM*} = \widehat{\boldsymbol{\Delta}}^{EM}$. ■

Acknowledgments. The authors are grateful to Professor Y. Kanemoto, University of Tokyo, for his kindness of giving us the data treated in Example 2. The research of the

Table 5: Estimates of Average Land Prices

Area	Year	$\overline{\boldsymbol{y}}_{i\cdot}$	pUB	pST	pEM	$\widehat{\boldsymbol{\beta}}_2\overline{\boldsymbol{b}}_{i\cdot}$
$i = 3$	2001	163,780	171,400	171,420	171,670	192,850
	1996	192,830	206,240	206,160	206,600	236,380
$i = 8$	2001	228,900	230,240	230,130	230,180	226,250
	1996	276,150	280,530	280,370	280,520	280,470
$i = 14$	2001	356,480	333,230	333,470	332,750	297,320
	1996	494,000	441,170	441,860	440,250	372,460
$i = 17$	2001	334,830	318,560	318,310	317,820	267,080
	1996	422,060	396,170	396,060	395,270	332,590
$i = 20$	2001	293,340	277,730	278,160	277,670	270,070
	1996	414,270	372,660	373,510	372,220	336,410
$i = 25$	2001	278,510	264,090	263,740	263,310	211,900
	1996	340,330	320,320	320,050	319,450	261,170
$i = 35$	2001	330,290	310,460	310,520	309,910	270,140
	1996	438,920	398,650	399,000	397,780	334,950
$i = 36$	2001	262,970	278,630	278,850	279,350	339,050
	1996	325,300	350,780	350,850	351,680	427,850
$i = 45$	2001	225,090	207,150	207,300	206,750	177,650
	1996	307,290	268,120	268,580	267,410	216,110

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 Nos. 11680320 and 13680371. The research of the second author was supported in part by Natural Sciences and Engineering Research Council of Canada. This work was done during the visit of the first author to the University of Toronto, 1999 and 2002 summer.

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.

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

Breiman, L. and Friedman, J.H. (1997). Predicting multivariate responses in multiple linear regression. *J. Roy. Statist. Soc. Ser. B*, **59** , 3-54.

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.

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. and Srivastava, M.S. (1999). Prediction in multivariate mixed linear models with equal replications. Discussion Paper CIRJE-F-62, Faculty of Economics, University of Tokyo.

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

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.

Rao, C.R. (1975). Simultaneous estimation of parameters in different linear models and applications to biometric problems. *Biometrics*, **31**, 545-554.

Reinsel, G.C. (1985). Mean squared error properties of empirical Bayes estimators in a multivariate random effects general linear model. *J. Amer. Statist. Assoc.*, **80**, 642-650.

Srivastava, M.S. (2002). *Methods of Multivariate Statistics*, Wiley, New York.

Srivastava, M.S. and Solanky, T.K.S. (2002). Predicting multivariate response for linear regression model. *Commun. Statist.-Comp. Simul.*, to appear.

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)

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