Bayesian Simultaneous Estimation for Means in *k* Sample Problems

By

Ryo Imai, Tatsuya Kubokawa, Malay Ghosh

April 2018

CENTER FOR RESEARCH AND EDUCATION FOR POLICY EVALUATION
DISCUSSION PAPER NO. 21

CENTER FOR RESEARCH AND EDUCATION FOR POLICY EVALUATION (CREPE)
THE UNIVERSITY OF TOKYO
http://www.crepe.e.u-tokyo.ac.jp/

Bayesian Simultaneous Estimation for Means in k Sample Problems

Ryo Imai*, Tatsuya Kubokawa† and Malay Ghosh‡

November 29, 2017

Abstract

This paper is concerned with the simultaneous estimation of k population means when one suspects that the k means are nearly equal. As an alternative to the preliminary test estimator based on the test statistics for testing hypothesis of equal means, we derive Bayesian and minimax estimators which shrink individual sample means toward a pooled mean estimator given under the hypothesis. Interestingly, it is shown that both the preliminary test estimator and the Bayesian minimax shrinkage estimators are further improved by shrinking the pooled mean estimator. The performance of proposed shrinkage estimators is investigated by simulation.

Key words and phrases: Admissibility, Bayes estimator, decision theory, empirical Bayes, k sample problem, minimaxity, pooled estimator, preliminary test estimator, quadratic loss, shrinkage estimator, simultaneous estimation.

1 Introduction

Consider the multivariate k sample problem with the following canonical form: p-variate random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ and positive scalar random variable S are mutually independently distributed as

$$\begin{aligned} \boldsymbol{X}_i \sim& \mathcal{N}_p(\boldsymbol{\mu}_i, \sigma^2 \boldsymbol{V}_i), \quad \text{for } i = 1, \ldots, k, \\ S/\sigma^2 \sim& \chi^2_n, \end{aligned} \tag{1.1}$$

where the p-variate means $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$ and the scale parameter σ^2 are unknown, and $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_k$ are $p \times p$ known and nonsingular symmetric matrices. In this model, we want to estimate $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$ simultaneously relative to the quadratic loss function

$$L(\boldsymbol{\delta}, \boldsymbol{\omega}) = \sum_{i=1}^k \|\boldsymbol{\delta}_i - \boldsymbol{\mu}_i\|^2_{\boldsymbol{V}_i^{-1}}/\sigma^2 = \sum_{i=1}^k (\boldsymbol{\delta}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{V}_i^{-1} (\boldsymbol{\delta}_i - \boldsymbol{\mu}_i)/\sigma^2, \tag{1.2}$$

*Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN. E-Mail: imronuronnar13@g.ecc.u-tokyo.ac.jp

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

‡Department of Statistics, University of Florida, 102 Griffin-Floyd Hall, Gainesville, Florida 32611. E-Mail: ghoshm@stat.ufl.edu

where $\|\boldsymbol{a}\|_{\boldsymbol{A}}^2 = \boldsymbol{a}^\top \boldsymbol{A}\boldsymbol{a}$ for the transpose $\boldsymbol{a}^\top$ of $\boldsymbol{a}$, $\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_k)$ for estimator $\boldsymbol{\delta}_i$ of $\boldsymbol{\mu}_i$, and $\boldsymbol{\omega} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k, \sigma^2)$ is a set of unknown parameters.

A typical example of the model (1.1) is a k sample problem. For $i = 1, \ldots, k$, a sample with size n_i is obtained from the i-th population, and

$$\boldsymbol{X}_{ij} \sim \mathcal{N}_p(\boldsymbol{\mu}_i, \sigma^2 \boldsymbol{V}_{i,0}), \quad j = 1, \ldots, n_i, \tag{1.3}$$

where $\boldsymbol{V}_{i,0}$ is a known matrix. In this case, $\boldsymbol{X}_i$, $\boldsymbol{V}_i$, S and n in (1.1) correspond to $\overline{\boldsymbol{X}}_i = n_i^{-1} \sum_{j=1}^{n_i} \boldsymbol{X}_{ij}$, $\boldsymbol{V}_{i,0}/n_i$, $\mathrm{tr}\,[\sum_{i=1}^k \boldsymbol{V}_{i,0}^{-1} \sum_{j=1}^{n_i} (\boldsymbol{X}_{ij} - \overline{\boldsymbol{X}}_i)(\boldsymbol{X}_{ij} - \overline{\boldsymbol{X}}_i)^\top]$ and $\sum_{i=1}^k (n_i - 1)p$, respectively.

Another example of (1.1) is k linear regression models such that

$$\boldsymbol{y}_i = \boldsymbol{Z}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, k, \tag{1.4}$$

where $\boldsymbol{y}_i$ is an n_i-variate vector of observations, $\boldsymbol{\beta}_i$ is a p-variate vector of regression coefficients, $\boldsymbol{Z}_i$ is an $n_i \times p$ matrix of explanatory variables and $\boldsymbol{\varepsilon}_i$ is an n_i-variate random vector having $\mathcal{N}_{n_i}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{n_i})$. In this case, $\boldsymbol{X}_i$, $\boldsymbol{V}_i$, S and n in (1.1) correspond to $\widehat{\boldsymbol{\beta}}_i = (\boldsymbol{Z}_i^\top \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i^\top \boldsymbol{y}_i$, $(\boldsymbol{Z}_i^\top \boldsymbol{Z}_i)^{-1}$, $\sum_{i=1}^k \|\boldsymbol{y}_i - \boldsymbol{Z}_i \widehat{\boldsymbol{\beta}}_i\|_{\boldsymbol{I}_{n_i}}^2$ and $\sum_{i=1}^k (n_i - p)$, respectively.

In this paper, we consider the case that the means $\boldsymbol{\mu}_i$'s are close to the hypothesis

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k.$$

For example, suppose we consider k laboratories, say Laboratory L_1, …, Laboratory L_k, which use similar instruments to measure several common characteristics. It is then suspected that the k population mean vectors are nearly equal. In such a situation, rather than estimating each mean by the corresponding sample mean separately, it might be more preferable to use a pooled mean estimator. A classical approach towards solution to this problem is a preliminary test estimator which uses the pooled mean estimator upon acceptance of the null hypothesis H_0, and uses separate sample means upon rejection of H_0. However, the preliminary test estimators are non-smooth and do not necessarily improve on sample means. As an alternative method, we consider Bayesian methods including hierarchical Bayes and empirical Bayes estimators.

A Bayesian approach to this problem uses the prior distribution of $\boldsymbol{\mu}_i$ assuming a multivariate normal distribution with mean $\boldsymbol{\nu}$ and variance τ^2 where τ^2 is assumed to have a hierarchical prior distribution. The mean $\boldsymbol{\nu}$ corresponds to $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k = \boldsymbol{\nu}$. Ghosh and Sinha (1988) considered this prior distribution in the framework of estimating the single mean $\boldsymbol{\mu}_1$ and showed that the resulting empirical and hierarchical Bayes estimators have reasonable forms of shrinking $\boldsymbol{X}_1$ towards the pooled estimator of $\boldsymbol{\nu}$. Also they derived conditions for those shrinkage estimators to be minimax. Shrinkage estimators with minimaxity has been studied in the literature since Stein (1956) established the inadmissibility of the standard estimator $\boldsymbol{X}_1$. For papers related to the context of this paper, see James and Stein (1961), Strawderman (1971, 73), Efron and Morris (1976) Stein (1981) and Sun (1996).

The Bayesian approach suggested in this paper consists of two kinds of shrinkage. One is to shrink the k sample means toward a pooled mean estimator, and the other is to shrink the p-dimensional pooled estimator toward a constant like zero. The former shrinkage is treated in Section 2.1, and a class of minimax estimators is derived. Out of the class, we develop hierarchical and empirical Bayes estimators with minimaxity. The latter shrinkage is discussed in Section 2.2, and we show that all the estimators derived by the former shrinkage can be further improved by using the latter shrinkage.

In Section 3, we find a Bayes and minimax estimator, namely an admissible and minimax estimator using both kinds of shrinkage. The Bayes estimator which consists of the two kinds of shrinkage has a new form which we cannot handle with conventional techniques given in Section 2. Thus, we need to extend the classes of minimax estimators in order to treat such a new type of shrinkage estimators. The minimaxity of the Bayes estimator is successfully established.

In Section 4, through simulation studies, we investigate the performance of several hierarchical Bayes and empirical Bayes estimators and the preliminary test estimators. The simulation results show that the hierarchical Bayes and empirical Bayes estimators for the two kinds of prior distributions have good performances.

2 Improvement by Two Kinds of Shrinkage

2.1 Shrinkage toward the pooled estimator

We begin with assuming the prior distribution

$$\begin{aligned} \boldsymbol{\mu}_i \mid \boldsymbol{\nu}, \tau^2 \sim & \mathcal{N}_p(\boldsymbol{\nu}, \tau^2 \boldsymbol{V}_i), \quad \text{for } i = 1, \ldots, k, \\ \boldsymbol{\nu} \sim & \text{Uniform}(\mathbb{R}^p), \end{aligned} \tag{2.1}$$

where Uniform($\mathbb{R}^p$) denotes the improper uniform distribution over $\mathbb{R}^p$. The posterior distribution of $\boldsymbol{\mu}_i$ given $\boldsymbol{X}_i$ and $\boldsymbol{\nu}$ and the marginal distribution of $\boldsymbol{X}_i$ given $\boldsymbol{\nu}$ are

$$\begin{aligned} \boldsymbol{\mu}_i \mid \boldsymbol{X}_i, \boldsymbol{\nu}, \tau^2, \sigma^2 \sim & \mathcal{N}_p\Big(\widehat{\boldsymbol{\mu}}_i^*(\sigma^2, \tau^2, \boldsymbol{\nu}), (\sigma^{-2} + \tau^{-2})^{-1}\Big), \\ \boldsymbol{X}_i \mid \boldsymbol{\nu}, \tau^2, \sigma^2 \sim & \mathcal{N}_p(\boldsymbol{\nu}, (\tau^2 + \sigma^2)\boldsymbol{V}_i), \end{aligned} \qquad i = 1, \ldots, k, \tag{2.2}$$

where $\widehat{\boldsymbol{\mu}}_i^*(\sigma^2, \tau^2, \boldsymbol{\nu})$ is

$$\widehat{\boldsymbol{\mu}}_i^*(\sigma^2, \tau^2, \boldsymbol{\nu}) = \boldsymbol{X}_i - \frac{\sigma^2}{\tau^2 + \sigma^2}(\boldsymbol{X}_i - \boldsymbol{\nu}). \tag{2.3}$$

Also, the posterior distribution of $\boldsymbol{\nu}$ given $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ and the marginal distribution of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ are

$$\begin{aligned} & \boldsymbol{\nu} \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k, \tau^2, \sigma^2 \sim \mathcal{N}_p(\widehat{\boldsymbol{\nu}}, (\tau^2 + \sigma^2)\boldsymbol{A}), \\ & f_\pi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \mid \tau^2, \sigma^2) \propto \frac{1}{(\tau^2 + \sigma^2)^{p(k-1)/2}} \exp\Big\{ - \frac{\sum_{i=1}^k \|\boldsymbol{x}_i - \widehat{\boldsymbol{\nu}}\|^2_{\boldsymbol{V}_i^{-1}}}{2(\tau^2 + \sigma^2)} \Big\}, \end{aligned} \tag{2.4}$$

where

$$\boldsymbol{A} = \Big(\sum_{i=1}^k \boldsymbol{V}_i^{-1} \Big)^{-1} \quad \text{and} \quad \widehat{\boldsymbol{\nu}} = \boldsymbol{A} \sum_{i=1}^k \boldsymbol{V}_i^{-1} \boldsymbol{X}_i. \tag{2.5}$$

Then, the Bayes estimator of $\boldsymbol{\mu}_i$ is derived from (2.3) relative to the quadratic loss (1.2) as

$$\widehat{\boldsymbol{\mu}}_i^B(\sigma^2, \tau^2) = \widehat{\boldsymbol{\mu}}_i^*(\sigma^2, \tau^2, \widehat{\boldsymbol{\nu}}) = \boldsymbol{X}_i - \frac{\sigma^2}{\tau^2 + \sigma^2}(\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}}). \tag{2.6}$$

Because $\tau^2+\sigma^2$ and σ^2 are unknown, we estimate $\tau^2+\sigma^2$ by $\sum_{i=1}^k \|\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}}\|^2_{\boldsymbol{V}_i^{-1}}/\{p(k-1)-2\}$ from the marginal likelihood in (2.4). When σ^2 is estimated by $\hat{\sigma}^2 = S/(n+2)$, the resulting empirical Bayes estimator is

$$\widehat{\boldsymbol{\mu}}_i^{EB1} = \boldsymbol{X}_i - \min\Big\{\frac{\{p(k-1)-2\}/(n+2)}{F}, 1\Big\}(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}}), \tag{2.7}$$

where

$$F = \sum_{i=1}^k \|\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}}\|^2_{\boldsymbol{V}_i^{-1}}/S.$$

Motivated from the empirical Bayes estimator, we consider the class of estimators

$$\widehat{\boldsymbol{\mu}}_i^S(\phi_0) = \boldsymbol{X}_i - \frac{\phi_0(F)}{F}(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}}), \tag{2.8}$$

where $\phi_0(F)$ is an absolutely continuous function. We can provide conditions on $\phi_0(F)$ for minimaxity of $\widehat{\boldsymbol{\mu}}_i^S(\phi_0)$. The proof of Theorem 2.1 is given in the appendix.

Theorem 2.1 *The estimator* $(\widehat{\boldsymbol{\mu}}_1^S(\phi_0),\ldots,\widehat{\boldsymbol{\mu}}_k^S(\phi_0))$ *is minimax relative to the quadratic loss* (1.2) *if* $\phi_0(F)$ *satisfies the following conditions:*
(a) $\phi_0(F)$ *is non-decreasing in* F*.*
(b) $0<\phi_0(F)\le 2\{p(k-1)-2\}/(n+2)$*.*

It follows from Theorem 2.1 that the empirical Bayes estimator $(\widehat{\boldsymbol{\mu}}_1^{EB1},\ldots,\widehat{\boldsymbol{\mu}}_k^{EB1})$ is minimax. Another minimax estimator is the hierarchical Bayes estimator. In addition to the prior distribution (2.1), we assume that

$$\begin{aligned}\pi(\tau^2 \mid \sigma^2) \propto& \Big(\frac{\sigma^2}{\tau^2+\sigma^2}\Big)^{a+1},\\ \pi(\sigma^2) \propto& (\sigma^2)^{c-2},\end{aligned} \tag{2.9}$$

where a and c are constants. From (2.4), the posterior distribution of (τ^2,σ^2) given $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_k,S$ is

$$\begin{aligned}&\pi(\tau^2,\sigma^2 \mid \boldsymbol{x}_1,\ldots,\boldsymbol{x}_k,S)\\ &\propto \Big(\frac{\sigma^2}{\tau^2+\sigma^2}\Big)^{p(k-1)/2+a+1}\Big(\frac{1}{\sigma^2}\Big)^{\{n+p(k-1)\}/2+2-c}\exp\Big\{-\frac{\sum_{i=1}^k\|\boldsymbol{x}_i-\widehat{\boldsymbol{\nu}}\|^2_{\boldsymbol{V}_i^{-1}}}{2(\tau^2+\sigma^2)}-\frac{S}{2\sigma^2}\Big\}.\end{aligned}$$

Making the transformation $\lambda=\sigma^2/(\tau^2+\sigma^2)$ and $\eta=1/\sigma^2$ with the Jacobian $|\partial(\tau^2,\sigma^2)/\partial(\lambda,\eta)| = 1/(\lambda^2\eta^3)$ gives

$$\pi(\lambda,\eta\mid \boldsymbol{x}_1,\ldots,\boldsymbol{x}_k,S)\propto \lambda^{p(k-1)/2+a-1}\eta^{\{n+p(k-1)\}/2-c-1}\exp\Big\{-\frac{\lambda\eta}{2}\sum_{i=1}^k\|\boldsymbol{x}_i-\widehat{\boldsymbol{\nu}}\|^2_{\boldsymbol{V}_i^{-1}}-\frac{\eta}{2}S\Big\},$$

where $0<\lambda<1$ and $\eta>0$. Since σ^2 has the prior distribution, from (2.6), the Bayes estimator of $\boldsymbol{\mu}_i$ relative to the quadratic loss (1.2) is written as

$$\begin{aligned}\widehat{\boldsymbol{\mu}}_i^{HB1} =& E[\boldsymbol{\mu}_i/\sigma^2 \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_k,S]/E[1/\sigma^2\mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_k,S]\\ =&\boldsymbol{X}_i-\frac{E[(\tau^2+\sigma^2)^{-1}\mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_k,S]}{E[(\sigma^2)^{-1}\mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_k,S]}(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}}),\end{aligned} \tag{2.10}$$

where

$$\frac{E[(\tau^2+\sigma^2)^{-1} \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_k,S]}{E[(\sigma^2)^{-1} \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_k,S]}
= \frac{\int_0^1\int_0^\infty \lambda^{p(k-1)/2+a}\eta^{\{n+p(k-1)\}/2-c}\exp\{-\frac{\eta S}{2}(\lambda F+1)\}d\eta d\lambda}{\int_0^1\int_0^\infty \lambda^{p(k-1)/2+a-1}\eta^{\{n+p(k-1)\}/2-c}\exp\{-\frac{\eta S}{2}(\lambda F+1)\}d\eta d\lambda}.$$

Making the transformations $x = F\lambda$ and $v = S\eta$ gives the expression

$$\frac{E[(\tau^2+\sigma^2)^{-1} \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_k,S]}{E[(\sigma^2)^{-1} \mid \boldsymbol{X}_1,\ldots,\boldsymbol{X}_k,S]} = \frac{\phi^{HB1}(F)}{F},$$

where

$$\phi^{HB1}(F) = \frac{\int_0^F x^{p(k-1)/2+a}/(1+x)^{\{n+p(k-1)\}/2+1-c}dx}{\int_0^F x^{p(k-1)/2+a-1}/(1+x)^{\{n+p(k-1)\}/2+1-c}dx} \tag{2.11}$$

We can derive a condition on a and c for minimaxity of the hierarchical Bayes estimator.

Theorem 2.2 *The hierarchical Bayes estimator* (2.10) *is minimax if* $p(k-1) \geq 3$ *and if* a *and* c *satisfy* $a + c < (n+2)/2$ *and*

$$\{2p(k-1)+n-2\}a + 2\{p(k-1)-2\}c \leq \frac{n-2}{2}p(k-1) - 2n. \tag{2.12}$$

Proof. It suffices to show that $\phi^{HB1}(F)$ satisfies the conditions (a) and (b) in Theorem 2.1. For the proof of (a), see the proof of Theorem 3.2. To check the condition (b), we note that

$$\int_0^\infty \frac{x^\ell}{(1+x)^m}dx = \int_0^1 w^\ell(1-w)^{m-\ell-2}dw = B(\ell+1, m-\ell-1), \tag{2.13}$$

for the beta function $B(\cdot,\cdot)$. Using the condition (a), we see that

$$\begin{aligned}\phi^{HB1}(F) \leq \lim_{F\to\infty}\phi^{HB1}(F) &= \frac{\int_0^\infty x^{p(k-1)/2+a}/(1+x)^{\{n+p(k-1)\}/2+1-c}dx}{\int_0^\infty x^{p(k-1)/2+a-1}/(1+x)^{\{n+p(k-1)\}/2+1-c}dx}\\ &= \frac{B(p(k-1)/2+a+1, n/2-a-c)}{B(p(k-1)/2+a, n/2-a-c+1)} = \frac{p(k-1)+2a}{n-2(a+c)}.\end{aligned}$$

Thus, the condition (b) is satisfied if $a + c < n/2$ and if

$$\frac{p(k-1)+2a}{n-2(a+c)} \leq 2\frac{p(k-1)-2}{n+2},$$

which is equivalently rewritten by (2.12). Hence, the minimaxity of the hierarchical Bayes estimator is established. □

Remark 2.1 Assuming the uniform distribution for $\boldsymbol{\nu}$ in (2.1) yields the Bayes estimator $\widehat{\boldsymbol{\mu}}_i^*(\sigma^2, \tau^2, \widehat{\boldsymbol{\nu}})$. It is interesting to note that this Bayes estimator is also interpreted as an empirical Bayes estimator, because when $\boldsymbol{\nu}$ is an unknown parameter, $\widehat{\boldsymbol{\nu}}$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\nu}$ in the marginal likelihood $\boldsymbol{X}_i \sim \mathcal{N}_p(\boldsymbol{\nu}, (\tau^2+\sigma^2)\boldsymbol{V}_i)$ for $i = 1,\ldots,k$. Thus, the estimator $\widehat{\boldsymbol{\mu}}_i^*(\sigma^2, \tau^2, \widehat{\boldsymbol{\nu}})$ is a Bayes empirical Bayes estimator.

2.2 Shrinkage of the pooled estimator

The empirical Bayes estimator $\widehat{\boldsymbol{\mu}}_i^{EB1}$ and the hierarchical Bayes estimator $\widehat{\boldsymbol{\mu}}_i^{HB1}$ shrink k individual estimators $\boldsymbol{X}_i$ toward the common mean $\widehat{\boldsymbol{\nu}}$. Because $\widehat{\boldsymbol{\nu}}$ is a p-dimensional estimator against the uniform prior over $\mathbb{R}^p$, we can consider to shrink $\widehat{\boldsymbol{\nu}}$ for further improvement. To this end, we first assume the prior distribution

$$\begin{aligned}\boldsymbol{\mu}_i \mid \boldsymbol{\nu}, \tau^2 \sim& \mathcal{N}_p(\boldsymbol{\nu}, \tau^2 \boldsymbol{V}_i), \quad \text{for } i = 1, \ldots, k,\\ \boldsymbol{\nu} \mid \gamma^2, \sim& \mathcal{N}_p(0, \gamma^2 \boldsymbol{A}),\end{aligned} \tag{2.14}$$

for $\gamma > 0$ and $\boldsymbol{A} = \Big(\sum_{i=1}^k \boldsymbol{V}_i^{-1}\Big)^{-1}$. It is noted that the normal prior is considered instead of the uniform prior in (2.1). Then the posterior distributions are given by

$$\begin{aligned}\boldsymbol{\mu}_i \mid \boldsymbol{X}_i, \boldsymbol{\nu}, \tau^2, \sigma^2 \sim& \mathcal{N}_p\Big(\widehat{\boldsymbol{\mu}}_i^B(\sigma^2, \tau^2, \boldsymbol{\nu}), (\sigma^{-2} + \tau^{-2})^{-1}\Big), \qquad i = 1, \ldots, k,\\ \boldsymbol{\nu} \mid \boldsymbol{X}, \tau^2, \gamma^2, \sigma^2 \sim& \mathcal{N}_p\Big(\frac{\gamma^2}{\gamma^2 + \tau^2 + \sigma^2}\widehat{\boldsymbol{\nu}}, \frac{\gamma^2(\tau^2 + \sigma^2)}{\gamma^2 + \tau^2 + \sigma^2}\boldsymbol{A}\Big),\end{aligned} \tag{2.15}$$

where $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$, and $\widehat{\boldsymbol{\mu}}_i^*(\sigma^2, \tau^2, \boldsymbol{\nu})$ is given in (2.3). The marginal density of $\boldsymbol{X}$ is

$$\begin{aligned}f_\pi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \mid \tau^2, \gamma^2, \sigma^2) \propto& \frac{1}{(\tau^2 + \sigma^2)^{p(k-1)/2}} \exp\Big\{ - \frac{\sum_{i=1}^k \|\boldsymbol{x}_i - \widehat{\boldsymbol{\nu}}\|^2_{\boldsymbol{V}_i^{-1}}}{2(\tau^2 + \sigma^2)}\Big\}\\ &\times \frac{1}{(\gamma^2 + \tau^2 + \sigma^2)^{p/2}} \exp\Big\{ - \frac{\|\widehat{\boldsymbol{\nu}}\|^2_{\boldsymbol{A}^{-1}}}{2(\gamma^2 + \tau^2 + \sigma^2)}\Big\}.\end{aligned} \tag{2.16}$$

Then from (2.3) and (2.15), the Bayes estimator of $\boldsymbol{\mu}_i$ is

$$\begin{aligned}\widehat{\boldsymbol{\mu}}_i^B(\sigma^2, \tau^2, \gamma^2) =& \boldsymbol{X}_i - \frac{\sigma^2}{\tau^2 + \sigma^2}(\boldsymbol{X}_i - E[\boldsymbol{\nu} \mid \boldsymbol{X}])\\ =& \boldsymbol{X}_i - \frac{\sigma^2}{\tau^2 + \sigma^2}\Big(\boldsymbol{X}_i - \frac{\gamma^2}{\gamma^2 + \tau^2 + \sigma^2}\widehat{\boldsymbol{\nu}}\Big)\\ =& \boldsymbol{X}_i - \frac{\sigma^2}{\tau^2 + \sigma^2}(\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}}) - \frac{\sigma^2}{\gamma^2 + \tau^2 + \sigma^2}\widehat{\boldsymbol{\nu}}.\end{aligned} \tag{2.17}$$

We shall estimate $\tau^2 + \sigma^2$, $\gamma^2 + \tau^2 + \sigma^2$ by $\sum_{i=1}^k \|\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}}\|^2_{\boldsymbol{V}_i^{-1}}/\{p(k-1) - 2\}$ and $\|\widehat{\boldsymbol{\nu}}\|^2_{\boldsymbol{A}^{-1}}/(p-2)$, respectively, from the marginal likelihood in (2.16). When σ^2 is estimated by $\hat{\sigma}^2 = S/(n+2)$, the resulting empirical Bayes estimator is

$$\widehat{\boldsymbol{\mu}}_i^{EB2} = \boldsymbol{X}_i - \min\Big\{\frac{\{p(k-1) - 2\}/(n+2)}{F}, 1\Big\}(\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}}) - \min\Big\{\frac{(p-2)/(n+2)}{G}, 1\Big\}\widehat{\boldsymbol{\nu}}, \tag{2.18}$$

where $F = \sum_{i=1}^k \|\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}}\|^2_{\boldsymbol{V}_i^{-1}}/S$ and

$$G = \|\widehat{\boldsymbol{\nu}}\|^2_{\boldsymbol{A}^{-1}}/S.$$

The estimator $\widehat{\boldsymbol{\mu}}_i^{EB2}$ enjoys shrinking $\boldsymbol{X}_i$ toward $\widehat{\boldsymbol{\nu}}$ and shrinking $\widehat{\boldsymbol{\nu}}$ toward zero. Motivated from the empirical Bayes estimator, we consider the class of double shrinkage estimators

$$\widehat{\boldsymbol{\mu}}_i^S(\phi_0, \psi_0) = \boldsymbol{X}_i - \frac{\phi_0(F)}{F}(\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}}) - \frac{\psi_0(G)}{G}\widehat{\boldsymbol{\nu}}, \tag{2.19}$$

where $\phi_0(F)$ and $\psi_0(G)$ are absolutely continuous functions. We can provide conditions on $\psi_0(G)$ for improving on $\widehat{\boldsymbol{\mu}}_i^S(\phi_0)$ in (2.8). The proof of Theorem 2.3 is given in the appendix.

Theorem 2.3 *The estimator* $(\widehat{\boldsymbol{\mu}}_1^S(\phi_0), \ldots, \widehat{\boldsymbol{\mu}}_k^S(\phi_0))$ *in* (2.8) *is improved on by the double shrinkage estimator* $(\widehat{\boldsymbol{\mu}}_1^S(\phi_0, \psi_0), \ldots, \widehat{\boldsymbol{\mu}}_k^S(\phi_0, \psi_0))$ *in* (2.19) *relative to the quadratic loss* (1.2) *if* $\psi_0(G)$ *satisfies the following conditions:*

(a) $\psi_0(G)$ *is non-decreasing in* G.

(b) $0 < \psi_0(G) \leq 2(p-2)/(n+2)$.

It follows from Theorem 2.3 that the empirical Bayes estimator $(\widehat{\boldsymbol{\mu}}_1^{EB1}, \ldots, \widehat{\boldsymbol{\mu}}_k^{EB1})$ in (2.7) is improved on by $(\widehat{\boldsymbol{\mu}}_1^{EB2}, \ldots, \widehat{\boldsymbol{\mu}}_k^{EB2})$ in (2.18). The preliminary test estimator based on the statistic F for testing the hypothesis $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$ is

$$\begin{aligned}\widehat{\boldsymbol{\mu}}_i^{PT} =& \begin{cases} \boldsymbol{X}_i & \text{if } F > (p(k-1)/n)F_{p(k-1),n,\alpha} \\ \widehat{\boldsymbol{\nu}} & \text{otherwise} \end{cases} \\ =& \boldsymbol{X}_i - I_{[F \leq (p(k-1)/n)F_{p(k-1),n,\alpha}]}(\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}}),\end{aligned} \tag{2.20}$$

where $F_{p(k-1),n,\alpha}$ is the upper α point of the F distribution with $(p(k-1), n)$ degrees of freedom, and $I_{[\cdot]}$ denotes the indicator function. Because the preliminary test estimator belongs to the class (2.8), Theorem 2.3 implies that the estimator $(\widehat{\boldsymbol{\mu}}_1^{PT}, \ldots, \widehat{\boldsymbol{\mu}}_k^{PT})$ can be improved on by $(\widehat{\boldsymbol{\mu}}_1^{PT*}, \ldots, \widehat{\boldsymbol{\mu}}_k^{PT*})$ with

$$\widehat{\boldsymbol{\mu}}_i^{PT*} = \widehat{\boldsymbol{\mu}}_i^{PT} - \min\Big\{\frac{(p-2)/(n+2)}{G}, 1\Big\}\widehat{\boldsymbol{\nu}}. \tag{2.21}$$

3 Bayes and minimax estimation

We shall derive the Bayes and minimax estimator. To this end, we assume the following prior distribution in addition to (2.14):

$$\begin{aligned}\pi(\tau^2 \mid \sigma^2) \propto& \Big(\frac{\sigma^2}{\tau^2+\sigma^2}\Big)^{a+1}, \\ \pi(\gamma^2 \mid \tau^2, \sigma^2) \propto& \Big(\frac{\sigma^2}{\gamma^2+\tau^2+\sigma^2}\Big)^{b+1}, \\ \pi(\sigma^2) \propto& (\sigma^2)^{c-3}, \quad \text{for } \sigma^2 \leq 1/L,\end{aligned} \tag{3.1}$$

where a, b and c are constants and L is a positive constant. When $a > 0$, $b > 0$ and $c > 0$, the prior distribution (3.1) is proper, because by making the transformation $\xi = \sigma^2/(\gamma^2 + \tau^2 + \sigma^2)$, $\lambda = \sigma^2/(\tau^2 + \sigma^2)$ and $\eta = 1/\sigma^2$ with the Jacobian $|\partial(\tau^2, \sigma^2)/\partial(\lambda, \eta)| = 1/(\lambda^2\xi^2\eta^4)$, we have

$$\begin{aligned}&\int_0^{1/L}\int_0^\infty\int_0^\infty \Big(\frac{\sigma^2}{\tau^2+\sigma^2}\Big)^{a+1}\Big(\frac{\sigma^2}{\gamma^2+\tau^2+\sigma^2}\Big)^{b+1}\Big(\frac{1}{\sigma^2}\Big)^{3-c} d\tau^2 d\gamma^2 d\sigma^2 \\ &= \int_L^\infty\int_0^1\int_0^1 \lambda^{a+1}\xi^{b+1}\eta^{3-c}\frac{1}{\lambda^2\xi^2\eta^4} d\lambda d\xi d\eta = \int_0^1 \lambda^{a-1}d\lambda\int_0^1 \xi^{b-1}d\xi\int_L^\infty \frac{1}{\eta^{1+c}}d\eta < \infty.\end{aligned}$$

From (2.15), the posterior distribution of $(\tau^2, \gamma^2, \sigma^2)$ given $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k, S$ is

$$\begin{aligned}
&\pi(\tau^2, \gamma^2, \sigma^2 \mid \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, S) \\
&\propto \Big(\frac{\sigma^2}{\tau^2+\sigma^2}\Big)^{p(k-1)/2+a+1} \exp\Big\{ -\frac{SF}{2(\tau^2+\sigma^2)}\Big\} \\
&\quad \times \Big(\frac{\sigma^2}{\gamma^2+\tau^2+\sigma^2}\Big)^{p/2+b+1} \exp\Big\{ -\frac{SG}{2(\gamma^2+\tau^2+\sigma^2)}\Big\} \times \Big(\frac{1}{\sigma^2}\Big)^{(pk+n)/2+3-c},
\end{aligned} \tag{3.2}$$

for $\sigma^2 < 1/L$. Using the transformation $\xi = \sigma^2/(\gamma^2+\tau^2+\sigma^2)$, $\lambda = \sigma^2/(\tau^2+\sigma^2)$ and $\eta = 1/\sigma^2$ again, we can rewrite the posterior distribution as

$$\pi(\xi, \lambda, \eta \mid \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, S) \propto \lambda^{p(k-1)/2+a-1}\xi^{p/2+b-1}\eta^{(n+pk)/2-c-1} \exp\Big\{ -\frac{S\eta}{2}(\lambda F + \xi G + 1)\Big\},$$

where $0 < \xi < 1$, $0 < \lambda < 1$ and $\eta \geq L$. Using the same arguments as in (2.10), from (2.17), the hierarchical Bayes estimator of $\boldsymbol{\mu}_i$ relative to the quadratic loss (1.2) is written as

$$\widehat{\boldsymbol{\mu}}_i^{HB2} = \boldsymbol{X}_i - \frac{E[(\tau^2+\sigma^2)^{-1} \mid \boldsymbol{X}, S]}{E[(\sigma^2)^{-1} \mid \boldsymbol{X}, S]}(\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}}) - \frac{E[(\gamma^2+\tau^2+\sigma^2)^{-1} \mid \boldsymbol{X}, S]}{E[(\sigma^2)^{-1} \mid \boldsymbol{X}, S]}\widehat{\boldsymbol{\nu}}, \tag{3.3}$$

where $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$,

$$\begin{aligned}
&\frac{E[(\tau^2+\sigma^2)^{-1} \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k, S]}{E[(\sigma^2)^{-1} \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k, S]} \\
&\quad = \frac{\int_0^1\int_0^1\int_L^\infty \lambda^{p(k-1)/2+a}\xi^{p/2+b-1}\eta^{(n+pk)/2-c} \exp\{-\frac{\eta S}{2}(\lambda F + \xi G + 1)\} d\eta d\xi d\lambda}{\int_0^1\int_0^1\int_L^\infty \lambda^{p(k-1)/2+a-1}\xi^{p/2+b-1}\eta^{(n+pk)/2-c} \exp\{-\frac{\eta S}{2}(\lambda F + \xi G + 1)\} d\eta d\xi d\lambda}.
\end{aligned}$$

Making the transformations $x = F\lambda$, $y = G\xi$ and $v = S\eta$ gives the expression

$$\frac{E[(\tau^2+\sigma^2)^{-1} \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k, S]}{E[(\sigma^2)^{-1} \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k, S]} = \frac{\phi^{HB2}(F, G, S)}{F},$$

where

$$\phi^{HB2}(F, G, S) = \frac{\int_0^F\int_0^G\int_{LS}^\infty x^{p(k-1)/2+a}y^{p/2+b-1}v^{(n+pk)/2-c} \exp\{-v(x+y+1)/2\} dvdydx}{\int_0^F\int_0^G\int_{LS}^\infty x^{p(k-1)/2+a-1}y^{p/2+b-1}v^{(n+pk)/2-c} \exp\{-v(x+y+1)/2\} dvdydx}. \tag{3.4}$$

Similarly

$$\frac{E[(\xi^2+\tau^2+\sigma^2)^{-1} \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k, S]}{E[(\sigma^2)^{-1} \mid \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k, S]} = \frac{\psi^{HB2}(F, G, S)}{G},$$

where

$$\psi^{HB2}(F, G, S) = \frac{\int_0^F\int_0^G\int_{LS}^\infty x^{p(k-1)/2+a-1}y^{p/2+b}v^{(n+pk)/2-c} \exp\{-v(x+y+1)/2\} dvdydx}{\int_0^F\int_0^G\int_{LS}^\infty x^{p(k-1)/2+a-1}y^{p/2+b-1}v^{(n+pk)/2-c} \exp\{-v(x+y+1)/2\} dvdydx}. \tag{3.5}$$

Thus, the hierarchical Bayes estimator in (3.3) is expressed as

$$\widehat{\boldsymbol{\mu}}_i^{HB2} = \boldsymbol{X}_i - \frac{\phi^{HB2}(F, G, S)}{F}(\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}}) - \frac{\psi^{HB2}(F, G, S)}{G}\widehat{\boldsymbol{\nu}}. \tag{3.6}$$

This estimator does not belong to the class (2.19), because shrinkage functions $\phi^{HB2}(F,G,S)$ and $\psi^{HB2}(F,G,S)$ include both F and G.

To show minimaxity of the hierarchical Bayes estimator (3.6), we begin by deriving an unbiased risk estimator for the following double shrinkage estimators with more general forms:

$$\widehat{\boldsymbol{\mu}}_i(\phi,\psi) = \boldsymbol{X}_i - \frac{\phi(F,G,S)}{F}(\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}}) - \frac{\psi(F,G,S)}{G}\widehat{\boldsymbol{\nu}}, \tag{3.7}$$

where $\phi(F,G,S)$ and $\psi(F,G,S)$ are absolutely continuous functions.

Theorem 3.1 *An unbiased estimator of the risk function of* (3.7) *relative to the loss* (1.2) *is*

$$\begin{aligned} UER(\phi,\psi) =& pk - \frac{[2\{p(k-1)-2\}-(n+2)\phi]\phi}{F} - 4\phi_F - \frac{4\phi}{F}(F\phi_F + G\phi_G) + \frac{4S\phi\phi_S}{F} \\ & - \frac{\{2(p-2)-(n+2)\psi\}\psi}{G} - 4\phi_G - \frac{4\psi}{G}(F\psi_F + G\psi_G) + \frac{4S\psi\psi_S}{G}, \end{aligned} \tag{3.8}$$

where $\phi_F = \partial\phi(F,G,S)/\partial F$, $\phi_G = \partial\phi(F,G,S)/\partial G$, $\phi_S = \partial\phi(F,G,S)/\partial S$, and ψ_F, ψ_G and ψ_S are defined similarly.

Considering the case with $\psi(F,G,S) = 0$ in (3.8), we can extend Theorem 2.1 to the general case.

Proposition 3.1 *The estimator $(\widehat{\boldsymbol{\mu}}_1(\phi),\ldots,\widehat{\boldsymbol{\mu}}_k(\phi))$ with the form $\widehat{\boldsymbol{\mu}}_i(\phi) = \boldsymbol{X}_i - \{\phi(F,G,S)/F\}(\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}})$ is minimax relative to the quadratic loss* (1.2) *if $\phi(F,G,S)$ satisfies the following conditions:*
(a) *$\phi(F,G,S)$ is non-decreasing in F and G, and non-increasing in S.*
(b) *$0 < \phi(F,G,S) \le 2\{p(k-1)-2\}/(n+2)$.*

The unbiased risk estimator (3.8) also enables us to extend Theorem 2.3 to the general case.

Proposition 3.2 *The estimator $(\widehat{\boldsymbol{\mu}}_1(\phi),\ldots,\widehat{\boldsymbol{\mu}}_k(\phi))$ given in Proposition 3.1 is improved on by the double shrinkage estimator $(\widehat{\boldsymbol{\mu}}_1(\phi,\psi),\ldots,\widehat{\boldsymbol{\mu}}_k(\phi,\psi))$ in* (3.7) *relative to the quadratic loss* (1.2) *if $\psi(F,G,S)$ satisfies the following conditions:*
(a) *$\psi(F,G,S)$ is non-decreasing in F and G, and non-increasing in S.*
(b) *$0 < \psi(F,G,S) \le 2(p-2)/(n+2)$.*

We now derive conditions under which the hierarchical Bayes estimator (3.6) is minimax, namely it satisfies the conditions in Propositions 3.1 and 3.2. Combining the condition for the minimaxity and the condition for the proper prior, we can get a condition for the Bayes estimator to be admissible and minimax.

Theorem 3.2 *The hierarchical Bayes estimator* (3.6) *is minimax if $p(k-1) \ge 3$, $p \ge 3$ and if a, b and c satisfy $a+b+c < n/2$ and*

$$\{2p(k-1)+n-2\}a + 2\{p(k-1)-2\}(b+c) \le p(k-1)(n-2)/2 - 2n, \tag{3.9}$$

$$(2p+n-2)b + 2(p-2)(a+c) \le p(n-2)/2 - 2n. \tag{3.10}$$

Further, (3.6) *is a Bayes minimax estimator, namely admissible and minimax, provided constants a, b and c are positive and satisfy the above conditions.*

The latter part of Theorem 3.2 follows from the fact that the prior distribution (3.1) is proper when $a > 0$, $c > 0$ and $L > 0$. Then, the Bayes estimator (3.6) is admissible and minimax. Since $k \geq 2$, we can find such constants if $p \geq 5$ and $n > 2p/(p-4)$. This shows that p is at least 5.

Remark 3.1 Although the condition of $L > 0$ is imposed technically in order to guarantee the admissibility, we have no information on L as well as the Bayes estimator based on triple integrals is computationally hard to derive. From a practical point of view, we set $L = 0$. In this case, the hierarchical prior (3.1) is improper and $\widehat{\boldsymbol{\mu}}_i^{HB2}$ is no longer guaranteed to be admissible. However, the minimaxity still holds, and the resulting generalized Bayes estimator is

$$\widehat{\boldsymbol{\mu}}_i^{HB2} = \boldsymbol{X}_i - \frac{\phi^{HB2}(F,G)}{F}(\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}}) - \frac{\psi^{HB2}(F,G)}{G}\widehat{\boldsymbol{\nu}}, \tag{3.11}$$

where $\phi^{HB2}(F,G,S)$ and $\psi^{HB2}(F,G,S)$ are expressed based on double integrals as

$$\phi^{HB2}(F,G) = \frac{\int_0^F \int_0^G x^{p(k-1)/2+a} y^{p/2+b-1}(x+y+1)^{-\{(n+pk)/2-c+1\}}dydx}{\int_0^F \int_0^G x^{p(k-1)/2+a-1} y^{p/2+b-1}(x+y+1)^{-\{(n+pk)/2-c+1\}}dydx},$$

$$\psi^{HB2}(F,G) = \frac{\int_0^F \int_0^G x^{p(k-1)/2+a-1} y^{p/2+b}(x+y+1)^{-\{(n+pk)/2-c+1\}}dydx}{\int_0^F \int_0^G x^{p(k-1)/2+a-1} y^{p/2+b-1}(x+y+1)^{-\{(n+pk)/2-c+1\}}dydx}.$$

In our simulation experiments given in the next section, we use this generalized Bayes estimator.

4 Simulation Study

We investigate the numerical performances of the risk functions of the preliminary-test estimator and several empirical and hierarchical Bayes estimators through simulation. The estimators which we compare are the following seven:

JS: the James-Stein estimator

$$\widehat{\boldsymbol{\mu}}_i^{JS} = \boldsymbol{X}_i - \frac{p-2}{n+2}\frac{S}{\|\boldsymbol{X}_i\|^2_{\boldsymbol{V}_i^{-1}}}\boldsymbol{X}_i,$$

PT: the preliminary-test estimator $\widehat{\boldsymbol{\mu}}_i^{PT}$ given in (2.20),
PT*: the preliminary-test and shrinkage estimator $\widehat{\boldsymbol{\mu}}_i^{PT*}$ in (2.21),
EB: the empirical Bayes estimator $\widehat{\boldsymbol{\mu}}_i^{EB1}$ in (2.7),
EB*: the improved empirical Bayes estimator $\widehat{\boldsymbol{\mu}}_{EB2}$ in (2.18)
HB1: the hierarchical Bayes estimator $\widehat{\boldsymbol{\mu}}_i^{HB1}$ in (2.10),
HB2: the hierarchical Bayes estimator $\widehat{\boldsymbol{\mu}}_i^{HB2}$ in (3.11).

The significance size α in the preliminary-test and related estimators $\widehat{\boldsymbol{\mu}}_i^{PT}$ and $\widehat{\boldsymbol{\mu}}_i^{PT*}$ is $\alpha = 0.05$. For the hierarchical Bayes estimators $\widehat{\boldsymbol{\mu}}_i^{HB1}$ and $\widehat{\boldsymbol{\mu}}_i^{HB2}$, we used the constants $a = b = c = 0.1$ and $L = 0$.

In this simulation, we generate random numbers of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ and S based on the model (1.1) for $p = k = 5$, $n = 20$, $\sigma^2 = 2$ and $\boldsymbol{V}_i = (0.1 \times i)\boldsymbol{I}_p$, $i = 1, \ldots, k$. For the mean vectors $\boldsymbol{\mu}_i$, we treat the eight cases:

$$
\begin{aligned}
(\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_5) = & (\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0}), (2\boldsymbol{j}_5,2\boldsymbol{j}_5,2\boldsymbol{j}_5,2\boldsymbol{j}_5,2\boldsymbol{j}_5),\\
& (-0.4\boldsymbol{j}_5,-0.2\boldsymbol{j}_5,\mathbf{0},0.2\boldsymbol{j}_5,0.4\boldsymbol{j}_5), (-\boldsymbol{j}_5,-0.5\boldsymbol{j}_5,\mathbf{0},0.5\boldsymbol{j}_5,\boldsymbol{j}_5), (-2\boldsymbol{j}_5,-\boldsymbol{j}_5,\mathbf{0},\boldsymbol{j}_5,2\boldsymbol{j}_5),\\
& (1.2\boldsymbol{j}_5,1.4\boldsymbol{j}_5,1.6\boldsymbol{j}_5,1.8\boldsymbol{j}_5,2.0\boldsymbol{j}_5), (\boldsymbol{j}_5,1.5\boldsymbol{j}_5,2\boldsymbol{j}_5,2.5\boldsymbol{j}_5,3\boldsymbol{j}_5), (\mathbf{0},\boldsymbol{j}_5,2\boldsymbol{j}_5,3\boldsymbol{j}_5,4\boldsymbol{j}_5),
\end{aligned}
$$

where $\boldsymbol{j}_p = (1,\ldots,1)^\top \in \mathbb{R}^p$. The first two are the cases of equal means, the next three are the cases that $\sum_{i=1}^5 \boldsymbol{\mu}_i = \mathbf{0}$ and the last three are unbalanced cases.

For each estimator $\boldsymbol{\delta} = (\widehat{\boldsymbol{\mu}}_1,\ldots,\widehat{\boldsymbol{\mu}}_5)$, based on 5,000 replication of simulation, we obtain an approximated value of the risk function $R(\boldsymbol{\omega},\boldsymbol{\delta}) = E[L(\boldsymbol{\delta},\boldsymbol{\omega})]$ for the loss function given in (1.2). Table 1 reports the percentage relative improvement in average loss (PRIAL) of each estimator $\boldsymbol{\delta}$ over $\boldsymbol{\delta}^U = (\boldsymbol{X}_1,\ldots,\boldsymbol{X}_5)$, defined by

$$
\mathrm{PRIAL} = 100\{R(\boldsymbol{\omega},\boldsymbol{\delta}^U) - R(\boldsymbol{\omega},\boldsymbol{\delta})\}/R(\boldsymbol{\omega},\boldsymbol{\delta}^U).
$$

Table 1: Values of PRIAL of estimators PT, PT*, EB1, EB2, HB1 and HB2

$(\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_5)$	JS	PT	PT*	EB1	EB2	HB1	HB2
$(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0})$	54.1	74.1	87.2	70.3	83.4	69.7	82.6
$(2,2,2,2,2)\otimes\boldsymbol{j}_5$	9.0	74.1	74.4	70.3	70.6	69.7	70.0
$(-0.4,-0.2,0.0,0.2,0.4)\otimes\boldsymbol{j}_5$	48.3	64.1	75.3	63.6	74.9	64.7	76.5
$(-1,-0.5,0,0.5,1)\otimes\boldsymbol{j}_5$	37.4	17.4	23.0	40.6	46.2	44.6	52.2
$(-2,-1,0,1,2)\otimes\boldsymbol{j}_5$	25.0	−11.6	−10.0	17.5	19.1	17.5	18.9
$(1.2,1.4,1.6,1.8,2.0)\otimes\boldsymbol{j}_5$	12.7	64.1	64.6	63.6	64.2	64.7	65.2
$(1,1.5,2,2.5,3)\otimes\boldsymbol{j}_5$	9.5	17.4	17.9	40.6	41.1	44.6	44.9
$(0,1,2,3,4)\otimes\boldsymbol{j}_5$	18.9	−11.6	−10.8	17.5	18.3	17.5	17.6

It is noted from Theorem 2.3 that PT and EB1 can be improved on by PT* and EB2. These results can be confirmed by the simulation result in Table 1. Concerning the hierarchical Bayes estimators HB1 and HB2, Theorem 3.2 shows that the estimator HB2 is minimax, while we could not demonstrate that HB2 dominates HB1. However, Table 1 illustrates that HB2 is better than HB1.

Concerning PT, EB1 and HB1, their values of PRIAL are high under the hypothesis H_0 : $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_5$, but decrease when the means are far away from the hypothesis. The empirical Bayes estimator EB1 and the hierarchical Bayes estimator HB1 are comparable, and EB2 and HB2 are also comparable. The simulation results in these setups illustrate that the hierarchical Bayes estimator HB2 and the empirical Bayes estimator EB2 have good performances.

5 Proofs

We here provide the proofs of Theorems 3.1 and 3.2. Theorems 2.1 and 2.3 can be derived from Theorem 3.1 through Propositions 3.1 and 3.2.

5.1 Proof of Theorem 3.1

For the proof, the Stein identity due to Stein (1973, 81) and the chi-square identity due to Efron and Morris (1976) are useful. See also Bilodeau and Kariya (1989) for a multivariate version of the Stein identity.

Lemma 5.1 (1) *Assume that* $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^\top$ *is a p-variate random vector having* $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and that* $\boldsymbol{h}(\cdot)$ *is an absolutely continuous function from* $\mathbb{R}^p$ *to* $\mathbb{R}^p$. *Then, we have the Stein identity*

$$E[(\boldsymbol{Y}-\boldsymbol{\mu})^\top \boldsymbol{h}(\boldsymbol{Y})] = E\big[\mathrm{tr}\,\{\boldsymbol{\Sigma}\boldsymbol{\nabla}_{\boldsymbol{Y}}\boldsymbol{h}(\boldsymbol{Y})^\top\}\big], \tag{5.1}$$

provided the expectations in both sides exist, where $\boldsymbol{\nabla}_{\boldsymbol{Y}} = (\partial/\partial Y_1, \ldots, \partial/\partial Y_p)^\top$.

(2) *Assume that* S *is a random variable such that* $S/\sigma^2 \sim \chi_n^2$ *and that* $g(\cdot)$ *is an absolutely continuous function from* $\mathbb{R}$ *to* $\mathbb{R}$. *Then, we have the chi-square identity*

$$E[Sg(S)] = \sigma^2 E[ng(S) + 2Sg'(S)], \tag{5.2}$$

provided the expectations in both sides exist.

The risk function is decomposed as

$$\begin{aligned}
R(\boldsymbol{\omega},\widehat{\boldsymbol{\mu}}(\phi,\psi)) =& \sum_{i=1}^k E[(\widehat{\boldsymbol{\mu}}_i(\phi,\psi)-\boldsymbol{\mu}_i)^\top \boldsymbol{V}_i^{-1}(\widehat{\boldsymbol{\mu}}_i(\phi,\psi)-\boldsymbol{\mu}_i)/\sigma^2]\\
=&\sum_{i=1}^k E[(\boldsymbol{X}_i-\boldsymbol{\mu}_i)^\top \boldsymbol{V}_i^{-1}(\boldsymbol{X}_i-\boldsymbol{\mu}_i)/\sigma^2]\\
&-2\sum_{i=1}^k E\Big[(\boldsymbol{X}_i-\boldsymbol{\mu}_i)^\top \boldsymbol{V}_i^{-1}(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}})\frac{\phi}{\sigma^2 F}\Big] - 2\sum_{i=1}^k E\Big[(\boldsymbol{X}_i-\boldsymbol{\mu}_i)^\top \boldsymbol{V}_i^{-1}\widehat{\boldsymbol{\nu}}\frac{\psi}{\sigma^2 G}\Big]\\
&+\frac{1}{\sigma^2}\sum_{i=1}^k E\Big[\Big(\frac{\phi}{F}(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}})+\frac{\psi}{G}\widehat{\boldsymbol{\nu}}\Big)^\top \boldsymbol{V}_i^{-1}\Big(\frac{\phi}{F}(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}})+\frac{\psi}{G}\widehat{\boldsymbol{\nu}}\Big)\Big]\\
=&I_1 - 2I_2 - 2I_3 + I_4, \qquad \text{(say)}
\end{aligned} \tag{5.3}$$

for $\phi = \phi(F,G,S)$ and $\psi = \psi(F,G,S)$.

It is easy to see $I_1 = pk$. Let $\boldsymbol{\nabla}_i = \partial/\partial\boldsymbol{X}_i = (\partial/\partial X_{i1}, \ldots, \partial/\partial X_{ip})^\top$ for $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^\top$. It is noted that $\boldsymbol{\nabla}_i F = 2\boldsymbol{V}_i^{-1}(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}})/S$ and $\boldsymbol{\nabla}_i G = 2\boldsymbol{V}_i^{-1}\boldsymbol{A}\widehat{\boldsymbol{\nu}}/S$. For I_2, it is observed that

$$\begin{aligned}
I_2 =& \sum_{i=1}^k E\Big[\mathrm{tr}\,\Big[\boldsymbol{\nabla}_i\Big\{(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}})^\top\frac{\phi(F,G,S)}{F}\Big\}\Big]\Big]\\
=&\sum_{i=1}^k E\Big[\mathrm{tr}\,\Big[\Big\{\boldsymbol{\nabla}_i(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}})^\top\Big\}\frac{\phi(F,G,S)}{F}\Big] + \mathrm{tr}\,\Big[(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}})\Big\{\boldsymbol{\nabla}_i\frac{\phi(F,G,S)}{F}\Big\}^\top\Big]\Big]\\
=&\sum_{i=1}^k E\Big[\mathrm{tr}\,(\boldsymbol{I}-\boldsymbol{A}\boldsymbol{V}_i^{-1})\frac{\phi}{F} + \mathrm{tr}\,\Big[(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}})\Big\{\frac{2\boldsymbol{V}_i^{-1}(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}})}{S}\Big(-\frac{\phi}{F^2}+\frac{\phi_F}{F}\Big)+\frac{2\boldsymbol{V}_i^{-1}\boldsymbol{A}\widehat{\boldsymbol{\nu}}}{S}\frac{\phi_G}{F}\Big\}^\top\Big]\Big]\\
=&E\Big[p(k-1)\frac{\phi}{F} + 2F\Big(-\frac{\phi}{F^2}+\frac{\phi_F}{F}\Big) + \sum_{i=1}^k\frac{(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}})^\top \boldsymbol{V}_i^{-1}\boldsymbol{A}\widehat{\boldsymbol{\nu}}}{S}\frac{\phi_G}{F}\Big].
\end{aligned}$$

Noting that $\sum_{i=1}^k (\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}})^\top \boldsymbol{V}_i^{-1} = \boldsymbol{0}$, one gets

$$I_2 = E[\{p(k-1)-2\}\phi/F + 2\phi_F]. \tag{5.4}$$

Similarly, we have

$$\begin{aligned}
I_3 =& \sum_{i=1}^k E\Big[\mathrm{tr}\,\Big[\boldsymbol{\nabla}_i\Big\{\widehat{\boldsymbol{\nu}}^\top \frac{\psi(F,G,S)}{G}\Big\}\Big]\Big] \\
=& \sum_{i=1}^k E\Big[\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{V}_i^{-1})\frac{\psi}{G} + \mathrm{tr}\,\Big[\widehat{\boldsymbol{\nu}}\Big\{\frac{2\boldsymbol{V}_i^{-1}\widehat{\boldsymbol{\nu}}}{S}\Big(-\frac{\psi}{G^2}+\frac{\psi_G}{G}\Big) + \frac{2\boldsymbol{V}_i^{-1}(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}})}{S}\frac{\phi_F}{G}\Big\}^\top\Big]\Big].
\end{aligned}$$

Since $\sum_{i=1}^k \widehat{\boldsymbol{\nu}}^\top \boldsymbol{V}_i^{-1}(\boldsymbol{X}_i - \widehat{\boldsymbol{\nu}}) = \boldsymbol{0}$, one gets

$$I_3 = E[(p-2)\psi/G + 2\psi_G]. \tag{5.5}$$

Concerning I_4, it is noted that

$$\begin{aligned}
I_4 =& \frac{1}{\sigma^2}E\Big[\frac{S}{F}\phi^2(F,G,S)\Big] + \frac{1}{\sigma^2}E\Big[\frac{S}{G}\psi^2(F,G,S)\Big] + 2\frac{1}{\sigma^2}E\Big[\sum_{i=1}^k(\boldsymbol{X}_i-\widehat{\boldsymbol{\nu}})^\top\boldsymbol{V}_i^{-1}\widehat{\boldsymbol{\nu}}\frac{\phi\psi}{FG}\Big] \\
=& \frac{1}{\sigma^2}E\Big[\frac{S}{F}\phi^2(F,G,S)\Big] + \frac{1}{\sigma^2}E\Big[\frac{S}{G}\psi^2(F,G,S)\Big].
\end{aligned}$$

Using the chi-square identity, we have

$$\begin{aligned}
E\Big[\frac{S}{\sigma^2}\frac{\phi^2}{F}\Big] =& E\Big[\frac{n\phi^2}{F} + 2S\frac{\phi^2}{SF} + 2S\frac{2\phi}{F}\Big\{-\frac{F}{S}\phi_F - \frac{G}{S}\phi_G + \phi_G\Big\}\Big] \\
=& E\Big[(n+2)\frac{\phi^2}{F} - 4\frac{\phi}{F}\big(F\phi_F + G\phi_G\big) + 4\frac{S}{F}\phi\phi_S\Big].
\end{aligned}$$

Similarly,

$$E\Big[\frac{S}{\sigma^2}\frac{\psi^2}{G}\Big] = E\Big[(n+2)\frac{\psi^2}{G} - 4\frac{\psi}{G}\big(F\psi_F + G\psi_G\big) + 4\frac{S}{G}\psi\psi_S\Big].$$

Thus, one gets

$$I_4 = E\Big[(n+2)\frac{\phi^2}{F} - 4\frac{\phi}{F}\big(F\phi_F+G\phi_G\big) + 4\frac{S}{F}\phi\phi_S + (n+2)\frac{\psi^2}{G} - 4\frac{\psi}{G}\big(F\psi_F+G\psi_G\big) + 4\frac{S}{G}\psi\psi_S\Big]. \tag{5.6}$$

Combining (5.4), (5.5) and (5.6) gives the expression in Theorem 3.1.

5.2 Proof of Theorem 3.2

It suffices to show that $\phi^{HB2}(F,G,S)$ satisfies the conditions (a) and (b) in Proposition 3.1, and that $\psi^{HB2}(F,G,S)$ satisfies the conditions (a) and (b) in Proposition 3.2. For simplicity, let $h(x,y,v) = x^\alpha y^\beta v^\gamma \exp\{-v(x+y+1)/2\}$ for $\alpha = p(k-1)+a-1$, $\beta = p/2+b-1$ and $\gamma = (n+pk)/2 - c$. Then, $\phi^{HB2}(F,G,S)$ and $\psi^{HB2}(F,G,S)$ are written as

$$\begin{aligned}
\phi^{HB2}(F,G,S) =& \int_0^F\int_0^G\int_{LS}^\infty xh(x,y,v)dvdydx\Big/\int_0^F\int_0^G\int_{LS}^\infty h(x,y,v)dvdydx, \\
\psi^{HB2}(F,G,S) =& \int_0^F\int_0^G\int_{LS}^\infty yh(x,y,v)dvdydx\Big/\int_0^F\int_0^G\int_{LS}^\infty h(x,y,v)dvdydx
\end{aligned}$$

We shall check the conditions (a) and (b) in Proposition 3.1 for $\phi^{HB2}(F,G,S)$. The proof for $\psi^{HB2}(F,G,S)$ is omitted, because it can be shown similarly.

We begin by showing that $\phi^{HB2}(F,G,S)$ is increasing in F. The derivative of $\phi^{HB2}(F,G,S)$ with respect to F is proportional to

$$\begin{aligned}&\int_0^G\int_{LS}^\infty Fh(F,y,v)dvdy\times\int_0^F\int_0^G\int_{LS}^\infty h(x,y,v)dvdydx\\&\quad-\int_0^F\int_0^G\int_{LS}^\infty xh(x,y,v)dvdydx\times\int_0^G\int_{LS}^\infty h(F,y,v)dvdy\\&\quad=\int_0^F(F-x)\Big\{\int_0^G\int_{LS}^\infty h(F,y,v)dvdy\times\int_0^G\int_{LS}^\infty h(x,y,v)dvdy\Big\}dx,\end{aligned}$$

which is non-negative.

We next show that $\phi^{HB2}(F,G,S)$ is increasing in G. The derivative of $\phi^{HB2}(F,G,S)$ with respect to G is proportional to

$$\begin{aligned}&\int_0^F\int_{LS}^\infty xh(x,G,v)dvdx\times\int_0^F\int_0^G\int_{LS}^\infty h(x,y,v)dvdydx\\&\quad-\int_0^F\int_0^G\int_{LS}^\infty xh(x,y,v)dvdydx\times\int_0^F\int_{LS}^\infty h(x,G,v)dvdx\\&\quad=\Big\{\int_0^F\int_{LS}^\infty h(x,G,v)dvdx\Big\}^2\\&\quad\times\Big\{E^*[X]E\Big[\frac{\int_0^G\int_{LS}^\infty h(X,y,v)dvdy}{\int_{LS}^\infty h(X,G,v)dv}\Big]-E^*\Big[X\frac{\int_0^G\int_{LS}^\infty h(X,y,v)dvdy}{\int_{LS}^\infty h(X,G,v)dv}\Big]\Big\},\end{aligned}$$

where $E^*[\cdot]$ denote the expectation with respect to the probability

$$P(X\in A)=\int_A\int_{LS}^\infty h(x,G,v)dvdx/\int_0^F\int_{LS}^\infty h(x,G,v)dvdx.$$

Let $g(X)=\int_0^G\int_{LS}^\infty h(X,y,v)dvdy/\int_{LS}^\infty h(X,G,v)dv$. If $g(X)$ is decreasing in X, it is seen that $g(X)$ and X are monotone in opposite directions, which implies that $E^*[Xg(X)]-E^*[X]E^*[g(X)]\le 0$ from the covariance inequality. Thus, we need to show that $g(X)$ is decreasing in X. Note that $g(x)$ is written as

$$g(x)=\int_0^G\int_{LS}^\infty y^\beta v^\gamma e^{-v(x+y+1)/2}dvdy/\int_{LS}^\infty G^\beta v^\gamma e^{-v(x+G+1)/2}dv.$$

The derivative $g'(x)$ is proportional to

$$\begin{aligned}&-\int_0^G\int_{LS}^\infty y^\beta v^{\gamma+1}e^{-v(x+y+1)/2}dvdy\int_{LS}^\infty v^\gamma e^{-v(x+G+1)/2}dv\\&\quad+\int_0^G\int_{LS}^\infty y^\beta v^\gamma e^{-v(x+y+1)/2}dvdy\int_{LS}^\infty v^{\gamma+1}e^{-v(x+G+1)/2}dv\\&\quad=\Big\{\int_{LS}^\infty v^\gamma e^{-v(x+G+1)/2}dv\Big\}^2\Big\{-E^\dagger\Big[V\frac{\int_0^G y^\beta e^{-V(x+y+1)/2}dy}{e^{-V(x+G+1)/2}}\Big]+E^\dagger[V]E^\dagger\Big[\frac{\int_0^G y^\beta e^{-V(x+y+1)/2}dy}{e^{-V(x+G+1)/2}}\Big]\Big\}\\&\quad=\Big\{\int_{LS}^\infty v^\gamma e^{-v(x+G+1)/2}dv\Big\}^2\Big\{-E^\dagger\Big[V\int_0^G y^\beta e^{(G-y)V}dy\Big]+E^\dagger[V]E^\dagger\Big[\int_0^G y^\beta e^{(G-y)V}dy\Big]\Big\},\end{aligned}$$

where $E^{\dagger}[\cdot]$ denotes the expectation with respect to the probability

$$P(V \in A) = \int_A v^{\gamma} e^{-v(x+G+1)/2} dv \Big/ \int_{LS}^{\infty} v^{\gamma} e^{-v(x+G+1)/2} dv$$

for fixed x. Because $\int_0^G y^{\beta} e^{(G-y)V} dy$ is increasing in V for $y < G$, the covariance inequality implies that

$$-E^{\dagger}\Big[V \int_0^G y^{\beta} e^{(G-y)V} dy\Big] + E^{\dagger}[V] E^{\dagger}\Big[\int_0^G y^{\beta} e^{(G-y)V} dy\Big] \le 0,$$

which means that $g'(x) \le 0$. Hence, it is established that $\phi^{HB2}(F, G, S)$ is increasing in G.

We now show that $\phi^{HB2}(F, G, S)$ is decreasing in S. The derivative of $\phi^{HB2}(F, G, S)$ with respect to S is proportional to

$$\begin{aligned}
-L &\int_0^F \int_0^G xh(x, y, LS) dy dx \int_0^F \int_0^G \int_{LS}^{\infty} h(x, y, v) dv dy dx \\
&\quad + L \int_0^F \int_0^G \int_{LS}^{\infty} xh(x, y, v) dv dy dx \int_0^F \int_0^G h(x, y, LS) dy dx \\
&= L\Big\{ \int_0^F \int_0^G h(x, y, LS) dy dx \Big\}^2 \\
&\quad \times \Big\{ - E^{**}[X] \times E^{**}\Big[\frac{\int_0^G \int_{LS}^{\infty} h(X, y, v) dv dy}{\int_0^G h(X, y, LS) dy}\Big] + E^{**}\Big[X \frac{\int_0^G \int_{LS}^{\infty} h(X, y, v) dv dy}{\int_0^G h(X, y, LS) dy}\Big]\Big\},
\end{aligned}$$

where $E^{**}[\cdot]$ is the expectation with respect to the probability

$$P(X \in A) = \int_A \int_0^G h(x, y, LS) dy dx \Big/ \int_0^F \int_0^G h(x, y, LS) dy dx.$$

Note that

$$\frac{\int_0^G \int_{LS}^{\infty} h(X, y, v) dv dy}{\int_0^G h(X, y, LS) dy} = \frac{\int_0^G \int_{LS}^{\infty} y^{\beta} v^{\gamma} e^{(LS-v)X - v(y+1)/2} dv dy}{\int_0^G y^{\beta} (LS)^{\gamma} e^{-LS(y+1)/2} dy},$$

which is decreasing in X. Thus, X and $\int_0^G \int_{LS}^{\infty} h(X, y, v) dv dy / \int_0^G h(X, y, LS) dy$ are monotone in opposite directions, so that the derivative $\phi^{HB2}(F, G, S)$ with respect to S is negative due to the covariance inequality. Hence, the condition (a) is satisfied for $\phi^{HB2}(F, G, S)$.

Finally, from the condition (a), we see that

$$\begin{aligned}
\phi^{HB2}(F, G, S) &\le \lim_{F \to \infty} \lim_{G \to \infty} \lim_{S \to 0} \phi^{HB2}(F, G, S) \\
&= \frac{\int_0^{\infty} \int_0^{\infty} \int_0^{\infty} x^{\alpha+1} y^{\beta} v^{\gamma} \exp\{-y(x+y+1)/2\} dv dy dx}{\int_0^{\infty} \int_0^{\infty} x^{\alpha} y^{\beta} v^{\gamma} \exp\{-y(x+y+1)/2\} dv dy dx} \\
&= \frac{\Gamma(\alpha+2)\Gamma(\beta+1)\Gamma(\gamma-\alpha-\beta-2)}{\Gamma(\alpha+1)\Gamma(\beta+1)\Gamma(\gamma-\alpha-\beta-1)} = \frac{\alpha+1}{\gamma-\alpha-\beta-2} = \frac{p(k-1)+2a}{n-2(a+b+c)}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\psi^{HB2}(F,G,S) \leq & \lim_{F\to\infty}\lim_{G\to\infty}\lim_{S\to 0}\psi^{HB2}(F,G,S) \\
= & \frac{\int_0^\infty\int_0^\infty\int_0^\infty x^\alpha y^{\beta+1}v^\gamma \exp\{-y(x+y+1)/2\}dvdydx}{\int_0^\infty\int_0^\infty x^\alpha y^\beta v^\gamma \exp\{-y(x+y+1)/2\}dvdydx} \\
= & \frac{\Gamma(\alpha+1)\Gamma(\beta+2)\Gamma(\gamma-\alpha-\beta-2)}{\Gamma(\alpha+1)\Gamma(\beta+1)\Gamma(\gamma-\alpha-\beta-1)} = \frac{\beta+1}{\gamma-\alpha-\beta-2} = \frac{p+2b}{n-2(a+b+c)}.
\end{aligned}$$

Thus, the condition (b) in Proposition 3.1 is satisfied if $a+b+c<n/2$ and if

$$\frac{p(k-1)+2a}{n-2(a+b+c)} \leq 2\frac{p(k-1)-2}{n+2},$$

which is equivalently rewritten by (3.9). Also, the condition (b) in Proposition 3.2 is satisfied if $a+b+c<n/2$ and if

$$\frac{p+2b}{n-2(a+b+c)} \leq 2\frac{p-2}{n+2},$$

which leads to (3.10). Hence, the minimaxity of the hierarchical Bayes estimator is established.

Acknowledgments

Research of the second author was supported in part by Grant-in-Aid for Scientific Research (15H01943 and 26330036) from Japan Society for the Promotion of Science.

References

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

[2] B. Efron, C. Morris, Families of minimax estimators of the mean of a multivariate normal distribution, Ann. Statist. 4 (1976) 11–21.

[3] M. Ghosh, B.K. Sinha, Empirical and hierarchical Bayes competitors of preliminary test estimators in two sample problems, J. Multivariate Anal. 27 (1988) 206–227.

[4] W. James, C. Stein, Estimation with quadratic loss, in: Proc. Fourth Berkeley Sympos. Math. Statist. and Prob., Vol.I, Univ. California Press, Berkeley, 1961, pp.361–379.

[5] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, Proc. Third Berkeley Symp. Math. Statist. Probab. 1 (1956) 197–206, Univ. California Press, Berkeley.

[6] C. Stein, Estimation of the mean of a multivariate normal distribution, Ann. Statist. 9 (1981) 1135–1151.

[7] W.E. Strawderman, Proper Bayes minimax estimators of the multivariate normal mean, Ann. Math. Statist. 42 (1971) 385–388.

[8] W.E. Strawderman, Proper Bayes minimax estimators of the multivariate normal mean vector for the case of common unknown variances, Ann. Statist. 1 (1973) 1189–1194.

[9] L. Sun, Shrinkage estimation in the two-way multivariate normal model, Ann. Statist. 24 (1996) 825-840.