Shrinkage estimation with singular priors and an application to small area estimation

Ryumei Nakada[a], Tatsuya Kubokawa[b,*], Malay Ghosh[c], Sayar Karmakar[d]

[a]*Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[b]*Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[c]*Department of Statistics, University of Florida, 102 Griffin-Floyd Hall, Gainesville, Florida 32611*
[d]*Department of Statistics, University of Florida, 102 Griffin-Floyd Hall, Gainesville, Florida 32611*

Abstract

The paper considers estimation of the multivariate normal mean under a multivariate normal prior with a singular precision matrix. Such a setup appears in the multi-task averaging, serial and spatial smoothing problems. The empirical and hierarchical Bayes estimators shrink the maximum likelihood estimator by projecting it to the null space of the precision matrix. Conditions for minimaxity are given for the estimators proposed in this paper. The singular prior is applied to the Fay-Herriot small area estimation model with random effects having the singular distribution. Second-order approximation of the conditional mean squared error of the empirical Bayes estimator and its second-order unbiased estimator are derived. Numerical simulations confirm that the derived estimators perform well under the situation when there is a spatial correlation in the sample.

Keywords: Gaussian Markov random field prior, minimaxity, multi-task averaging, quadratic loss function, risk function, Stein estimator
2010 MSC: 62C20, 62F15, 62J07

1. Introduction

Stein [19], in his seminal paper, made the surprising discovery of the inadmissibility of the sample mean for estimating the multivariate normal mean in three or higher dimension under squared error loss. An explicit estimator dominating the sample mean was later provided by James and Stein [12]. These papers spurred a long and continued area of research which has not abated even today. A notable contribution in this regard is due to Efron and Morris, who in a series of articles (see for example, Efron and Morris [5]) provided an empirical Bayes interpretation of the James-Stein estimator. This led to a wider acceptance of the James-Stein estimator because of its applicability in real life problems.

The sample mean is a minimax estimator of the multivariate normal mean in any arbitrary dimension under a fairly general loss, including but not limited to the squared error loss. Thus any estimator dominating the sample mean also possesses the minimaxity property. Indeed, over the years, a long list of minimax estimators dominating the sample mean have been proposed under different scenarios.

James-Stein estimators, or its variants, are examples of shrinkage estimators, shrinking the sample mean towards a point or some surface. In the context of small area estimation, Fay and Herriot [6] proposed empirical Bayes shrinkage estimators which shrink area level sample means towards some regression vector.

The Fay-Herriot estimators, however, do not dominate the sample means in the original sense of Stein[19]. Indeed, to our knowledge, we are not aware of any estimators which dominate the sample mean under the Fay-Herriot model.

---

*Corresponding author. Email address: `tatsuya@e.u-tokyo.ac.jp`

Among others, we may refer to the estimators of Fay and Herriot [6], and also the ones proposed by Prasad and Rao [16] and Datta and Lahiri [3].

One of our original objectives was to revisit the Fay-Herriot model and find (if any) empirical or hierarchical Bayes estimators dominating the sample mean in a frequentist set up as considered by Stein and others. Our estimators are different from the ones existing in the literature. In the process, we were able to prove a result in a general framework going above and beyond the Fay-Herriot model. We can include as examples the multi-task averaging as proposed by Feldman, Gupta and Frigyik [7], the serial smoothing as considered by Yanagimoto and Yanagimoto [21], and even a spatial analogue of the latter. In addition, we are able to prove that our sufficient conditions ensuring dominance over the sample mean are exactly necessary for a wide range of empirical Bayes estimator and near necessary for the hierarchical Bayes estimator. Specifically, we have been able to prove that the unbiased estimators of the risks of our estimators are smaller than those of the sample mean under squared error loss.

The prior considered in this paper is Gaussian with a singular precision matrix. This occurs quite naturally in the Fay-Herriot model as well as in the examples mentioned earlier. Besides, our model includes the ones considered earlier in Besag, Green, Higdon and Mengersen [2], Ghosh, Natarajan, Stroud and Carlin [9] and Ghosh, Natarajan, Waller and Kim [10].

Some of the highlights of our findings are as follows. First, we have been able to show that our sufficient conditions for dominating the sample mean are very close to being necessary. This is quite different from the exiting literature where one finds only conditions sufficient for the dominance. Moreover, we show that the necessary and sufficient conditions for both empirical and hierarchical Bayes estimator are related to a single restriction on eigenvalue ratios of prior covariance matrix (cf. Remarks 1, 2 and inequality in (20)). Secondly, ours is the first result under the Fay-Herriot model where minimaxity of empirical Bayes estimators is provided. Interestingly, even in this random effect model necessary for small area estimation, the dominance condition is exactly the same as above. Overall both these results generalizes Stein's conditions for minimaxity for a general class of prior precision/covariance matrix.

The paper is organized as follows: Section 2 discusses the general framework of estimation of the multivariate normal mean under the improper singular prior model. Four examples of the singular prior distribution are also given. In Section 3, conditions for minimaxity of the empirical and hierarchical Bayes estimators are provided. Section 4 discusses the application of the singular prior model to small-area estimation, where area-level spatial information is built into random effects. Borrowing the arguments of Datta, Rao and Smith [4], we derive the second-order unbiased estimator of the conditional MSE. In Section 5, we investigate the performance of the empirical Bayes estimators through numerical and empirical studies. The concluding remarks are provided in Section 6. All the proofs are given in the Appendix.

2. Bayes Estimation for Singular Priors

2.1. Empirical and hierarchical Bayes estimators

Let y be an m-variate random vector such that the conditional distribution of y given θ is an m-variate normal distribution $\mathcal{N}_m(\theta, \Sigma)$, where θ is an m-variate mean vector and Σ is an $m \times m$ known covariance matrix. Consider an $m \times m$ non-negative definite and singular matrix C with rank r for $r < m$. Assume that θ has a prior distribution such that the conditional density of θ given τ^2 is $\pi(\theta \mid \tau^2, C)$, where

$$\pi(\theta \mid \tau^2, C) = \frac{c_0}{(2\pi)^{r/2}\tau^r} \exp\left(-\frac{1}{2\tau^2}\theta^\top C\theta\right), \tag{1}$$

where c_0^2 is the product of all the positive eigenvalues of C. It is noted that $\pi(\theta \mid \tau^2, C)$ is an improper prior, because the matrix C is singular. It is noted that the singular multivariate normal distribution suggested in the literature restricts the support to a subspace of θ, because of the integrability. This complicates Bayesian calculus, and the Bayes estimator is not written in a closed form. By using the improper singular prior (1) without a restricted support, the posterior distribution of θ given y is $\theta \mid y, \tau^2 \sim \mathcal{N}_m(\widehat{\theta}^B(\tau^2), (\Sigma^{-1} + \tau^{-2}C)^{-1})$, where $\widehat{\theta}^B(\tau^2)$ is the Bayes estimator of θ in the case of known τ^2, given by

$$\widehat{\theta}^B(\tau^2) = (\Sigma^{-1} + \tau^{-2}C)^{-1}\Sigma^{-1}y = y - (\tau^2\Sigma^{-1} + C)^{-1}Cy. \tag{2}$$

2

The marginal density of $\boldsymbol{y}$ is

$$f_\pi(\boldsymbol{y} \mid \tau^2) = \frac{c_0}{(2\pi)^{r/2}\tau^r |\boldsymbol{\Sigma}|^{1/2}|\boldsymbol{\Sigma}^{-1} + \tau^{-2}\boldsymbol{C}|^{1/2}} \exp\Big[-\frac{1}{2}\boldsymbol{y}^\top\{\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{-1} + \tau^{-2}\boldsymbol{C})^{-1}\boldsymbol{\Sigma}^{-1}\}\boldsymbol{y}\Big]. \tag{3}$$

To simplify the Bayes estimator and the marginal density, we consider the canonical form. Let $\boldsymbol{H}$ be an $m \times r$ matrix such that

$$\boldsymbol{\Sigma}^{1/2}\boldsymbol{C}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^\top, \ \boldsymbol{H}^\top\boldsymbol{H} = \boldsymbol{I}_r \ \text{and} \ \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_r).$$

Then, $(\tau^2\boldsymbol{I}_m + \boldsymbol{\Sigma}^{1/2}\boldsymbol{C}\boldsymbol{\Sigma}^{1/2})^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{C}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{H}(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\boldsymbol{H}^\top$, $|\boldsymbol{\Sigma}||\boldsymbol{\Sigma}^{-1} + \tau^{-2}\boldsymbol{C}| = |\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda}|/(\tau^{2r}|\boldsymbol{\Lambda}|)$ and $\boldsymbol{I}_m - (\boldsymbol{I}_m + \tau^{-2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{C}\boldsymbol{\Sigma}^{1/2})^{-1} = \boldsymbol{H}(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\boldsymbol{H}^\top$. Thus, the Bayes estimator (2) and the marginal density (3) can be rewritten as

$$\widehat{\boldsymbol{\theta}}^{\text{B}}(\tau^2) = \boldsymbol{y} - \boldsymbol{\Sigma}^{1/2}\boldsymbol{H}(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\boldsymbol{H}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}, \tag{4}$$

$$f_\pi(\boldsymbol{y} \mid \tau^2) = \frac{c_0|\boldsymbol{\Lambda}|^{1/2}}{(2\pi)^{r/2}|\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda}|^{1/2}} \exp\Big\{-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{H}(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\boldsymbol{H}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}\Big\}. \tag{5}$$

The unknown variance τ^2 can be estimated from the marginal distribution of $\boldsymbol{y}$. When estimator of τ^2 is denoted by $\hat{\tau}^2 = \hat{\tau}^2(\boldsymbol{y})$, the empirical Bayes estimator is given by

$$\widehat{\boldsymbol{\theta}}^{\text{EB}} = \widehat{\boldsymbol{\theta}}^{\text{B}}(\hat{\tau}^2) = \boldsymbol{y} - \boldsymbol{\Sigma}^{1/2}\boldsymbol{H}(\boldsymbol{I}_r + \hat{\tau}^2\boldsymbol{\Lambda})^{-1}\boldsymbol{H}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}. \tag{6}$$

Some specific estimators of τ^2 are given in Section 3.

The hierarchical Bayes estimator of $\boldsymbol{\theta}$ can be derived for the hierarchical prior distribution such that the first stage prior of $\boldsymbol{\theta}$ given τ^2 is given by (1) and the second stage prior of τ^2 is $\pi(\tau^2)$. Then from (4) and (5), we get the hierarchical Bayes estimator

$$\widehat{\boldsymbol{\theta}}^{\text{HB}} = \boldsymbol{y} - \boldsymbol{\Sigma}^{1/2}\boldsymbol{H}E[(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1} \mid \boldsymbol{y}]\boldsymbol{H}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{\Sigma}^{1/2}\boldsymbol{H}\frac{\int(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}f_\pi(\boldsymbol{y}, \tau^2)d\tau^2}{\int f_\pi(\boldsymbol{y}, \tau^2)d\tau^2}\boldsymbol{H}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}, \tag{7}$$

where $f_\pi(\boldsymbol{y}, \tau^2) = f_\pi(\boldsymbol{y} \mid \tau^2)\pi(\tau^2)$.

Finally, we give interesting expressions for the prior distribution (1) and the Bayes estimator (2). Let $\boldsymbol{P}_1$ be an $m \times r$ matrix such that

$$\boldsymbol{C} = \boldsymbol{P}_1\boldsymbol{\Lambda}_0^{-1}\boldsymbol{P}_1^\top \quad \text{and} \quad \boldsymbol{P}_1^\top\boldsymbol{P}_1 = \boldsymbol{I}_r, \tag{8}$$

and $\boldsymbol{\Lambda}_0^{-1}$ is a diagonal matrix of eigenvalues of $\boldsymbol{C}$. Let $\boldsymbol{P}_2$ be an $m \times (m - r)$ matrix such that $\boldsymbol{P}_2^\top\boldsymbol{P}_2 = \boldsymbol{I}_{m-r}$ and $\boldsymbol{P}_1^\top\boldsymbol{P}_2 = \boldsymbol{O}_{r \times (m-r)}$. It is noted that $\boldsymbol{P}_2$ is not unique.

Proposition 1. *Let $\boldsymbol{\Omega} = \boldsymbol{C}^+ + \boldsymbol{P}_2\boldsymbol{P}_2^\top = \boldsymbol{P}_1\boldsymbol{\Lambda}_0\boldsymbol{P}_1^\top + \boldsymbol{P}_2\boldsymbol{P}_2^\top$, where $\boldsymbol{C}^+$ is the Moore-Penrose inverse of $\boldsymbol{C}$.*

(i) *The singular prior distribution* (1) *is expressed as a mixture of a non-singular multivariate normal distribution and a uniform prior, namely,*

$$\boldsymbol{\theta} \mid \boldsymbol{\xi}, \tau^2 \sim \mathcal{N}_m(\boldsymbol{P}_2\boldsymbol{\xi}, \tau^2\boldsymbol{\Omega}), \quad \boldsymbol{\xi} \sim \text{Uniform}(\mathbb{R}^{m-r}). \tag{9}$$

(ii) *The Bayes estimator* (2) *is expressed as*

$$\widehat{\boldsymbol{\theta}}^{\text{B}}(\tau^2) = \boldsymbol{y} - \boldsymbol{\Sigma}(\boldsymbol{\Sigma} + \tau^2\boldsymbol{\Omega})^{-1}\{\boldsymbol{y} - \boldsymbol{P}_2\widehat{\boldsymbol{\xi}}(\tau^2)\},$$

where $\widehat{\boldsymbol{\xi}}(\tau^2) = \{\boldsymbol{P}_2^\top(\boldsymbol{\Sigma} + \tau^2\boldsymbol{\Omega})^{-1}\boldsymbol{P}_2\}^{-1}\boldsymbol{P}_2^\top(\boldsymbol{\Sigma} + \tau^2\boldsymbol{\Omega})^{-1}\boldsymbol{y}$.

2.2. Examples: the multi-task averaging, the smoothness prior and the Fay-Herriot model

We here provide four examples of the singular Bayesian model (1).

Example 1 (Multi-task averaging). The multi-task averaging was studied by Feldman, Gupta and Frigyik [7], [8]. Consider observable data y_{ij}'s for $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n_i\}$ with sampling errors σ_i^2's for $i \in \{1, \ldots, m\}$. Then, the problem of the multi-task averaging is to find the solution on θ_i's of minimizing

$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} \frac{(y_{ij} - \theta_i)^2}{\sigma_i^2} + \frac{\tau^2}{2} \sum_{a=1}^{m} \sum_{b=1}^{m} w_{ab}(\theta_a - \theta_b)^2, \tag{10}$$

where w_{ab} is a nonnegative constant satisfying $w_{ab} = w_{ba}$. The parameter τ^2 controls the penalty for sum of pairwise difference between θ_a and θ_b, and w_{ab} gives the strength of association between θ_a and θ_b. The $m \times m$ matrix $W = (w_{ij})$ is called the pairwise task similarity matrix in the context of the multi-task averaging. It is noted that the penalty term can be rewritten as

$$\sum_{a=1}^{m} \sum_{b=1}^{m} w_{ab}(\theta_a - \theta_b)^2 = 2\theta^\top C\theta, \tag{11}$$

where C is the matrix given by

$$C = (c_{ij}) = \text{diag}\Big(\sum_{j=1}^{m} w_{1j}, \ldots, \sum_{j=1}^{m} w_{pj}\Big) - W.$$

Since equation (11) holds for any θ and LHS of (11) is nonnegative for nonnegative weights w_{ab}, the matrix C is nonnegative definite. Thus, the function given in (10) is expressed as

$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} \frac{(y_{ij} - \theta_i)^2}{\sigma_i^2} + \tau^2 \theta^\top C\theta.$$

The optimal θ in term of minimizing this function is

$$\widehat{\theta}^{\text{MTA}} = \Big\{ I_m + \tau^2 \text{diag}\Big(\frac{\sigma_1^2}{n_1}, \ldots, \frac{\sigma_m^2}{n_m}\Big)C \Big\}^{-1} \overline{y}.$$

The variables y and Σ in Section 2.1 correspond to $\overline{y}$ and $\text{diag}(\sigma_1^2/n_1, \ldots, \sigma_m^2/n_m)$. When τ^2 is known, from (2), the optimal estimator $\widehat{\theta}^{\text{MTA}}$ is the Bayes estimator in the framework given in Section 2.1 for the prior distribution in (1).

It is noted that $\hat{\theta}^{\text{MTA}}$ has a similar form to the ridge-type empirical Bayes estimator given in Kubokawa and Srivastava [13], which treated the case that the covariance matrix Σ of y is ill-conditioned. In contrast, our paper handles the situation that the precision matrix C in the prior distribution of θ is singular.

Example 2 (Smoothing). Consider serial data (i, y_i), $i \in \{1, \ldots, m\}$, where $y = (y_1, \ldots, y_m)^\top$ is distributed as $\mathcal{N}_m(\theta, \sigma^2 I_m)$. For $\theta = (\theta_1, \ldots, \theta_m)^\top$, we consider the smoothness prior $\sum(\theta_i - 2\theta_{i+1} + \theta_{i+2})^2$, which has been studied in the literature. For example, see Yanagimoto and Yanagimoto [21]. Define C by

$$C = \begin{pmatrix} 1 & -2 & 1 & & & & \\ -2 & 5 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & 0 \\ & & & & & & \\ & 0 & & & & & \\ & & & & 1 & -2 & 1 \end{pmatrix}. \tag{12}$$

Then, it can be shown that $\sum(\theta_i - 2\theta_{i+1} + \theta_{i+2})^2 = \theta^\top C\theta$. In this case, the rank of C is $r = m - 2$. Let P_1 be an $m \times (m-2)$ matrix such that $C = P_1 \Lambda_0^{-1} P_1^\top$, $P_1^\top P_1 = I_{m-2}$ and Λ_0^{-1} is a diagonal matrix of eigenvalues of C. Then, we get the Bayes estimator in the framework given in Section 2.1 for the prior distribution

$$\pi(\theta \mid \tau^2, C) = \frac{1}{(2\pi)^{(m-2)/2} \tau^r |\Lambda_0|^{-1/2}} \exp\Big(-\frac{1}{2\tau^2} \theta^\top C\theta\Big).$$

4

From (2) and Proposition 1, the Bayes estimator of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}}^{\text{B}}(\tau^2) = \boldsymbol{y} - \sigma^2(\tau^2\boldsymbol{I}_m + \sigma^2\boldsymbol{C})^{-1}\boldsymbol{C}\boldsymbol{y} = \boldsymbol{y} - \sigma^2(\sigma^2\boldsymbol{I}_m + \tau^2\boldsymbol{\Omega})^{-1}\{\boldsymbol{y} - \boldsymbol{P}_2\widehat{\boldsymbol{\xi}}(\tau^2)\},$$

where $\boldsymbol{\Omega} = \boldsymbol{C}^+ + \boldsymbol{P}_2\boldsymbol{P}_2^\top$ and $\widehat{\boldsymbol{\xi}}(\tau^2) = \{\boldsymbol{P}_2^\top(\sigma^2\boldsymbol{I}_m + \tau^2\boldsymbol{\Omega})^{-1}\boldsymbol{P}_2\}^{-1}\boldsymbol{P}_2^\top(\sigma^2\boldsymbol{I}_m + \tau^2\boldsymbol{\Omega})^{-1}\boldsymbol{y}$. In this model, $\boldsymbol{P}_2 = (\boldsymbol{p}_{21}, \boldsymbol{p}_{22})$ are given by $\boldsymbol{p}_{21}^\top = (1, \ldots, 1)/\sqrt{m}$ and $\boldsymbol{p}_{22}^\top = (1 - (m+1)/2, \ldots, i - (m+1)/2, \ldots, m - (m+1)/2)/\sqrt{m(m+1)(m-1)/12}$.

Example 3 (Spatial smoothing). In this example, we consider a spatial analogue of the smoothing prior prorposed by Yanagimo and Yanagimoto [21]. For the simplicity of exposition we restrict ourselves to 2 dimensional spatial example but this is easily generalizable to higher dimensions. Note that, the serial data prior in Example 2 is motivated from the simple linear regression model $\theta_i = \mu + b_i$ as in such a case $\theta_i - 2\theta_{i+1} + \theta_{i+2}$ will be zero. Thus putting a penalty of $\sum_{i=1}^{m-2}(y_i - 2y_{i+1} + y_{i+2})^2$ will ensure that the model adheres to the linear model of the serial data.

The corresponding spatial analogue should adhere to $\theta_{ij} = \mu + a_i + b_j$, where θ_{ij} refers to the conditional mean of the observation in the (i, j)th cell. In view of the one-dimensional prior we now put a prior

$$\sum_{i=1}^{m}\sum_{j=1}^{m-2}(\theta_{ij} - 2\theta_{i,(j+1)} + \theta_{i,(j+2)})^2 + \sum_{i=1}^{m-2}\sum_{j=1}^{m}(\theta_{ij} - 2\theta_{(i+1),j} + \theta_{(i+2),j})^2 = \boldsymbol{\theta}^\top\boldsymbol{C}_2\boldsymbol{\theta}, \quad \text{(say)}$$

which ensures both in-row smoothness and in-column smoothness. Then it can be shown that $\boldsymbol{C}_2 = \boldsymbol{C} \otimes \boldsymbol{I}_m + \boldsymbol{I}_m \otimes \boldsymbol{C}$ for $\boldsymbol{C}$ defined in (12).

Example 4 (Fay-Herriot model with uniform prior for regression coefficients). The Fay-Herriot model used in small area estimation is $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v} + \boldsymbol{\epsilon}$, where $\boldsymbol{y}$ is an m-variate vector of observations, $\boldsymbol{X}$ is an $m \times p$ known matrix of covariates, and $\boldsymbol{\beta}$ is a p-variate vector of regression coefficients. Here, $\boldsymbol{v}$ and $\boldsymbol{\epsilon}$ are mutually independently distributed as $\boldsymbol{v} \sim \mathcal{N}_m(\boldsymbol{0}_m, \tau^2\boldsymbol{I}_m)$ and $\boldsymbol{\epsilon} \sim \mathcal{N}_m(\boldsymbol{0}_m, \boldsymbol{D})$ for $\boldsymbol{D} = \text{diag}(D_1, \ldots, D_m)$. Let $\boldsymbol{\theta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v}$. Assume that $\boldsymbol{\beta}$ is uniformly distributed over $\mathbb{R}^p$, namely $\boldsymbol{\beta} \sim \text{Uniform}(\mathbb{R}^m)$. Then, the marginal prior distribution of $\boldsymbol{\theta}$ is

$$\pi(\boldsymbol{\theta} \mid \tau^2, \boldsymbol{I}_m - \boldsymbol{P}_X) = \frac{1}{(2\pi)^{(m-p)/2}\tau^{m-p}|\boldsymbol{X}^\top\boldsymbol{X}|^{1/2}} \exp\left\{-\frac{1}{2\tau^2}\boldsymbol{\theta}^\top(\boldsymbol{I}_m - \boldsymbol{P}_X)\boldsymbol{\theta}\right\}, \tag{13}$$

where $\boldsymbol{P}_X = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$. This is a singular prior and belongs to the class (1) with $\boldsymbol{C} = \boldsymbol{I}_m - \boldsymbol{P}_X$. Let $\boldsymbol{P}_1$ be an $m \times (m-p)$ matrix such that $\boldsymbol{C} = \boldsymbol{P}_1\boldsymbol{P}_1^\top$. The matrices $\boldsymbol{\Lambda}_0^{-1}$ and $\boldsymbol{P}_2$ in Proposition 1 correspond to $\boldsymbol{\Lambda}_0 = \boldsymbol{I}_{p-m}$ and $\boldsymbol{P}_2 = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1/2}$. In this case, $\boldsymbol{\Omega} = \boldsymbol{I}_m$ and $\boldsymbol{P}_2\widehat{\boldsymbol{\xi}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\tau^2)$ for $\widehat{\boldsymbol{\beta}}(\tau^2) = \{\boldsymbol{X}^\top(\boldsymbol{D} + \tau^2\boldsymbol{I}_m)^{-1}\boldsymbol{X}\}^{-1}\boldsymbol{X}^\top(\boldsymbol{D} + \tau^2\boldsymbol{I}_m)^{-1}\boldsymbol{y}$. Thus, from Proposition 1, the Bayes estimator of $\boldsymbol{\theta}$ is $\widehat{\boldsymbol{\theta}}^{\text{B}}(\tau^2) = \boldsymbol{y} - (\tau^2\boldsymbol{D}^{-1} + \boldsymbol{C})^{-1}\boldsymbol{C}\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{D}(\boldsymbol{D} + \tau^2\boldsymbol{I}_m)^{-1}\{\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\tau^2)\}$, which is the best linear unbiased predictor of $\boldsymbol{\theta}$.

3. Conditions for Minimaxity

The maximum likelihood estimator $\boldsymbol{y}$ is minimax, but inadmissible for $m \geq 3$. Since this fact was discovered by Stein [19] and the improved estimator was derived in a closed form by James and Stein [12], many articles have developed deep, extensive and applicable studies. Most studies on minimaxity treat the homogeneous case which corresponds to $\lambda_1 = \cdots = \lambda_r$. We here provide conditions for minimaxity of the empirical and hierarchical Bayes estimators suggested in (6) and (7) when $\lambda_1, \ldots, \lambda_r$ are possibly different. The estimators are evaluated by the frequentist's risk function under the quadratic loss $R(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) = E[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top\boldsymbol{\Sigma}^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \mid \boldsymbol{\theta}]$. Without any loss of generality, assume that $\lambda_1, \ldots, \lambda_r$ are ordered as $\lambda_1 \geq \cdots \geq \lambda_r$ for $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_r)$. Conditions for minimaxity of the empirical Bayes estimators depend on estimator $\hat{\tau}^2$.

Let $\boldsymbol{z} = \boldsymbol{H}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}$. Since $\boldsymbol{z} \sim \mathcal{N}_r(\boldsymbol{0}_r, \boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})$, we have $E[\boldsymbol{z}^\top(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\boldsymbol{z}] = r$. This equality suggests the estimating equation $\boldsymbol{z}^\top(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\boldsymbol{z} = r$. Thus, by replacing $\boldsymbol{z}$ and r with $\boldsymbol{H}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}$ and a for a positive constant a suitably chosen, one gets the estimator $\hat{\tau}_1^2 = \max(\hat{\tau}_*^2, 0)$, where $\hat{\tau}_*^2$ is the solution of the equation

$$\boldsymbol{y}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{H}(\boldsymbol{I}_r + \hat{\tau}_*^2\boldsymbol{\Lambda})^{-1}\boldsymbol{H}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y} = a. \tag{14}$$

5

As another estimator, we can suggest the closed-form estimator

$$\hat{\tau}_2^2 = \boldsymbol{y}^\top \boldsymbol{\Sigma}^{-1/2} \boldsymbol{H} \boldsymbol{H}^\top \boldsymbol{\Sigma}^{-1/2} \boldsymbol{y}/a, \tag{15}$$

for a positive constant a. Using similar arguments as in Shinozaki and Chang [18] and Kubokawa and Strawderman [14], we derive conditions on a for minimaxity of the empirical Bayes estimators.

Proposition 2. *The empirical Bayes estimator $\widehat{\boldsymbol{\theta}}^{\mathrm{EB1}} = \widehat{\boldsymbol{\theta}}^{\mathrm{B}}(\hat{\tau}_1^2)$ for $\hat{\tau}_1^2$ in (14) is minimax if constant a satisfies the condition*

$$0 < a \le 2\Big(\sum_{i=1}^{r} \frac{\lambda_r}{\lambda_i} - 2 \Big). \tag{16}$$

Also, the empirical Bayes estimator $\widehat{\boldsymbol{\theta}}^{\mathrm{EB2}} = \widehat{\boldsymbol{\theta}}^{\mathrm{B}}(\hat{\tau}_2^2)$ for $\hat{\tau}_2^2$ in (15) is minimax if

$$0 < a \le 2\lambda_r \Big(\sum_{i=1}^{r} \frac{\lambda_r}{\lambda_i} - 2 \Big). \tag{17}$$

When $a = r$ for $\hat{\tau}_1^2$ and $a = \mathrm{tr}(\boldsymbol{\Lambda})$ for $\hat{\tau}_2^2$, conditions (16) and (17) are written as $\sum_{i=1}^{r}(\lambda_r/\lambda_i - 1/2) \ge 2$ and $\sum_{i=1}^{r}\{\lambda_r/\lambda_i - \lambda_i/(2\lambda_r)\} \ge 2$, both of which are satisfied for $r \ge 4$ in the case of $\lambda_1 \cdots = \lambda_r$. Although these conditions seem restrictive, it is noted that they are close to necessary conditions as demonstrated below.

Let $\widehat{\Delta}(\boldsymbol{z})$ be an unbiased estimator of the risk difference $\Delta = R(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{\mathrm{EB}}) - R(\boldsymbol{\theta}, \boldsymbol{y})$. The conditions (16) and (17) can be derived as sufficient conditions for $\widehat{\Delta}(\boldsymbol{z}) \le 0$ for all $\boldsymbol{z}$. It is interesting to remark that the conditions (16) and (17) are necessary and sufficient conditions for $\widehat{\Delta}(\boldsymbol{z}) \le 0$.

Proposition 3. *Let $\widehat{\Delta}^{\mathrm{EB1}}(\boldsymbol{z})$ and $\widehat{\Delta}^{\mathrm{EB2}}(\boldsymbol{z})$ be unbiased estimators of the risk differences for the empirical Bayes estimators $\widehat{\boldsymbol{\theta}}^{\mathrm{EB1}}$ and $\widehat{\boldsymbol{\theta}}^{\mathrm{EB2}}$. Then, condition (16) is necessary and sufficient for $\widehat{\Delta}^{\mathrm{EB1}}(\boldsymbol{z}) \le 0$, and condition (17) is necessary and sufficient for $\widehat{\Delta}^{\mathrm{EB2}}(\boldsymbol{z}) \le 0$.*

We also propose a general estimator $\widehat{\boldsymbol{\theta}}^{\mathrm{EBG}}$ based on the following general estimator of τ^2:

$$\hat{\tau}_G^2 = \sum_{i=1}^{r} a_i z_i^2, \tag{18}$$

for $\boldsymbol{z} = (z_1, \ldots, z_r)^\top$, where one practical choice of a_i could be proportional to $1/\lambda_i$. In the following proposition we derive a condition for the general estimator $\widehat{\boldsymbol{\theta}}^{\mathrm{EBG}} = \widehat{\boldsymbol{\theta}}^{\mathrm{B}}(\hat{\tau}_G^2)$ with $\hat{\tau}_G^2$ in (18).

Proposition 4. *The empirical Bayes estimator $\widehat{\boldsymbol{\theta}}^{\mathrm{EBG}} = \widehat{\boldsymbol{\theta}}^{\mathrm{B}}(\hat{\tau}_G^2)$ for $\hat{\tau}_G^2$ in (18) is minimax if*

$$0 < \frac{1}{a_{(r)}} \le 2\lambda_r \Big(\sum_{i=1}^{r} \frac{\lambda_r}{\lambda_i} - 2 \Big). \tag{19}$$

where $a_{(r)} = \min(a_1, \ldots, a_r)$. Moreover, for particular sequences with $a_1 \ge \cdots \ge a_r$ the condition in (19) is also necessary for an unbiased estimator of the risk difference to be negative.

Remark 1. When $\lambda_1 = \cdots = \lambda_r$, the conditions (16) and (17) are satisfied for $r > 2$, and improvements are guaranteed for large r. However, those sufficient conditions for improvement are a little restrictive in the case of different λ_i's with large λ_1 and small λ_r. Such undesirable properties of empirical Bayes estimators have been recognized by Shinozaki and Chang [18]. However, this restriction is minor for our examples in Section 2.2. For example, in the smoothing example, we have $\boldsymbol{C} = \boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^\top$ with $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$. The $\boldsymbol{C}$ matrix in (12) satisfies $\sum_{i=1}^{r} \lambda_r/\lambda_i > 2$ for $r > 6$. For the spatial smoothing example, the restriction on sample size weakens to any $r > 2$. We would also like to emphasize that

$$\sum_{i=1}^{r} \lambda_r/\lambda_i > 2 \tag{20}$$

6

holds for C matrices corresponding to non-square grids as well as long as the number of rows and number of columns are not too far apart from each other. This shows the applicability of the empirical Bayes estimator in a large class of real applications.

We next derive conditions for minimaxity of the hierarchical Bayes estimator (7). Minimaxity of hierarchical Bayes estimators has been studied by Berger and Robert [1] and Kubokawa and Strawderman [14] for slightly different priors. We here consider the prior density of τ^2 given by

$$\pi(\tau^2) \propto |I_r + \tau^2\Lambda|^{-a+1/2}I(\tau^2 \geq \tau_0^2), \tag{21}$$

where a and τ_0^2 are known non-negative constants and $I(\tau^2 \geq \tau_0^2)$ denotes the indicator function with $I(\tau^2 \geq \tau_0^2) = 1$ for $\tau^2 \geq \tau_0^2$. The hierarchical Bayes estimator is expressed as

$$\widehat{\boldsymbol{\theta}}^{\text{HB}} = \boldsymbol{y} - \Sigma^{1/2}H\frac{\int(I_r + \tau^2\Lambda)^{-1}f_\pi(\boldsymbol{y}, \tau^2)d\tau^2}{\int f_\pi(\boldsymbol{y}, \tau^2)d\tau^2}H^\top\Sigma^{-1/2}\boldsymbol{y},$$

where

$$f_\pi(\boldsymbol{y}, \tau^2) = \frac{c_0|\Lambda|^{1/2}I(\tau^2 > \tau_0^2)}{(2\pi)^{r/2}|I_r + \tau^2\Lambda|^a}\exp\Big\{-\frac{1}{2}\boldsymbol{y}^\top\Sigma^{-1/2}H(I_r + \tau^2\Lambda)^{-1}H^\top\Sigma^{-1/2}\boldsymbol{y}\Big\}. \tag{22}$$

Then, the condition for minimaxity of the hierarchical Bayes estimator is given in the following proposition.

Proposition 5. *The hierarchical Bayes estimator* (7) *with the prior* (21) *for* τ^2 *is minimax if a satisfies either of the following conditions:*

$$\frac{1}{r} < a \leq \frac{\sum_{i=1}^r(1 + \tau_0^2\lambda_1)/(1 + \tau_0^2\lambda_i) - 1}{r(2\lambda_1/\lambda_r - 1)}, \quad \frac{1}{r} < a \leq \frac{1}{2r}\Big(\sum_{i=1}^r\frac{\lambda_r}{\lambda_i} - \sum_{i=1}^r\frac{1}{1 + \tau_0^2\lambda_i}\Big). \tag{23}$$

A necessary condition for an unbiased estimator of the risk difference to be negative is

$$\frac{1}{r} < a \leq \frac{1}{r}\Big(\sum_{i=1}^r\frac{\lambda_r}{\lambda_i} - 1\Big). \tag{24}$$

Remark 2. There is a gap between (23) and (24). Our conjecture is this gap is due to the additional parameter τ_0 in our prior setting. In the special case of $\lambda_1 = \cdots = \lambda_r$, the first sufficient condition in (23) is identical to (24).

Also, note that (20) implies it is possible to choose a and τ_0 to satisfy either of equations in (23). One can also see that (24) implies (20). Thus our results are applicable for a large class of prior covariances where (20) is satisfied.

4. Small Area Estimation with Singular Priors

4.1. An area-level model and empirical Bayes estimation

Small area estimation has been studied actively and extensively in recent years. Small area refers to a small geographical area or a group for which little information is obtained from the sample survey. When only a few observations are available from a given small area, the direct estimator based only on the data from the small area is likely to be unreliable, so that the relevant supplementary information such as data from other related small areas is used via suitable linking models to increase the precision of the estimate. For good reviews and motivations, see Ghosh and Rao [11], Pfeffermann [15] and Rao and Molina [17].

The basic model used for analyzing area-level data is the Fay-Herriot model. In this model, the random effects represent area effects and are mutual independent between different areas. For incorporating relatedness or similarities between areas, we here extend the Fay-Heriot model to the case that the random effects have spatial correlations by assuming the singular prior distribution (1) or (9).

The area-level model we investigate is

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{v} + \boldsymbol{\epsilon}, \tag{25}$$

7

or component-wise $y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + v_i + \varepsilon_i$ for $i \in \{1, \ldots, m\}$, where $\boldsymbol{y} = (y_1, \ldots, y_m)^\top$, $\boldsymbol{v} = (v_1, \ldots, v_m)^\top$, $\boldsymbol{\epsilon} = (\varepsilon_1, \ldots, \varepsilon_m)^\top$, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)^\top$ is an $m \times p$ known matrix of covariates, and $\boldsymbol{\beta}$ is a p-variate vector of regression coefficients. Here, $\boldsymbol{\epsilon}$ is independent of $\boldsymbol{v}$ and distributed as $\mathcal{N}_m(\boldsymbol{0}_m, \boldsymbol{D})$ for $\boldsymbol{D} = \text{diag}(D_1, \ldots, D_m)$. We incorporate a spatial information into the distribution of the random effects $\boldsymbol{v}$. Like the multi-tast averaging, let w_{ij} be a strength of association between the areas i and j for $i, j \in \{1, \ldots, m\}$. Then, $\sum_{i=1}^m \sum_{j=1}^m w_{ij}(v_i - v_j)^2 = 2\boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{v}$ for

$$\boldsymbol{A} = (a_{ij}) = \text{diag}\Big(\sum_{j=1}^m w_{1j}, \ldots, \sum_{j=1}^m w_{mj} \Big) - (w_{ij}), \tag{26}$$

where $\text{rank}(\boldsymbol{A}) = r$. As a prior distribution of $\boldsymbol{v}$, we assume the density

$$\pi(\boldsymbol{v} \mid \tau^2, \boldsymbol{A}) = \frac{a_0}{(2\pi)^{r/2}\tau^r} \exp\Big(-\frac{1}{2\tau^2} \boldsymbol{v}^\top \boldsymbol{A} \boldsymbol{v} \Big), \tag{27}$$

where a_0^2 is the product of all the positive eigenvalues of $\boldsymbol{A}$. This is a singular normal distribution. Thus, the conditional distribution of $\boldsymbol{v}$ given $\boldsymbol{y}$ is $\boldsymbol{v} \mid \boldsymbol{y} \sim \mathcal{N}((\boldsymbol{D}^{-1} + \tau^{-2}\boldsymbol{A})^{-1}\boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), (\boldsymbol{D}^{-1} + \tau^{-2}\boldsymbol{A})^{-1})$ and the marginal density of $\boldsymbol{y}$ is

$$\frac{a_0}{(2\pi)^{r/2}\tau^r |\boldsymbol{I}_m + \tau^{-2}\boldsymbol{D}\boldsymbol{A}|^{1/2}} \exp\Big\{ -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{A}(\tau^2\boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}\boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \Big\}. \tag{28}$$

Let $\boldsymbol{\theta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v}$. Then, $\boldsymbol{\theta} \mid \boldsymbol{y} \sim \mathcal{N}(\widehat{\boldsymbol{\theta}}^{\text{B}}(\boldsymbol{\beta}, \tau^2), (\boldsymbol{D}^{-1} + \tau^{-2}\boldsymbol{A})^{-1})$, where $\widehat{\boldsymbol{\theta}}^{\text{B}}(\boldsymbol{\beta}, \tau^2)$ is the Bayes estimator given by

$$\widehat{\boldsymbol{\theta}}^{\text{B}}(\boldsymbol{\beta}, \tau^2) = \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{D}^{-1} + \tau^{-2}\boldsymbol{A})^{-1}\boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{y} - (\tau^2\boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}).$$

If $r > p$ and $\boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X}$ is non-singular, the MLE of $\boldsymbol{\beta}$ based on (28) is

$$\widehat{\boldsymbol{\beta}}(\tau^2) = \{\boldsymbol{X}^\top \boldsymbol{A}(\tau^2\boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}\boldsymbol{D}^{-1}\boldsymbol{X}\}^{-1}\boldsymbol{X}^\top \boldsymbol{A}(\tau^2\boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}\boldsymbol{D}^{-1}\boldsymbol{y}.$$

The resulting empirical Bayes estimator of $\boldsymbol{\theta}$ for known τ^2 is then

$$\widehat{\boldsymbol{\theta}}^{\text{EB}}(\tau^2) = \boldsymbol{y} - (\tau^2\boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}\boldsymbol{A}\{\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\tau^2)\}. \tag{29}$$

In the case of known τ^2, it is remarked that the empirical Bayes estimator $\widehat{\boldsymbol{\theta}}^{\text{EB}}(\tau^2)$ is identical to the hierarchial Bayes estimator against the uniform prior for $\boldsymbol{\beta}$. In fact, by the identity

$$(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{A}(\tau^2\boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}\boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$
$$= \{\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\tau^2)\}^\top \boldsymbol{A}(\tau^2\boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}\boldsymbol{D}^{-1}\{\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\tau^2)\} + \{\widehat{\boldsymbol{\beta}}(\tau^2) - \boldsymbol{\beta}\}^\top \boldsymbol{X}^\top \boldsymbol{A}(\tau^2\boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}\boldsymbol{D}^{-1}\boldsymbol{X}\{\widehat{\boldsymbol{\beta}}(\tau^2) - \boldsymbol{\beta}\},$$

it follows from (28) that with a uniform prior for $\boldsymbol{\beta}$,

$$\boldsymbol{\beta} \mid \boldsymbol{y}, \tau^2 \sim \mathcal{N}(\widehat{\boldsymbol{\beta}}(\tau^2), \{\boldsymbol{X}^\top \boldsymbol{A}(\tau^2\boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}\boldsymbol{D}^{-1}\boldsymbol{X}\}^{-1}).$$

Thus for known τ^2, the hierarchical Bayes estimator $\widehat{\boldsymbol{\theta}}^{\text{HB}}(\tau^2)$ is identical to the empirical Bayes estimator $\widehat{\boldsymbol{\theta}}^{\text{EB}}(\tau^2)$.

Since τ^2 is unknown, we estimate it from (28). To this end, writing $\boldsymbol{A}(\tau^2\boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}\boldsymbol{D}^{-1} = \boldsymbol{D}^{-1} - \boldsymbol{D}^{-1}(\boldsymbol{D}^{-1} + \tau^{-2}\boldsymbol{A})^{-1}\boldsymbol{D}^{-1}$ and as before $\boldsymbol{D}^{1/2}\boldsymbol{A}\boldsymbol{D}^{1/2} = \boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^\top$ with $\boldsymbol{H}^\top \boldsymbol{H} = \boldsymbol{I}_r$ and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_r)$, it follows after simplification $\boldsymbol{A}(\tau^2\boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}\boldsymbol{D}^{-1} = \boldsymbol{D}^{-1/2}\boldsymbol{H}(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\boldsymbol{H}^\top \boldsymbol{D}^{-1/2}$. Letting $\boldsymbol{z} = \boldsymbol{H}^\top \boldsymbol{D}^{-1/2}\boldsymbol{y}$ and $\widetilde{\boldsymbol{X}} = \boldsymbol{H}^\top \boldsymbol{D}^{-1/2}\boldsymbol{X}$, we have

$$\boldsymbol{z} \sim \mathcal{N}_r(\widetilde{\boldsymbol{X}}\boldsymbol{\beta}, \boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda}). \tag{30}$$

Note that $(\boldsymbol{z} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta})^\top (\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}(\boldsymbol{z} - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}) = \{\boldsymbol{z} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}(\tau^2)\}^\top (\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\{\boldsymbol{z} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}(\tau^2)\} + \{\widehat{\boldsymbol{\beta}}(\tau^2) - \boldsymbol{\beta}\}^\top \widetilde{\boldsymbol{X}}^\top (\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\{\widehat{\boldsymbol{\beta}}(\tau^2) - \boldsymbol{\beta}\}$ and $\{\boldsymbol{z} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}(\tau^2)\}^\top (\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\{\boldsymbol{z} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}(\tau^2)\} = \boldsymbol{z}^\top \widetilde{\boldsymbol{P}}(\tau^2)\boldsymbol{z}$ for

$$\widetilde{\boldsymbol{P}} = \widetilde{\boldsymbol{P}}(\tau^2) = (\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1} - (\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\widetilde{\boldsymbol{X}}\{\widetilde{\boldsymbol{X}}^\top (\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\widetilde{\boldsymbol{X}}\}^{-1}(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}. \tag{31}$$

Then, $\boldsymbol{z}^\top \widetilde{\boldsymbol{P}}(\tau^2)\boldsymbol{z} \sim \chi_{r-p}^2$. The parameter τ^2 can be estimated based on this distribution or the distribution of $\boldsymbol{z}$ in (30).

4.2. Measuring the uncertainty

We now evaluate the prediction error of the empirical Bayes estimators. It is noted that the joint distribution of (y, v) and the marginal distribution of y are not integrable due to the singularity of A given in (26). Let $\tilde{y} = \widetilde{H}^\top D^{-1/2} y$ for an $m \times (m - r)$ matrix $\widetilde{H}$ such that $\widetilde{H}^\top \widetilde{H} = I_{m-r}$ and $\widetilde{H}^\top H = O_{(m-r)\times r}$. Then y is decomposed as $(z, \tilde{y})$ for $z = H^\top D^{-1/2} y$, and the marginal density of y can be expressed as $f_\pi(y \mid \beta, \tau^2)dy = f(z \mid \beta, \tau^2)dzd\tilde{y}$, where $f(z \mid \beta, \tau^2)$ is the density of distribution (30) and does not depend on $\tilde{y}$. Since $\tilde{y}$ does not have an integrable density, we consider a conditional mean squared error matrix (MSE) given $\tilde{y}$, defined by

$$\mathbf{cMSE}(\beta, \tau^2, \widehat{\theta}) = E[(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)^\top \mid \tilde{y}].$$

To evaluate the conditional MSE of the empirical Bayes estimator asymptotically, we assume the following conditions on m, r, D, X and $\hat{\tau}^2$:

(A1) $m \to \infty$ and $r/m \to c$ for some constant $c \in (0, 1]$;

(A2) $0 < \min_{1\le j\le m}(D_j) \le D_i \le \max_{1\le j\le m}(D_j) < \infty$ for all $i \in \{1, \ldots, m\}$, where D_i does not depend on m;

(A3) $X^\top X/m$ converges to a positive definite matrix, and $\max_{1\le i\le m}\{x_i^\top(X^\top X)^{-1}x_i\} = O(m^{-1})$ for $X^\top = (x_1, \ldots, x_m)$;

(A4) $\hat{\tau}^2$ is an estimator of τ^2 such that $\hat{\tau}^2$ is a function of z, $\hat{\tau}^2(-z) = \hat{\tau}^2(z)$ and $\hat{\tau}^2(z - \widetilde{X}\alpha) = \hat{\tau}^2(z)$ for any α. Also, $\hat{\tau}^2$ satisfies that $\hat{\tau}^2 - \tau^2 = O_p(m^{-1/2})$, $E[(\hat{\tau}^2 - \tau^2)^2 \mid z_i, z_j] = E[(\hat{\tau}^2 - \tau^2)^2] + O_p(r^{-3/2})$ for $i, j = 1, \ldots, r$, and the bias $\mathrm{Bias}(\hat{\tau}^2)$ and variance $\mathrm{Var}(\hat{\tau}^2)$ are of order $O(r^{-1})$.

The second-order approximation of the conditional mean squared error of the empirical Bayes estimator and its second-order unbiased estimator are given under the conditions (A1)-(A4). Let

$$
\begin{aligned}
G_1(\tau^2) =& \tau^2(\tau^2 D^{-1} + A)^{-1}, \quad G_2(\tau^2) = (\tau^2 D^{-1} + A)^{-1}AX\{X^\top D^{-1}(\tau^2 D^{-1} + A)^{-1}AX\}^{-1}XA(\tau^2 D^{-1} + A)^{-1}, \\
G_3^*(\tau^2) =& D^{1/2}H(I_r + \tau^2\Lambda)^{-1}\Lambda(I_r + \tau^2\Lambda)^{-1}(I_r + \tau^2\Lambda)^{-1}H^\top D^{1/2}\mathrm{Var}(\hat{\tau}^2), \\
G_4(\tau^2) =& (\tau^2 D^{-1} + A)^{-1}A(\tau^2 D^{-1} + A)^{-1}\mathrm{Bias}(\hat{\tau}^2).
\end{aligned}
\tag{32}
$$

Proposition 6. *Under conditions* (A1)-(A4), *the conditional MSE of the empirical Bayes estimator* $\widehat{\theta}^{\mathrm{EB}}(\hat{\tau}^2)$ *is approximated as*

$$\mathbf{cMSE}(\tau^2, \widehat{\theta}^{\mathrm{EB}}(\hat{\tau}^2)) = G_1(\tau^2) + G_2(\tau^2) + G_3^*(\tau^2) + [O(m^{-3/2})]_{m\times m}, \tag{33}$$

where the notation $M = [O_p(m^{-3/2})]_{m\times m}$ *means that any* (i, j) *element of* $m \times m$ *matrix* M *is of order* $O_p(m^{-3/2})$. *Further, a second-order unbiased estimator of the conditional MSE is*

$$\mathbf{cmse}(\hat{\tau}^2, \widehat{\theta}^{\mathrm{EB}}(\hat{\tau}^2)) = G_1(\hat{\tau}^2) + G_2(\hat{\tau}^2) + 2G_3^*(\hat{\tau}^2) - G_4(\hat{\tau}^2). \tag{34}$$

In the context of small area estimation, one has interest in the empirical Bayes estimate of each small area and the corresponding MSE estimate. The conditional MSE estimate of $\hat{\theta}_i^{\mathrm{EB}}(\hat{\tau}^2)$ in the i-th area is provided from Proposition 6 as the (i, i)-element of $\mathbf{cmse}(\hat{\tau}^2, \widehat{\theta}^{\mathrm{EB}}(\hat{\tau}^2))$.

Related to the arguments of minimaxity in Section 3, we here investigate the risk function of the empirical Bayes estimators, where estimator $\widehat{\theta}$ is evaluated by the risk $E[L_w(\theta, \widehat{\theta})]$ for the weighted loss function

$$L_w(\theta, \widehat{\theta}) = (\widehat{\theta} - \theta)^\top D^{-1/2}HH^\top D^{-1/2}(\widehat{\theta} - \theta). \tag{35}$$

Proposition 7. *Under conditions* (A1)-(A4), *the following results hold:*

(i) $E[L_w(\theta, \widehat{\theta}^{\mathrm{EB}}(\hat{\tau}^2))] = \mathrm{tr}\,[\mathbf{cMSE}(\tau^2, \widehat{\theta}^{\mathrm{EB}}(\hat{\tau}^2))D^{-1}] - (m - r)$ *and* $E[L_w(\theta, y)] = r$;

(ii) $E[L_w(\theta, \widehat{\theta}^{\mathrm{EB}}(\hat{\tau}^2))] \le E[L_w(\theta, y)]$ *up to second-order if*

$$\sup_{\tau^2}\left\{\mathrm{Var}(\hat{\tau}^2)\mathrm{tr}\,[\{(I_r + \tau^2\Lambda)^{-1}\Lambda\}^2(I_r + \tau^2\Lambda)^{-1}](1 + \tau^2\lambda_r)\right\} \le \sum_{i=1}^r \frac{\lambda_r}{\lambda_i} - p, \tag{36}$$

where $\Lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_r)$ *with* $\lambda_1 \ge \cdots \ge \lambda_r$.

Typical estimators of τ^2 are the restricted maximum likelihood (REML) estimator $\hat{\tau}^2_{\text{RE}}$, the Fay-Herriot estimator $\hat{\tau}^2_{\text{FH}}$ and the Prasad-Rao estimator $\hat{\tau}^2_{\text{PR}}$, respectively, given by $\hat{\tau}^2_{\text{RE}} = \max(\hat{\tau}^2_*, 0)$, $\hat{\tau}^2_{\text{FH}} = \max(\hat{\tau}^2_{**}, 0)$ and $\hat{\tau}^2_{\text{PR}} = \max(0, [z^\top\{I_r - \widetilde{X}(\widetilde{X}^\top\widetilde{X})^{-1}\widetilde{X}^\top\}z - (r-p)]/r_*)$, where $r_* = \text{tr}[\{I_r - \widetilde{X}(\widetilde{X}^\top\widetilde{X})^{-1}\widetilde{X}^\top\}\Lambda]$, and $\hat{\tau}^2_*$ and $\hat{\tau}^2_{**}$ are the solutions of the equations

$$\{z - \widetilde{X\beta}(\tau^2_*)\}^\top(I_r + \tau^2_*\Lambda)^{-1}\Lambda(I_r + \tau^2_*\Lambda)^{-1}\{z - \widetilde{X\beta}(\tau^2_*)\} = \text{tr}\{\widetilde{P}(\tau^2)\Lambda\},$$

$$\{z - \widetilde{X\beta}(\tau^2_{**})\}^\top(I_r + \tau^2_{**}\Lambda)^{-1}\{z - \widetilde{X\beta}(\tau^2_{**})\} = r - p,$$

for $\widetilde{P}(\tau^2)$ given in (31). According to Prasad and Rao [16], Datta and Lahiri [3] and Datta, et al. [4], their asymptotic variances are $\text{Var}(\hat{\tau}^2_{\text{RE}}) \approx 2/\text{tr}[\{(I_r + \tau^2\Lambda)^{-1}\Lambda\}^2]$, $\text{Var}(\hat{\tau}^2_{\text{FH}}) \approx 2r/[\text{tr}\{(I_r + \tau^2\Lambda)^{-1}\Lambda\}]^2$ and $\text{Var}(\hat{\tau}^2_{\text{PR}}) \approx 2\text{tr}[\{(I_r + \tau^2\Lambda)\Lambda^{-1}\}^2]/r^2$. Note that $\text{tr}[\{(I_r + \tau^2\Lambda)^{-1}\Lambda\}^2(I_r + \tau^2\Lambda)^{-1}](1 + \tau^2\lambda_r) \leq \text{tr}[\{(I_r + \tau^2\Lambda)^{-1}\Lambda\}^2]$ for $\hat{\tau}^2_{\text{RE}}$.

For $\hat{\tau}^2_{\text{FH}}$ and $\hat{\tau}^2_{\text{PR}}$, note that, $\text{Var}(\hat{\tau}^2_{\text{FH}}) \leq \text{Var}(\hat{\tau}^2_{\text{PR}})$ since for positive quantities $\eta_1, \eta_2, \ldots \eta_r$, $(\sum_{i=1}^r \eta_i)^2 \sum_{k=1}^r \eta_k^{-2} \geq r^3$. With that, for $\hat{\tau}^2_{\text{PR}}$, LHS of (36) reads

$$\frac{2}{r^2}\sum_{i=1}^r \frac{(1+\tau^2\lambda_i)^2}{\lambda_i^2}\sum_{i=1}^r \frac{(1+\tau^2\lambda_r)\lambda_i^2}{(1+\tau^2\lambda_i)^3} \leq \frac{2}{r^2}\sum_{i=1}^r \frac{(1+\tau^2\lambda_i)^2}{\lambda_i^2}\sum_{i=1}^r \frac{\lambda_i^2}{(1+\tau^2\lambda_i)^2}$$

$$= \frac{2}{r^2}\left(r + \sum_{i=1}^r\sum_{j=1,j\neq i}^r \frac{(1+\tau^2\lambda_i)^2\lambda_j^2}{(1+\tau^2\lambda_j)^2\lambda_i^2}\right) \leq \frac{2}{r^2}\left(\frac{r^2+r}{2} + \sum_{i=1}^r\sum_{j=1,j<i}^r \frac{\lambda_j^2}{\lambda_i^2}\right)$$

in the light of the inequality $(1 + \tau^2\lambda_i)/(1 + \tau^2\lambda_j) \leq \max(1, \lambda_i/\lambda_j)$. Then from Proposition 7, we get conditions for the dominance.

Proposition 8. *Under conditions* (A1)-(A4)*, the following results hold:*
(i) $\widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2_{\text{RE}})$ *dominates* y *in second-order if* $p \leq \sum_{i=1}^r \lambda_r/\lambda_i - 2$;
(ii) $\widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2_{\text{FH}})$ *and* $\widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2_{\text{PR}})$ *dominate* y *in second-order if*

$$\frac{2}{r^2}\left(\frac{r^2+r}{2} + \sum_{i=1}^r\sum_{j=1,j<i}^r \frac{\lambda_j^2}{\lambda_i^2}\right) \leq \sum_{i=1}^r \lambda_r/\lambda_i - p.$$

Remark 3. Note that, for the special case of $\lambda_1 = \ldots = \lambda_r$ the conditions in (i) and (ii) above both coincide to Stein's condition for minimaxity $r - p \geq 2$. Also, it is interesting to note that the condition $p \leq \sum_{i=1}^r \lambda_r/\lambda_i - 2$ for the dominance of $\widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2_{\text{RE}})$ over y corresponds to the conditions for minimaxity given in Propositions 2 and 4 in the case of $p = 0$. Among the three estimators of τ^2, the usage of $\widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2_{\text{PR}})$ and $\widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2_{\text{FH}})$ is a bit restrictive due to the higher variance of $\hat{\tau}^2_{\text{PR}}$ and $\hat{\tau}^2_{\text{FH}}$. This comparison is also an interesting find in the light of minimaxity conditions.

5. Numerical Studies

5.1. Performance of the empirical Bayes estimators

In this section, we investigate the performance of the empirical Bayes estimators through simulation and empirical studies. We first compare the weighted mean squared errors of the empirical Bayes estimator $\widehat{\boldsymbol{\theta}}^{\text{EB}} = \widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2)$, given in (29), in the Fay-Herriot model (25). For the j-th component ε_j of the error term $\boldsymbol{\epsilon}$, we treat the following three distributions:

(A) $\varepsilon_j \sim \sqrt{D_j} \times \mathcal{N}(0,1)$,
(B) $\varepsilon_j \sim \sqrt{D_j} \times \sqrt{3/5}\text{t}(5)$,
(C) $\varepsilon_j \sim \sqrt{D_j} \times (\chi^2(2) - 2)/2$,

where t(5) and $\chi^2(2)$ are t-distribution with degrees of freedom 5 and χ^2-distribution with degrees of freedom 2, respectively. Note that the mean and the variance of each distribution are normalized to be 0 and 1.

In this simulation experiment, for simplicity, we utilize specific singular precision $A = \mathrm{diag}\,(I_r, O_{m-r})$ for the random effect v. For estimating τ^2, we employ the Prasad-Rao type estimator $\hat{\tau}^2_{\mathrm{PR}}$. The performance of estimator $\widehat{\theta}$ is measured in terms of the weighted mean squared error (MSE) given by wMSE $= E[L_w(\theta, \widehat{\theta})]/r$ for the weigted loss $L_w(\theta, \widehat{\theta})$ given in (35), or equivalently, $(1/r)\sum_{i=1}^{r} E[(\hat{\theta}_i - \theta_i)^2/D_i]$ in this setting, which measures the estimation error of the square integrable part of θ. This implies that the problem is reduced to the estimation of $\theta_1 = X_1\beta + v_1$, where $X^\top = (X_1^\top, X_2^\top)$ and $v^\top = (v_1^\top, v_2^\top)$ for $r \times p$ matrix X_1 and $r \times 1$ vector v_1. Since $v_1 \sim \mathcal{N}_r(0, \tau^2 D_1)$ for $D_1 = \mathrm{diag}\,(d_1, \ldots, d_r)$, one can generate values of v_1 from the normal distribution $\mathcal{N}_r(0, \tau^2 D_1)$.

The average values of the weighted MSE based on $1,000$ replications are reported in Table 1 for $m \in \{16, 64, 256\}$ and $r/m \in \{0.25, 0.5, 0.75, 1.0\}$, where $\beta = 0$, $\tau^2 = 1$ and $D_i = (i+1)/m$ for variance D_i of ε_i. It is observed from the table that the empirical Bayes estimator $\widehat{\theta}^{\mathrm{EB}}$ performs well even for t(5) and $\chi^2(2)$ distributions, namely it is robust against distributions with heavy tails or skewness. Also, the non-normal cases sometimes outperform the normal case. A similar observation was pointed out in the small area estimation by Datta, et al. [4]. It is noted that the weighted MSE decreases in the normal case as r/m approaches to one.

Table 1: Weighted mean squared errors for $m \in \{16, 64, 256\}$ and $r/m \in \{0.25, 0.5, 0.75, 1.0\}$ under normal, t(5) and $\chi^2(2)$ distributions for ϵ.

ϵ	normal				t(5)				$\chi^2(2)$			
r/m	0.25	0.5	0.75	1.0	0.25	0.5	0.75	1.0	0.25	0.5	0.75	1.0
$m = 16$	1.094	0.882	0.882	0.825	0.750	0.964	0.879	0.782	0.878	0.741	0.859	0.804
$m = 64$	0.991	0.952	0.935	0.876	0.923	0.953	0.905	0.879	0.932	0.964	0.900	0.799
$m = 256$	0.999	0.963	0.948	0.907	0.927	0.974	0.917	0.900	1.011	1.007	0.926	0.905

5.2. Performance under misspecifeid setups of covariance structures

We here investigate the performance of the empirical Bayes estimator in the case that the covariance structure in A is misspecified for the distribution of v. The model we examine is the Fay-Herriot model (25) with $m = 50$, $\tau^2 = 1$, $\beta = 0$ and $D_i = (i+1)/m$.

We treat the three cases of covariance structures of A: $A_1 = I_m$, and A_2 and A_3 are induced by (26) from the following weight matrices for pairwise differences: $W_2 = (w_{2ij})$, $W_3 = (w_{3ij})$ for

$$w_{2ij} := \frac{1}{|i-j|+1}, \quad w_{3ij} := \frac{1}{\exp(|i-j|)},$$

respectively. For clarification, we write $\widehat{\theta}^{\mathrm{EB}}_{(Aj)}$ as the empirical Bayes estimator $\widehat{\theta}^{\mathrm{EB}}$ which assumes the covariance structure A_j, where τ^2 is estimated by the Prasad-Rao type estimator $\hat{\tau}^2_{\mathrm{PR}}$. Note that the estimator $\widehat{\theta}^{\mathrm{EB}}_{(A1)}$ with $A = I_m$ corresponds to the usual small-area estimator under homoskedastic random effect v.

As the true covariance structure, we consider the following four cases:

Case 1: $G_1 = (g_{1ij})$, $g_{1ij} := \delta_{ij}$,
Case 2: $G_2 = (g_{2ij})$, $g_{2ij} := 0.4^{|i-j|}$,
Case 3: $G_3 = (g_{2ij})$, $g_{2ij} := 0.8^{|i-j|}$,
Case 4: $G_4 = (g_{3ij})$, $g_{3ij} := 0.95^{|i-j|}$,

where $\delta_{ij} = 1$ for $i = j$ and $= 0$ for $i \neq j$. Then, we can sample a random variable v from $\mathcal{N}_m(0_m, G_i)$ for $i \in \{1, \ldots, 4\}$. Since these distributions are not improper, we can evaluate the unconditional mean squared error $E[(\widehat{\theta} - \theta)^\top (\widehat{\theta} - \theta)]/m$. The average values of the mean squared errors of the empirical Bayes estimators $\widehat{\theta}^{\mathrm{EB}}_{(A1)}$, $\widehat{\theta}^{\mathrm{EB}}_{(A2)}$, $\widehat{\theta}^{\mathrm{EB}}_{(A3)}$ based on $1,000$ replications are reported in Table 2. It can be observed that when there is a strong spatial correlation in random effects, the empirical Bayes estimators assuming the models with singular random effects perform better. For larger

Table 2: MSE of three estimators $\widehat{\theta}^{EB}_{(A1)}, \widehat{\theta}^{EB}_{(A2)}, \widehat{\theta}^{EB}_{(A3)}$ for four cases $G_1, \ldots, G_4$ of covariance matrices in normal, $t(5)$ and $\chi^2(2)$ distributions.

ϵ	normal			$t(5)$			$\chi^2(2)$		
	$\widehat{\theta}^{EB}_{(A1)}$	$\widehat{\theta}^{EB}_{(A2)}$	$\widehat{\theta}^{EB}_{(A3)}$	$\widehat{\theta}^{EB}_{(A1)}$	$\widehat{\theta}^{EB}_{(A2)}$	$\widehat{\theta}^{EB}_{(A3)}$	$\widehat{\theta}^{EB}_{(A1)}$	$\widehat{\theta}^{EB}_{(A2)}$	$\widehat{\theta}^{EB}_{(A3)}$
G_1	**0.3167**	0.3227	0.4054	**0.3251**	0.3302	0.4099	**0.3155**	0.3216	0.3887
G_2	0.3149	**0.3085**	0.3380	0.3063	**0.3008**	0.3267	0.3335	**0.3225**	0.3411
G_3	0.2897	0.2515	**0.2100**	0.3131	0.2715	**0.2209**	0.3212	0.2816	**0.2220**
G_4	0.2411	0.1869	**0.1215**	0.2541	0.3132	**0.1214**	0.2567	0.2433	**0.1243**

Table 3: Values of similarity $\text{Sim}(A, G^{-1})$ of matrices G^{-1} and A for $A_1 = I_m$ and singular matrices A_2 and A_3 induced by (26) from $W_2 = (w_{2ij})$ and $W_3 = (w_{3ij})$.

	A_1	A_2	A_3
G_1	**0**	2.58	10.1
G_2	22.3	14.7	**3.65**
G_3	490	344	**86.2**
G_4	9470	6670	**1710**

correlation, $\widehat{\theta}^{EB}_{(A3)}$ performs well. This is because the structure of A_3 compared to A_2 is more similar to the structure of the precision whose inverse has the exponential-decay elements.

To explain the results given in Table 2, it is interesting to describe the distance between precision structures $\tau^{-2}A_i$ and G_j^{-1}. As such a measure, we consider the distance $\text{tr}\,(\tau^{-2}A - G^{-1})^2$. Minimizing $\text{tr}\,(\tau^{-2}A - G^{-1})^2$ with respect to τ^2 gives

$$\text{Sim}(A, G^{-1}) = \text{tr}\left\{\frac{\text{tr}\,(AG^{-1})}{\text{tr}\,(A^2)}A - G^{-1}\right\}^2. \tag{37}$$

We use this quantity as a measure of similarity of matrices $\tau^{-2}A$ and G^{-1}. The values of $\text{Sim}(A_i, G_j^{-1})$ for $i \in \{1, 2, 3\}$ and $j \in \{1, 2, 3, 4\}$ are given in Table 3. This result implies the structure of G_1^{-1} is similar to that of the matrix $\tau^{-2}A_1$. Similarly, G_2^{-1}, G_3^{-1} and G_4^{-1} are similar to $\tau^{-2}A_3$, respectively. For example, in the two cases G_3 and G_4, the performance of $\widehat{\theta}^{EB}_{(A3)}$ is better in Table 2, because the two cases are closer to $\tau^{-2}A_3$ in Table 3. In the case of $\tau^{-2}A_2$, however, $\widehat{\theta}^{EB}_{(A2)}$ is better than $\widehat{\theta}^{EB}_{(A3)}$ under G_2, while $\tau^{-2}A_3$ is closer to G_2^{-1} than $\tau^{-2}A_2$. It may be noted that $\text{Sim}(A, G^{-1})$ describes the distance between precision structures, but it does not measure the distributional discrepancy between the improper prior $\pi(v \mid \tau^2, A)$ in (27) and the proper prior $\pi(v \mid G)$ with nonsingular covariance matrix G.

5.3. An example

This example, primarily for illustration, uses the Fay-Herriot model (25) and data from the Survey of Family Income and Expenditure (SFIE) in Japan. The survey data are reported every month as average for $m = 47$ capital cities of prefectures, and we use the data of average spendings on 'education' and 'health' in November, 2019. It is noted that the sample sizes in SFIE are around 50 for each prefecture and thus is unreliable.

In the area-level model (25), the variances D_i are estimated from the past expenditure data from 2010 to 2018 and treated as known. Concerning the 'education' spending, as covariate x_i, we use data from the National Survey of Family Income and Expenditure (NSFIE) of year 2009, which is carried out every five years. The average spending data in NSFIE are more reliable than SFIE, because the sample sizes are much larger. The parameter τ^2 is estimated by the Prasad-Rao type estimator.

We here consider several covariance structures in A. Define d_{ij} and e_{ij} by the following:

$$d_{ij} = \text{distance between capital cities of prefectures } i \text{ and } j,$$

$$e_{ij} = I(\text{prefecture } i \text{ is adjacent to prefecture } j),$$

where $I(C) = 1$ if C is true, and $I(C) = 0$, otherwise. Based on d_{ij} and e_{ij}, we set the following five structures of A, where A_2, A_3, A_4 and A_5 are induced by (26) from the pairwise similarity matrices W_2, W_3, W_4 and W_5, respectively.

Case 1: $A = A_1 := I_m$,
Case 2: $A = A_2$, $W_2 = (w_{2ij})$, $w_{2ij} = 1/d_{ij}$ for $i \neq j$,
Case 3: $A = A_3$, $W_3 = (w_{3ij})$, $w_{3ij} = \exp(-d_{ij}/100)$ for $i \neq j$,
Case 4: $A = A_4$, $W_4 = (w_{4ij})$, $w_{4ij} = I(d_{ij} < 500) + I(d_{ij} < 1000)$ for $i \neq j$,
Case 5: $A = A_5$, $W_5 = (e_{ij})$.

Table 4: Estimates of 'education' spending with square roots of cMSE estimates where cMSE estimates are defined in (34) and y_i's are observed values. Values of $\widehat{\theta}^{EB}_{(A1)}, \ldots, \widehat{\theta}^{EB}_{(A5)}$ for selected prefectures and the corresponding square roots of cMSE estimates are given for $A_1 = I_m$, and singular matrices $A_2, \ldots, A_5$ induced from $W_2, \ldots, W_5$.

Pre	Mie	Shiga	Kyoto	Osaka	Hyogo	Nara	Wakayama
$D_i^{1/2}$	4922	5888	3933	6168	15211	10473	8102
y_i	21045	10563	19823	19872	11604	23310	5180
A_1	16861 (2376)	17718 (3121)	17642 (2737)	17297 (2935)	17421 (2775)	18281 (2669)	14481 (1936)
A_2	17015 (2164)	18198 (2645)	17729 (2282)	17358 (2264)	17748 (2464)	18475 (2554)	14694 (1842)
A_3	16521 (2255)	17191 (2961)	17137 (2591)	17128 (2685)	17108 (2649)	17726 (2618)	14706 (1861)
A_4	18049 (1956)	17911 (2724)	17861 (2326)	17298 (2358)	17628 (2362)	19238 (2416)	15322 (1542)
A_5	17387 (2565)	18150 (3133)	18031 (2998)	17874 (2908)	18100 (3444)	19332 (3960)	16120 (3070)

The empirical Bayes estimates and squared roots of estimates of cMSE with the observed y_i and squared root of dispersion $D_i^{1/2}$ are provided in Tables 4 for the 'education' spending and in Table 5 for the 'health' spending only for seven prefectures Mie, Shiga, Kyoto, Osaka, Hyogo, Nara and Wakayama in the Kansai region, where squared roots of estimates of cMSE are given inside the parentheses. It is observed that the cMSE estimates of the empirical Bayes estimators assuming the singular precision matrices for the random effect mostly dominates that of the usual small-area estimator under homoskedastic random effects. Also the total averages of square roots of cMSE estimates for individual prefectures are given in Table 6. The singular precision matrix A_4 gives the smallest total average of square roots of cMSE estimates, so that the structure of A_4 is recommendable among $A_1, \ldots, A_5$.

Table 5: Estimates of 'health' spending with square roots of cMSE estimates where cMSE estimates are defined in (34) and y_i's are observed values. Values of $\widehat{\theta}^{EB}_{(A1)}, \ldots, \widehat{\theta}^{EB}_{(A5)}$ for selected prefectures and the corresponding square roots of cMSE estimates are given for $A_1 = I_m$, and singular matrices $A_2, \ldots, A_5$ induced from $W_2, \ldots, W_5$.

Pre	Mie	Shiga	Kyoto	Osaka	Hyogo	Nara	Wakayama
$D_i^{1/2}$	2461	1217	9113	3939	1955	2542	2146
y_i	9542	12668	11575	6931	15220	12163	9909
A_1	11580 (1468)	12242 (1439)	11793 (1097)	11972 (1503)	12464 (1497)	12120 (1366)	11567 (1446)
A_2	11799 (1262)	12185 (1065)	11940 (936)	12080 (1149)	12373 (1206)	12126 (1116)	11806 (1259)
A_3	12510 (789)	12332 (824)	12339 (1047)	12579 (650)	12384 (724)	12383 (899)	12373 (827)
A_4	11526 (662)	12460 (699)	12031 (937)	11539 (552)	12164 (604)	11891 (765)	13085 (687)
A_5	11844 (1225)	11885 (1057)	11740 (1035)	11764 (1072)	11965 (1338)	11811 (1301)	11638 (1335)

13

Table 6: Total averages of square roots of cMSE estimates of $\widehat{\theta}^{\text{EB}}_{(A1)}, \ldots, \widehat{\theta}^{\text{EB}}_{(A5)}$, namely, values of $\text{tr}\{(\mathbf{cmse})^{1/2}\}/47$ for $\mathbf{cmse}$ defined in (34) are given for $A_1, \ldots, A_5$.

	A_1	A_2	A_3	A_4	A_5
'education'	2391	2486	2406	2211	3454
'health'	1415	1370	880	803	1449

6. Concluding Remarks

In estimation of the normal mean vector, we have considered the improper prior distributions with singular precision matrices which appear in the multi-task averaging, the smoothness prior and the Fay-Herriot model. Conditions for minimaxity of the empirical and hierarchical Bayes estimators have been derived. Their minimaxity is guaranteed by $\sum_{i=1}^{r} \lambda_r / \lambda_i > 2$ where $\lambda_1 \geq \cdots \geq \lambda_r$ are the eigenvalues of prior covariance matrix. Especially for the empirical Bayes estimators, these conditions are necessary and sufficient for the dominance.

As an important application of the singular prior model, we have discussed the Fay-Herriot model with spatial correlation in random effects. Assuming the singular precision for the random effects, we have provided the empirical Bayes estimator, the second-order approximation of the conditional mean squared error and the second-order unbiased estimator of cMSE. It is interesting to note that the second-order approximation of the weighted MSE of the empirical Bayes estimator with the REML estimator of τ^2 is smaller than that of y if $\sum_{i=1}^{r} \lambda_r / \lambda_i > 2$, which is the same condition as given above. The performance has been investigated through simulation and empirical studies, and it has been shown that the empirical Bayes estimators with singular priors have some improvements when there are strong spatial correlations.

A. Appendix

Proof of Proposition 1 Note that $\mathbf{\Omega} P_2 = P_2$ and $\mathbf{\Omega}^{-1} P_2 = P_2$, because $\mathbf{\Omega} = P_1 \Lambda_0 P_1^\top + P_2 P_2^\top$ and $\mathbf{\Omega}^{-1} = P_1 \Lambda_0^{-1} P_1^\top + P_2 P_2^\top$. Since $C = P_1 \Lambda_0^{-1} P_1^\top$ and $P_2^\top P_2 = I_{m-r}$, it is observed that

$$(\theta - P_2 \xi)^\top \mathbf{\Omega}^{-1} (\theta - P_2 \xi) = \theta^\top (P_1 \Lambda_0^{-1} P_1^\top + P_2 P_2^\top) \theta - 2\theta^\top P_2 \xi + \xi^\top \xi = (\xi - P_2^\top \theta)^\top (\xi - P_2^\top \theta) + \theta^\top C \theta,$$

which shows Part (i).

For Part (ii), it is observed that

$$
\begin{aligned}
&(\theta - P_2 \xi)^\top (\tau^2 \mathbf{\Omega})^{-1} (\theta - P_2 \xi) + (y - \theta)^\top \Sigma^{-1} (y - \theta) \\
&= \theta^\top \{\Sigma^{-1} + (\tau^2 \mathbf{\Omega})^{-1}\} \theta - 2\theta^\top (\Sigma^{-1} y + \tau^{-2} P_2 \xi) + \tau^{-2} \xi^\top P_2^\top P_2 \xi + y^\top \Sigma^{-1} y \\
&= \{\theta - \widehat{\theta}^{\text{B}}(\tau^2, \xi)\}^\top \{\Sigma^{-1} + (\tau^2 \mathbf{\Omega})^{-1}\} \{\theta - \widehat{\theta}^{\text{B}}(\tau^2, \xi)\} + (y - P_2 \xi)^\top (\Sigma + \tau^2 \mathbf{\Omega})^{-1} (y - P_2 \xi),
\end{aligned}
$$

for $\widehat{\theta}^{\text{B}}(\tau^2, \xi) = (\Sigma^{-1} + \tau^{-2} \mathbf{\Omega}^{-1})^{-1} (\Sigma^{-1} y + \tau^{-2} P_2 \xi)$. Thus, $\theta \mid y, \xi \sim \mathcal{N}(\widehat{\theta}^{\text{B}}(\tau^2, \xi), \{\Sigma^{-1} + (\tau^2 \mathbf{\Omega})^{-1}\}^{-1})$ and $E[\theta \mid y, \xi] = \widehat{\theta}^{\text{B}}(\tau^2, \xi)$. It is noted that $\{\Sigma^{-1} + (\tau^2 \mathbf{\Omega})^{-1}\}^{-1} = (\tau^2 \mathbf{\Omega})(\Sigma + \tau^2 \mathbf{\Omega})^{-1} \Sigma = \Sigma(\Sigma + \tau^2 \mathbf{\Omega})^{-1} (\tau^2 \mathbf{\Omega})$. Thus,

$$E[\theta \mid y, \xi] = (\tau^2 \mathbf{\Omega})(\Sigma + \tau^2 \mathbf{\Omega})^{-1} y + \Sigma(\Sigma + \tau^2 \mathbf{\Omega})^{-1} P_2 \xi = y - \Sigma(\Sigma + \tau^2 \mathbf{\Omega})^{-1} (y - P_2 \xi). \tag{38}$$

Also, $y \mid \xi \sim \mathcal{N}(P_2 \xi, \tau^2 \mathbf{\Omega} + \Sigma)$ and $\xi \sim$ Uniform. Writing

$$(y - P_2 \xi)^\top (\Sigma + \tau^2 \mathbf{\Omega})^{-1} (y - P_2 \xi) = \xi^\top P_2^\top (\Sigma + \tau^2 \mathbf{\Omega})^{-1} P_2 \xi - 2\xi^\top P_2^\top (\Sigma + \tau^2 \mathbf{\Omega})^{-1} y + y^\top (\Sigma + \tau^2 \mathbf{\Omega})^{-1} y,$$

we can see that $\boldsymbol{\xi} \mid \boldsymbol{y} \sim \mathcal{N}(\{\boldsymbol{P}_2^\top(\boldsymbol{\Sigma} + \tau^2\boldsymbol{\Omega})^{-1}\boldsymbol{P}_2\}^{-1}\boldsymbol{P}_2^\top(\boldsymbol{\Sigma} + \tau^2\boldsymbol{\Omega})^{-1}\boldsymbol{y}, \{\boldsymbol{P}_2^\top(\boldsymbol{\Sigma} + \tau^2\boldsymbol{\Omega})^{-1}\boldsymbol{P}_2\}^{-1})$. Hence,

$$E[\boldsymbol{\xi} \mid \boldsymbol{y}] = \{\boldsymbol{P}_2^\top(\boldsymbol{\Sigma} + \tau^2\boldsymbol{\Omega})^{-1}\boldsymbol{P}_2\}^{-1}\boldsymbol{P}_2^\top(\boldsymbol{\Sigma} + \tau^2\boldsymbol{\Omega})^{-1}\boldsymbol{y}. \tag{39}$$

Combine (38) and (39) to prove Part (ii) of Proposition 1. $\qquad\square$

Proof of Proposition 2 Let $\boldsymbol{z} = (z_1, \ldots, z_r)^\top = \boldsymbol{H}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}$ and $\boldsymbol{\eta} = \boldsymbol{H}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}$. It is noted that $\boldsymbol{z} \sim \mathcal{N}_r(\boldsymbol{\eta}, \boldsymbol{I}_r)$, and the risk difference of $\widehat{\boldsymbol{\theta}}^{\text{EB1}}$ and $\boldsymbol{y}$ can be written as $\Delta = R(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{\text{EB1}}) - R(\boldsymbol{\theta}, \boldsymbol{y}) = E[-2(\boldsymbol{z}-\boldsymbol{\eta})^\top(\boldsymbol{I}_r+\hat{\tau}_1^2\boldsymbol{\Lambda})^{-1}\boldsymbol{z}+\boldsymbol{z}^\top(\boldsymbol{I}_r+\hat{\tau}_1^2\boldsymbol{\Lambda})^{-2}\boldsymbol{z}]$. The Stein identity given in Stein [20] is used to rewrite the first term as $E[(\boldsymbol{z}-\boldsymbol{\eta})^\top(\boldsymbol{I}_r+\hat{\tau}_1^2\boldsymbol{\Lambda})^{-1}\boldsymbol{z}] = E[\boldsymbol{\nabla}_z^\top\{(\boldsymbol{I}_r+\hat{\tau}_1^2\boldsymbol{\Lambda})^{-1}\boldsymbol{z}\}]$ for $\boldsymbol{\nabla}_z = (\partial/\partial z_1, \ldots, \partial/\partial z_r)^\top$. Then, the risk difference is expressed as $\Delta = E[\widehat{\Delta}]$, where

$$\widehat{\Delta} = -2\sum_{i=1}^r \frac{1}{1 + \lambda_i\hat{\tau}_1^2} + 2\sum_{i=1}^r \frac{\lambda_i z_i}{(1 + \lambda_i\hat{\tau}_1^2)^2}\frac{\partial\hat{\tau}_1^2}{\partial z_i} + \sum_{i=1}^r \frac{z_i^2}{(1 + \lambda_i\hat{\tau}_1^2)^2}. \tag{40}$$

Note that $\hat{\tau}_1^2 = \max(\hat{\tau}_*^2, 0)$ and $\hat{\tau}_*^2$ is the solution of the equation $\sum_{i=1}^r z_i^2/(1 + \lambda_i\hat{\tau}_*^2) = a$. When $\hat{\tau}_*^2 < 0$, we have $\hat{\tau}_1^2 = 0$ and $a = \sum_{i=1}^r z_i^2/(1 + \lambda_i\hat{\tau}_*^2) > \sum_{i=1}^r z_i^2$. Thus, in the case of $\hat{\tau}_*^2 < 0$, it can be seen that $\widehat{\Delta} = -2r + \sum_{i=1}^r z_i^2 < -2r + a$, which is not positive if $a \leq 2r$. We next consider the case of $\hat{\tau}_*^2 > 0$. From the implicit function theorem, the partial derivative of $\hat{\tau}_*^2$ is

$$\frac{\partial\hat{\tau}_*^2}{\partial z_i} = 2\frac{z_i/(1 + \lambda_i\hat{\tau}_*^2)}{\sum_{j=1}^r \lambda_j z_j^2/(1 + \lambda_j\hat{\tau}_*^2)^2},$$

which is substituted into (40) to get

$$\widehat{\Delta} = -2\sum_{i=1}^r \frac{1}{1 + \lambda_i\hat{\tau}_*^2} + 4\frac{\sum_{i=1}^r \lambda_i z_i^2/(1 + \lambda_i\hat{\tau}_*^2)^3}{\sum_{j=1}^r \lambda_j z_j^2/(1 + \lambda_j\hat{\tau}_*^2)^2} + \sum_{i=1}^r \frac{z_i^2}{(1 + \lambda_i\hat{\tau}_*^2)^2}. \tag{41}$$

To evaluate the second term, we use the equalities

$$\sum_{i=1}^r \frac{\lambda_i z_i^2}{(1 + \lambda_i\hat{\tau}_*^2)^3} = \frac{1}{\hat{\tau}_*^2}\sum_{i=1}^r \frac{z_i^2}{(1 + \lambda_i\hat{\tau}_*^2)^2} - \frac{1}{\hat{\tau}_*^2}\sum_{i=1}^r \frac{z_i^2}{(1 + \lambda_i\hat{\tau}_*^2)^3},$$

$$\sum_{i=1}^r \frac{\lambda_i z_i^2}{(1 + \lambda_i\hat{\tau}_*^2)^2} = \frac{1}{\hat{\tau}_*^2}\sum_{i=1}^r \frac{z_i^2}{1 + \lambda_i\hat{\tau}_*^2} - \frac{1}{\hat{\tau}_*^2}\sum_{i=1}^r \frac{z_i^2}{(1 + \lambda_i\hat{\tau}_*^2)^2}.$$

We here demonstrate that

$$\frac{\sum_{i=1}^r z_i^2/(1 + \lambda_i\hat{\tau}_*^2)^2 - \sum_{i=1}^r z_i^2/(1 + \lambda_i\hat{\tau}_*^2)^3}{\sum_{i=1}^r z_i^2/(1 + \lambda_i\hat{\tau}_*^2) - \sum_{i=1}^r z_i^2/(1 + \lambda_i\hat{\tau}_*^2)^2} \leq \frac{\sum_{i=1}^r z_i^2/(1 + \lambda_i\hat{\tau}_*^2)^2}{\sum_{i=1}^r z_i^2/(1 + \lambda_i\hat{\tau}_*^2)}. \tag{42}$$

This inequality is equivalent to

$$\Big\{\sum_{i=1}^r \frac{z_i^2}{(1 + \lambda_i\hat{\tau}_*^2)^2}\Big\}^2 \leq \sum_{i=1}^r \frac{z_i^2}{1 + \lambda_i\hat{\tau}_*^2}\sum_{i=1}^r \frac{z_i^2}{(1 + \lambda_i\hat{\tau}_*^2)^3},$$

which can be verified by usisng the covariance inequality. Hence, we get the inequality (42). From (41), (42) and the definition of $\hat{\tau}_*^2$, it follows that

$$\widehat{\Delta} \leq -2\sum_{i=1}^r \frac{1}{1 + \lambda_i\hat{\tau}_*^2} + \Big(\frac{4}{a} + 1\Big)\sum_{i=1}^r \frac{z_i^2}{(1 + \lambda_i\hat{\tau}_*^2)^2}.$$

Since $\lambda_1 \geq \cdots \geq \lambda_r$, we have $\sum_{i=1}^r z_i^2/(1 + \lambda_i\hat{\tau}_*^2)^2 \leq a/(1 + \lambda_r\hat{\tau}_*^2)$, so that $\widehat{\Delta} \leq 0$ if

$$a \leq 2\Big(\sum_{i=1}^r \frac{1 + \lambda_r\hat{\tau}_*^2}{1 + \lambda_i\hat{\tau}_*^2} - 2\Big),$$

15

which leads to the condition in (16).

We next show the minimaxity of $\widehat{\theta}^{\text{EB2}}$, where $\hat{\tau}_2^2 = z^\top z/a$. Similar to (40), an unbiased estimator of the the risk difference is

$$\widehat{\Delta} = -2\sum_{i=1}^{r}\frac{1}{1+\lambda_i\hat{\tau}_2^2} + \frac{4}{a}\sum_{i=1}^{r}\frac{\lambda_i z_i^2}{(1+\lambda_i\hat{\tau}_2^2)^2} + \sum_{i=1}^{r}\frac{z_i^2}{(1+\lambda_i\hat{\tau}_2^2)^2}. \tag{43}$$

It is here observed that

$$\sum_{i=1}^{r}\frac{\lambda_i z_i^2}{(1+\lambda_i\hat{\tau}_2^2)^2} = \frac{1}{\hat{\tau}_2^2}\sum_{i=1}^{r}\frac{z_i^2}{1+\lambda_i\hat{\tau}_2^2} - \frac{1}{\hat{\tau}_2^2}\sum_{i=1}^{r}\frac{z_i^2}{(1+\lambda_i\hat{\tau}_2^2)^2} \le \frac{1}{\hat{\tau}_2^2}\sum_{i=1}^{r}\frac{z_i^2}{1+\lambda_i\hat{\tau}_2^2},$$

so that $\widehat{\Delta}$ is evaluated as

$$\begin{aligned}
\widehat{\Delta} \le & -2\sum_{i=1}^{r}\frac{1}{1+\lambda_i\hat{\tau}_2^2} + \frac{4}{a}\frac{1}{\hat{\tau}_2^2}\sum_{i=1}^{r}\frac{z_i^2}{1+\lambda_i\hat{\tau}_2^2} + \sum_{i=1}^{r}\frac{z_i^2}{(1+\lambda_i\hat{\tau}_2^2)^2} \\
\le & -2\sum_{i=1}^{r}\frac{1}{1+\lambda_i\hat{\tau}_2^2} + \frac{4}{a}\frac{1}{\hat{\tau}_2^2}\frac{a\hat{\tau}_2^2}{1+\lambda_r\hat{\tau}_2^2} + \frac{a\hat{\tau}_2^2}{(1+\lambda_r\hat{\tau}_2^2)^2}.
\end{aligned}$$

Hence, $\widehat{\Delta} \le 0$ if

$$a \le 2\lambda_r\Big(\sum_{i=1}^{r}\frac{1+\lambda_r\hat{\tau}_2^2}{1+\lambda_i\hat{\tau}_2^2} - 2\Big),$$

which leads to the condition in (17). $\qquad\square$

Proof of Proposition 3 For the empirical Bayes estimator $\widehat{\theta}^{\text{EB1}}$, we consider the unbiased estimator of the risk difference, denoted by $\widehat{\Delta}^{\text{EB1}}(z)$, given by (41). For a necessary condition, assume that $\widehat{\Delta}^{\text{EB1}}(z) \le 0$ for any z. Take $z_1 = \cdots = z_{r-1} = 0$ and $z_r = z$. Then,

$$0 \ge \widehat{\Delta}^{\text{EB1}}(0,\ldots,0,z) = -2\sum_{i=1}^{r}\frac{1}{1+\lambda_i\hat{\tau}_*^2} + 4\frac{1}{1+\lambda_r\hat{\tau}_*^2} + \frac{a}{1+\lambda_r\hat{\tau}_*^2},$$

because $z^2/(1+\lambda_r\hat{\tau}_*^2) = a$. Noting that $\hat{\tau}_*^2 \to \infty$ as $z \to \infty$, we can see that

$$0 \ge \lim_{z\to\infty}(1+\lambda_r\hat{\tau}_*^2)\widehat{\Delta}^{\text{EB1}}(0,\ldots,0,z) = -2\lim_{z\to\infty}\sum_{i=1}^{r}\frac{1+\lambda_r\hat{\tau}_*^2}{1+\lambda_i\hat{\tau}_*^2} + 4 + a = -2\sum_{i=1}^{r}\frac{\lambda_r}{\lambda_i} + 4 + a,$$

which yields the necessary condition given in (16).

For the empirical Bayes estimator $\widehat{\theta}^{\text{EB2}}$, from (43), the same arguments as used above give

$$0 \ge \widehat{\Delta}^{\text{EB2}}(0,\ldots,0,z) = -2\sum_{i=1}^{r}\frac{1}{1+\lambda_i\hat{\tau}_2^2} + \frac{4}{a}\frac{\lambda_r a\hat{\tau}_2^2}{(1+\lambda_r\hat{\tau}_2^2)^2} + \frac{a\tau_2^2}{(1+\lambda_r\hat{\tau}_2^2)^2},$$

because $\hat{\tau}_2^2 = z^2/a$. Thus,

$$0 \ge \lim_{z\to\infty}(1+\lambda_r\hat{\tau}_*^2)\widehat{\Delta}^{\text{EB2}}(0,\ldots,0,z) = -2\lim_{z\to\infty}\sum_{i=1}^{r}\frac{1+\lambda_r\hat{\tau}_*^2}{1+\lambda_i\hat{\tau}_*^2} + 4 + \frac{a}{\lambda_r} = -2\sum_{i=1}^{r}\frac{\lambda_r}{\lambda_i} + 4 + \frac{a}{\lambda_r},$$

which yields the necessary condition given in (17). $\qquad\square$

Proof of Proposition 4 Similar to (40), an unbiased estimator of the the risk difference is

$$\widehat{\Delta} = -2\sum_{i=1}^{r}\frac{1}{1+\lambda_i\hat{\tau}_G^2} + 4\sum_{i=1}^{r}\frac{a_i\lambda_i z_i^2}{(1+\lambda_i\hat{\tau}_G^2)^2} + \sum_{i=1}^{r}\frac{z_i^2}{(1+\lambda_i\hat{\tau}_G^2)^2}.$$

It is here observed that

$$\sum_{i=1}^{r} \frac{a_i \lambda_i z_i^2}{(1 + \lambda_i \hat{\tau}_G^2)^2} = \frac{1}{\hat{\tau}_G^2} \sum_{i=1}^{r} \frac{a_i z_i^2}{1 + \lambda_i \hat{\tau}_G^2} - \frac{1}{\hat{\tau}_G^2} \sum_{i=1}^{r} \frac{a_i z_i^2}{(1 + \lambda_i \hat{\tau}_G^2)^2} \le \frac{1}{\hat{\tau}_G^2} \sum_{i=1}^{r} \frac{a_i z_i^2}{1 + \lambda_i \hat{\tau}_G^2},$$

so that $\widehat{\Delta}$ is evaluated as

$$\widehat{\Delta} \le -2 \sum_{i=1}^{r} \frac{1}{1 + \lambda_i \hat{\tau}_G^2} + \frac{4}{\hat{\tau}_G^2} \sum_{i=1}^{r} \frac{a_i z_i^2}{a_i (1 + \lambda_i \hat{\tau}_G^2)} + \sum_{i=1}^{r} \frac{a_i z_i^2}{a_i (1 + \lambda_i \hat{\tau}_G^2)^2}$$

$$\le -2 \sum_{i=1}^{r} \frac{1}{1 + \lambda_i \hat{\tau}_G^2} + \frac{4}{a_{(r)}} \frac{1}{1 + \lambda_r \hat{\tau}_G^2} + \frac{\hat{\tau}_G^2}{a_{(r)} (1 + \lambda_r \hat{\tau}_G^2)^2}.$$

Hence, $\widehat{\Delta} \le 0$ if

$$\frac{1}{a_{(r)}} \le 2 \lambda_r \Big(\sum_{i=1}^{r} \frac{1 + \lambda_r \hat{\tau}_2^2}{1 + \lambda_i \hat{\tau}_2^2} - 2 \Big),$$

which leads to the condition in (19). For showing the necessary condition, recall that we assumed $a_1 \ge \ldots \ge a_r$. From (43), the same arguments as used above give

$$0 \ge \widehat{\Delta}^{\mathrm{EBG}}(0, \ldots, 0, z) = -2 \sum_{i=1}^{r} \frac{1}{1 + \lambda_i \hat{\tau}_G^2} + 4 \frac{\lambda_r \hat{\tau}_G^2}{(1 + \lambda_r \hat{\tau}_G^2)^2} + \frac{\hat{\tau}_G^2}{a_r (1 + \lambda_r \hat{\tau}_G^2)^2},$$

because $\hat{\tau}_G^2 = a_r z^2$. Thus,

$$0 \ge \lim_{z \to \infty} (1 + \lambda_r \hat{\tau}_*^2) \widehat{\Delta}^{\mathrm{EBG}}(0, \ldots, 0, z) = -2 \lim_{z \to \infty} \sum_{i=1}^{r} \frac{1 + \lambda_r \hat{\tau}_*^2}{1 + \lambda_i \hat{\tau}_*^2} + 4 + \frac{1}{a_r \lambda_r} = -2 \sum_{i=1}^{r} \frac{\lambda_r}{\lambda_i} + 4 + \frac{1}{a_r \lambda_r},$$

which yields the necessity of the condition given in (19). $\qquad \square$

Proof of Proposition 5 Let $z = (z_1, \ldots, z_r)^\top = H^\top \Sigma^{-1/2} y$. Then, the marginal density (22) of z is written as

$$f_\pi(z \mid \tau^2) = \frac{c_0 |\Lambda|^{1/2}}{(2\pi)^{r/2} |I_r + \tau^2 \Lambda|^a} \exp\Big\{ -\frac{1}{2} z^\top (I_r + \tau^2 \Lambda)^{-1} z \Big\} I(\tau^2 > \tau_0^2).$$

It is noted that $\int f_\pi(z \mid \tau^2) d\tau^2 < \infty$ when $ar > 1$. Using the same arguments as in the proof of Proposition 2, we observe that an unbiased estimator of the risk difference $\Delta = R(\theta, \widehat{\theta}^{\mathrm{HB}}) - R(\theta, y)$ is

$$\widehat{\Delta} = -2 \sum_{i=1}^{r} \frac{\partial}{\partial z_i} \Big\{ \frac{\int (1 + x\lambda_i)^{-1} f_\pi(z \mid x) dx}{\int f_\pi(z \mid x) dx} z_i \Big\} + \sum_{i=1}^{r} \Big\{ \frac{\int (1 + x\lambda_i)^{-1} f_\pi(z \mid x) dx}{\int f_\pi(z \mid x) dx} \Big\}^2 z_i^2.$$

It is observed that

$$\sum_{i=1}^{r} \frac{\partial}{\partial z_i} \Big\{ \frac{\int (1 + x\lambda_i)^{-1} f_\pi(z \mid x) dx}{\int f_\pi(z \mid x) dx} z_i \Big\} = \sum_{i=1}^{r} \frac{\int (1 + x\lambda_i)^{-1} f_\pi(z \mid x) dx}{\int f_\pi(z \mid x) dx}$$

$$- \sum_{i=1}^{r} \frac{\int (1 + x\lambda_i)^{-2} f_\pi(z \mid x) dx z_i^2}{\int f_\pi(z \mid x) dx} + \sum_{i=1}^{r} \Big\{ \frac{\int (1 + x\lambda_i)^{-1} f_\pi(z \mid x) dx}{\int f_\pi(z \mid x) dx} \Big\}^2 z_i^2,$$

so that $\widehat{\Delta}$ is rewritten as

$$\widehat{\Delta} = -2 \sum_{i=1}^{r} \frac{\int (1 + x\lambda_i)^{-1} f_\pi(z \mid x) dx}{\int f_\pi(z \mid x) dx} + 2 \sum_{i=1}^{r} \frac{\int (1 + x\lambda_i)^{-2} z_i^2 f_\pi(z \mid x) dx}{\int f_\pi(z \mid x) dx} - \sum_{i=1}^{r} \Big\{ \frac{\int (1 + x\lambda_i)^{-1} f_\pi(z \mid x) dx}{\int f_\pi(z \mid x) dx} \Big\}^2 z_i^2. \quad (44)$$

We evaluate the second and third terms in RHS of (44). Note that

$$\frac{d}{dx}\prod_{i=1}^{r}(1+x\lambda_i)^{-a} = -\sum_{i=1}^{r}\frac{a\lambda_i}{1+x\lambda_i}\prod_{i=1}^{r}(1+x\lambda_i)^{-a},$$

$$\frac{d}{dx}e^{-(1/2)\sum_{i=1}^{r}z_i^2/(1+x\lambda_i)} = \frac{1}{2}\sum_{i=1}^{r}\frac{\lambda_i z_i^2}{(1+x\lambda_i)^2}e^{-(1/2)\sum_{i=1}^{r}z_i^2/(1+x\lambda_i)}.$$

Using the integration by parts, we get

$$\int\sum_{i=1}^{r}\frac{\lambda_i z_i^2}{(1+x\lambda_i)^2}f_\pi(z\mid x)dx = -2f_\pi(z\mid\tau_0^2) + 2a\int\sum_{i=1}^{r}\frac{\lambda_i}{1+x\lambda_i}f_\pi(z\mid x)dx, \tag{45}$$

which is used to evaluate the second term in RHS of (44) as

$$\int\sum_{i=1}^{r}\frac{z_i^2}{(1+x\lambda_i)^2}f_\pi(z\mid x)dx \le \frac{1}{\lambda_r}\int\sum_{i=1}^{r}\frac{\lambda_i z_i^2}{(1+x\lambda_i)^2}f_\pi(z\mid x)dx$$

$$= -\frac{2}{\lambda_r}f_\pi(z\mid\tau_0^2) + \frac{2a}{\lambda_r}\int\sum_{i=1}^{r}\frac{\lambda_i}{1+x\lambda_i}f_\pi(z\mid x)dx. \tag{46}$$

The third term in RHS of (44) is rewritten as

$$\sum_{i=1}^{r}\left\{\frac{\int(1+x\lambda_i)^{-1}f_\pi(z\mid x)dx}{\int f_\pi(z\mid x)dx}\right\}^2 z_i^2$$

$$= \sum_{i=1}^{r}\frac{\int(1+x\lambda_i)^{-1}f_\pi(z\mid x)dx}{\{\int f_\pi(z\mid x)dx\}^2}\left\{\int\frac{z_i^2}{(1+x\lambda_i)^2}f_\pi(z\mid x)dx + \int\frac{\lambda_i z_i^2}{(1+x\lambda_i)^2}xf_\pi(z\mid x)dx\right\}$$

$$\ge \frac{\int(1+x\lambda_1)^{-1}f_\pi(z\mid x)dx}{\{\int f_\pi(z\mid x)dx\}^2}\left(\frac{1}{\lambda_1}I_1 + I_2\right),$$

where $I_1 = \int\sum_{i=1}^{r}\lambda_i z_i^2(1+x\lambda_i)^{-2}f_\pi(z\mid x)dx$ and $I_2 = \int\sum_{i=1}^{r}\lambda_i z_i^2(1+x\lambda_i)^{-2}xf_\pi(z\mid x)dx$. The term I_1 is evaluated by the integration-by-part (45). Similarly,

$$I_2 = -2\tau_0^2 f_\pi(z\mid\tau_0^2) - 2\int f_\pi(z\mid x)dx + 2a\int\sum_{i=1}^{r}\frac{x\lambda_i}{1+x\lambda_i}f_\pi(z\mid x)dx.$$

Hence,

$$\sum_{i=1}^{r}\left\{\frac{\int(1+x\lambda_i)^{-1}f_\pi(z\mid x)dx}{\int f_\pi(z\mid x)dx}\right\}^2 z_i^2$$

$$\ge \frac{\int(1+x\lambda_1)^{-1}f_\pi(z\mid x)dx}{\{\int f_\pi(z\mid x)dx\}^2}\left\{-2\frac{1+\tau_0^2\lambda_1}{\lambda_1}f_\pi(z\mid\tau_0^2) - 2\int f_\pi(z\mid x)dx + \frac{2a}{\lambda_1}\int\sum_{i=1}^{r}\frac{\lambda_i(1+x\lambda_1)}{1+x\lambda_i}f_\pi(z\mid x)dx\right\}. \tag{47}$$

Combining (44), (46) and (47), we have

$$\widehat{\Delta} \le \frac{1}{\int f_\pi(z\mid x)dx}\left\{-2\int\sum_{i=1}^{r}\frac{1}{1+x\lambda_i}f_\pi(z\mid x)dx - \frac{4}{\lambda_r}f_\pi(z\mid\tau_0^2) + \frac{4a}{\lambda_r}\int\sum_{i=1}^{r}\frac{\lambda_i}{1+x\lambda_i}f_\pi(z\mid x)dx\right\}$$

$$- \frac{\int(1+x\lambda_1)^{-1}f_\pi(z\mid x)dx}{\{\int f_\pi(z\mid x)dx\}^2}\left\{-2\frac{1+\tau_0^2\lambda_1}{\lambda_1}f_\pi(z\mid\tau_0^2) - 2\int f_\pi(z\mid x)dx + \frac{2a}{\lambda_1}\int\sum_{i=1}^{r}\frac{\lambda_i(1+x\lambda_1)}{1+x\lambda_i}f_\pi(z\mid x)dx\right\}.$$

18

Noting that

$$-\frac{4}{\lambda_r}\frac{f_\pi(z \mid \tau_0^2)}{\int f_\pi(z \mid x)dx} + \frac{2}{\lambda_1}\frac{\int (1 + \tau_0^2\lambda_1)/(1 + x\lambda_1)f_\pi(z \mid x)dx}{\{\int f_\pi(z \mid x)dx\}^2}f_\pi(z \mid \tau_0^2) \le 0,$$

we can see that $\widehat{\Delta} \le 0$ if

$$a\Big(\frac{2}{\lambda_r} - \frac{1}{\lambda_1}\Big)\sum_{i=1}^r \frac{\lambda_i(1 + x\lambda_1)}{1 + x\lambda_i} \le \sum_{i=1}^r \frac{1 + x\lambda_1}{1 + x\lambda_i} - 1,$$

for any $x > \tau_0^2$. Thus, we get the sufficient condition

$$\frac{1}{r} < a \le \frac{\sum_{i=1}^r (1 + \tau_0^2\lambda_1)/(1 + \tau_0^2\lambda_i) - 1}{r(2\lambda_1/\lambda_r - 1)}.$$

Another approach to investigating minimaxity is the super-harmonic property of the prior distribution by Stein [20]. Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_r)^\top = \boldsymbol{H}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}$. Then, the prior of $\boldsymbol{\xi}$ is

$$\pi(\boldsymbol{\xi}) = \int_{\tau_0^2}^\infty \frac{1}{\tau^r \prod_{i=1}^r (1 + \tau^2\lambda_i)^{a-1/2}}e^{-\sum_{i=1}^r \xi_i^2/(2\tau^2\lambda_i)}d\tau^2 = \int_0^{t_0} \frac{t^{ra-2}}{\prod_{i=1}^r (t + \lambda_i)^{a-1/2}}e^{-t\sum_{i=1}^r \xi_i^2/(2\lambda_i)}dt,$$

where $t = 1/\tau^2$ and $t_0 = 1/\tau_0^2$. It is noted that $\pi(\boldsymbol{\xi})$ exists for $ra > 1$. The super-harmonic condition of the prior is given by $\sum_{j=1}^r (\partial^2/\partial\xi_j^2)\pi(\boldsymbol{\xi}) \le 0$. It is observed that

$$\sum_{j=1}^r \frac{\partial^2\pi(\boldsymbol{\xi})}{\partial\xi_j^2} = \int_0^{t_0} \Big(-\sum_{i=1}^r \frac{1}{\lambda_i} + t\sum_{i=1}^r \frac{\xi_i^2}{\lambda_i^2}\Big)\frac{t^{ra-1}}{\prod_{i=1}^r (t + \lambda_i)^{a-1/2}}e^{-t\sum_{i=1}^r \xi_i^2/(2\lambda_i)}dt$$

$$\le \int_0^{t_0} \Big(-\sum_{i=1}^r \frac{1}{\lambda_i} + \frac{t}{\lambda_r}\sum_{i=1}^r \frac{\xi_i^2}{\lambda_i}\Big)\frac{t^{ra-1}}{\prod_{i=1}^r (t + \lambda_i)^{a-1/2}}e^{-t\sum_{i=1}^r \xi_i^2/(2\lambda_i)}dt. \qquad (48)$$

Using the integration by parts, we get

$$\frac{1}{\lambda_r}\int_0^{t_0}\sum_{i=1}^r \frac{\xi_i^2}{\lambda_i}\frac{t^{ra}}{\prod_{i=1}^r (t + \lambda_i)^{a-1/2}}e^{-t\sum_{i=1}^r \xi_i^2/(2\lambda_i)}dt$$

$$= \frac{1}{\lambda_r}\Big[-2\frac{t^{ra}}{\prod_{i=1}^r (t + \lambda_i)^{a-1/2}}e^{-t\sum_{i=1}^r \xi_i^2/(2\lambda_i)}\Big]_{t=0}^{t_0} + \frac{2}{\lambda_r}\int_0^{t_0}\Big\{ra - \Big(a - \frac{1}{2}\Big)\sum_{i=1}^r \frac{t}{t + \lambda_i}\Big\}\frac{t^{ra-1}}{\prod_{i=1}^r (t + \lambda_i)^{a-1/2}}e^{-t\sum_{i=1}^r \xi_i^2/(2\lambda_i)}dt.$$

Then from (48),

$$\sum_{j=1}^r \frac{\partial^2\pi(\boldsymbol{\xi})}{\partial\xi_j^2} \le \int_0^{t_0}\Big[-\sum_{i=1}^r \frac{1}{\lambda_i} + \frac{2}{\lambda_r}\Big\{ra - \Big(a - \frac{1}{2}\Big)\sum_{i=1}^r \frac{t}{t + \lambda_i}\Big\}\Big]\frac{t^{ra-1}}{\prod_{i=1}^r (t + \lambda_i)^{a-1/2}}e^{-t\sum_{i=1}^r \xi_i^2/(2\lambda_i)}dt,$$

which implies that $\sum_{j=1}^r (\partial^2/\partial\xi_j^2)\pi(\boldsymbol{\xi}) \le 0$ if

$$\Big(\sum_{i=1}^r \frac{\lambda_i}{t + \lambda_i}\Big)a \le \frac{1}{2}\sum_{i=1}^r \frac{\lambda_r}{\lambda_i} - \frac{1}{2}\sum_{i=1}^r \frac{t}{t + \lambda_i}, \qquad (49)$$

for any $0 < t < t_0 = 1/\tau_0^2$. An sufficient condition for (49) is

$$\frac{1}{r} \le a \le \frac{1}{2r}\Big(\sum_{i=1}^r \frac{\lambda_r}{\lambda_i} - \sum_{i=1}^r \frac{t_0}{t_0 + \lambda_i}\Big),$$

which is given in (23).

For the necessary condition, take $z_1 = \cdots = z_{r-1} = 0$ and $z_r = z$ in (44). For $z_0 = (0, \ldots, 0, z)$,

$$0 \geq \widehat{\Delta}(z_0) = -2 \sum_{i=1}^{r} \frac{\int (1 + x\lambda_i)^{-1} f_\pi(z_0 \mid x) dx}{\int f_\pi(z_0 \mid x) dx} + 2 \frac{\int (1 + x\lambda_r)^{-2} z^2 f_\pi(z_0 \mid x) dx}{\int f_\pi(z_0 \mid x) dx} - \left\{ \frac{\int (1 + x\lambda_r)^{-1} f_\pi(z_0 \mid x) dx}{\int f_\pi(z_0 \mid x) dx} \right\}^2 z^2,$$

where

$$f_\pi(z_0 \mid x) \propto \frac{1}{\prod_{i=1}^{r}(1 + x\lambda_i)^a} \exp\left(-\frac{1}{2} \frac{z^2}{1 + x\lambda_r}\right) I(x > \tau_0^2).$$

Making the transformation $u = z^2/\{2(1 + x\lambda_r)\}$, we can rewrite $\widehat{\Delta}(z_0)$ as

$$\widehat{\Delta}(z_0) = -2 \sum_{i=1}^{r} \int_0^{u_z} \frac{2\lambda_r u}{\lambda_i z^2 - 2(\lambda_i - \lambda_r)u} g(u) du \frac{1}{\int_0^{u_z} g(u) du} + \frac{8}{z^2} \frac{\int_0^{u_z} u^2 g(u) du}{\int_0^{u_z} g(u) du} - \frac{4}{z^2} \left\{ \frac{\int_0^{u_z} u g(u) du}{\int_0^{u_z} g(u) du} \right\}^2,$$

where $u_z = z^2/\{2(1 + \tau_0^2 \lambda_r)\}$ and

$$g(u) = \frac{1}{u^2} \prod_{i=1}^{r} \left\{ \frac{2\lambda_r u}{\lambda_i z^2 - 2(\lambda_i - \lambda_r)u} \right\}^a e^{-u}.$$

Thus,

$$0 \geq \lim_{z \to \infty} \widehat{\Delta}(z_0) = -4 \sum_{i=1}^{r} \frac{\lambda_r}{\lambda_i} \frac{\int_0^\infty u^{ar-1} e^{-u} du}{\int_0^\infty u^{ar-2} e^{-u} du} + 8 \frac{\int_0^\infty u^{ar} e^{-u} du}{\int_0^\infty u^{ar-2} e^{-u} du} - 4 \left(\frac{\int_0^\infty u^{ar-1} e^{-u} du}{\int_0^\infty u^{ar-2} e^{-u} du} \right)^2 = 4(ar - 1)\left(-\sum_{i=1}^{r} \frac{\lambda_r}{\lambda_i} + ar + 1 \right),$$

which yields the necessary condition (24). $\qquad \square$

Proof of Proposition 6 Following Prasad and Rao [16], Datta, et al. [4] and others, under condition (A4), the conditional MSE of $\widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2)$ is $\mathbf{cMSE}(\tau^2, \widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2)) = E[\{\widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2) - \boldsymbol{\theta}\}\{\widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2) - \boldsymbol{\theta}\}^\top \mid \tilde{\boldsymbol{y}}] = \boldsymbol{G}_1(\tau^2) + \boldsymbol{G}_2(\tau^2) + \boldsymbol{G}_3(\tau^2)$ for $\boldsymbol{G}_1(\tau^2) = E[\{\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{\text{B}}(\boldsymbol{\beta}, \tau^2)\}\{\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{\text{B}}(\boldsymbol{\beta}, \tau^2)\}^\top \mid \tilde{\boldsymbol{y}}]$, $\boldsymbol{G}_2(\tau^2) = E[\{\widehat{\boldsymbol{\theta}}^{\text{EB}}(\tau^2) - \widehat{\boldsymbol{\theta}}^{\text{B}}(\boldsymbol{\beta}, \tau^2)\}\{\widehat{\boldsymbol{\theta}}^{\text{EB}}(\tau^2) - \widehat{\boldsymbol{\theta}}^{\text{B}}(\boldsymbol{\beta}, \tau^2)\}^\top \mid \tilde{\boldsymbol{y}}]$ and $\boldsymbol{G}_3(\tau^2) = E[\{\widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2) - \widehat{\boldsymbol{\theta}}^{\text{EB}}(\tau^2)\}\{\widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2) - \widehat{\boldsymbol{\theta}}^{\text{EB}}(\tau^2)\}^\top \mid \tilde{\boldsymbol{y}}]$. For $\boldsymbol{G}_1(\tau^2)$, from the independence of z and $\tilde{\boldsymbol{y}}$, we have

$$\boldsymbol{G}_1(\tau^2) = E^z E^{\theta|(z, \bar{y})}[\{\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{\text{B}}(\boldsymbol{\beta}, \tau^2)\}\{\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{\text{B}}(\boldsymbol{\beta}, \tau^2)\}^\top \mid (z, \tilde{\boldsymbol{y}})]] = E^z[\tau^2(\tau^2 \boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}] = \tau^2(\tau^2 \boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}. \quad (50)$$

Since the terms inside the expectations of $\boldsymbol{G}_2(\tau^2)$ and $\boldsymbol{G}_3(\tau^2)$ do not depend on $\tilde{\boldsymbol{y}}$, the conditional expectations are written as the corresponding unconditional expectations. Using the same arguments as in Prasad and Rao [16], Datta, et al. [4] and others, we can see that $\boldsymbol{G}_2(\tau^2)$ is expressed as in (32) and $\boldsymbol{G}_3(\tau^2)$ is approximated as $\boldsymbol{G}_3^*(\tau^2) + O(m^{3/2})$ for $\boldsymbol{G}_3^*(\tau^2)$ given in (32). This yields (33).

For Part (ii), by the Taylor series expansion, it is noted that

$$\boldsymbol{G}_1(\hat{\tau}^2) = \boldsymbol{G}_1(\tau^2) + (\tau^2 \boldsymbol{D}^{-1} + \boldsymbol{A})^{-1} \boldsymbol{A} (\tau^2 \boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}(\hat{\tau}^2 - \tau^2)$$
$$- (\tau^2 \boldsymbol{D}^{-1} + \boldsymbol{A})^{-1} \boldsymbol{A} (\tau^2 \boldsymbol{D}^{-1} + \boldsymbol{A})^{-1} \boldsymbol{D}^{-1} (\tau^2 \boldsymbol{D}^{-1} + \boldsymbol{A})^{-1}(\hat{\tau}^2 - \tau^2)^2 + [O_p(m^{-3/2})]_{m \times m}.$$

Since $\hat{\tau}^2$ is a function of z and independent of $\tilde{\boldsymbol{y}}$, we have

$$E[\boldsymbol{G}_1(\hat{\tau}^2) \mid \tilde{\boldsymbol{y}}] = E[\boldsymbol{G}_1(\hat{\tau}^2)] = \boldsymbol{G}_1(\tau^2) + \boldsymbol{G}_4(\tau^2) - \boldsymbol{G}_3^*(\tau^2) + [O(m^{-3/2})]_{m \times m},$$

which leads to the second-order unbiased estimator given in (34). $\qquad \square$

Proof of Proposition 7 For Part (i), it is observed that $E[L_w(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2))] = \tau^2 \text{tr}(\boldsymbol{I}_r + \tau^2 \boldsymbol{\Lambda})^{-1} + g(\tau^2)$ for $g(\tau^2) = E[\|(\boldsymbol{I}_r + \hat{\tau}^2)^{-1}\{z - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}(\hat{\tau}^2)\} - (\boldsymbol{I}_r + \tau^2)^{-1}\{z - \widetilde{\boldsymbol{X}}\boldsymbol{\beta}(\tau^2)\}\|^2]$, where $\|\cdot\|$ is the Euclidean norm. On the other hand, it can be seen from (50) that $\text{tr}[\mathbf{cMSE}(\tau^2, \widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2)) \boldsymbol{D}^{-1}] = \tau^2 \text{tr}\{(\tau^2 \boldsymbol{D}^{-1} + \boldsymbol{A})^{-1} \boldsymbol{D}^{-1}\} + g(\tau^2)$. Since $\tau^2 \text{tr}\{(\tau^2 \boldsymbol{D}^{-1} + \boldsymbol{A})^{-1} \boldsymbol{D}^{-1}\} = \tau^2 \text{tr}(\boldsymbol{I}_r + \tau^2 \boldsymbol{\Lambda})^{-1} + \text{tr}(\widetilde{\boldsymbol{H}} \widetilde{\boldsymbol{H}}^\top)$ and $\text{tr}(\widetilde{\boldsymbol{H}} \widetilde{\boldsymbol{H}}^\top) = m - r$, one gets the relation $E[L_w(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2))] = \text{tr}[\mathbf{cMSE}(\tau^2, \widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2)) \boldsymbol{D}^{-1}] - (m - r)$. It is easy to see that $E[L_w(\boldsymbol{\theta}, \boldsymbol{y})] = r$

20

For Part (ii), from the result in Part (i) and the second-order approximation in Proposition 6, it follows that $\text{tr}\{\mathbf{cMSE}(\tau^2, \widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2))\boldsymbol{D}^{-1}\} \leq m$ up to second order if and only if

$$\text{Var}(\hat{\tau}^2)\text{tr}\left[\{(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}\}^2(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\right]$$

$$\leq \text{tr}\{(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\} - \text{tr}\left[\{\widetilde{\boldsymbol{X}}^{\top}(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\widetilde{\boldsymbol{X}}\}^{-1}\widetilde{\boldsymbol{X}}^{\top}(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-2}\widetilde{\boldsymbol{X}}\right]. \tag{51}$$

Noting that $\text{tr}\left[\{\widetilde{\boldsymbol{X}}^{\top}(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\widetilde{\boldsymbol{X}}\}^{-1}\widetilde{\boldsymbol{X}}^{\top}(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-2}\widetilde{\boldsymbol{X}}\right] \leq p/(1 + \tau^2\lambda_r)$ and $\text{tr}\{(\boldsymbol{I}_r + \tau^2\boldsymbol{\Lambda})^{-1}\}(1 + \tau^2\lambda_r) \geq \sum_{i=1}^{r}\lambda_r/\lambda_i$, we can see that RHS in (51) is larger than $(\sum_{i=1}^{r}\lambda_i/\lambda_r - p)/(1 + \tau^2\lambda_r)$. Thus, $\widehat{\boldsymbol{\theta}}^{\text{EB}}(\hat{\tau}^2)$ dominates $\boldsymbol{y}$ up to second order under the condition (36). $\qquad\square$

References

[1] J.O. Berger, C. Robert, Subjective hierarchical Bayes estimation of a multivariate normal mean: on the frequentist interface, Ann. Statist. 18 (1990) 617–651.

[2] J. Besag, P. Green, D. Higdon, K. Mengersen, Bayesian computation and stochastic systems, Statist. Sci. 10 (1995) 3–66.

[3] G.S. Datta, P. Lahiri, A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems, Statist. Sinica 10 (2000) 613–627.

[4] G.S. Datta, J.N.K. Rao, D.D. Smith, On measuring the variability of small area estimators under a basic area level model, Biometrika 92 (2005) 183–196.

[5] B. Efron, C. Morris, Stein's estimation rule and its competitors - An empirical Bayes approach, J. Amer. Statist. Assoc. 68 (1973) 117–130.

[6] R.E. Fay, R.A. Herriot, Estimates of income for small places: an application of James-Stein procedures to census data, J. Amer. Statist. Assoc. 74 (1979) 269–277.

[7] S. Feldman, M.R. Gupta, B.A. Frigyik, Multi-task averaging, Advances Neural Information Processing Systems (NIPS), (2012).

[8] S. Feldman, M.R. Gupta, B.A. Frigyik, Revisiting Stein's paradox: multi-task averaging, J. Machine Learn. Res. 15 (2014) 3621–3662.

[9] M. Ghosh, K. Natarajan, T.W.F. Stroud, B.P. Carlin, Generalized linear models for small-area estimation, J. Amer. Statist. Assoc. 93 (1998) 273–282.

[10] M. Ghosh, K. Natarajan, L.A. Waller, D. Kim, Hierarchical Bayes GLMs for the analysis of spatial data: An application to disease mapping, J. Statist. Plann. Inf. 75 (1999) 305–318.

[11] M. Ghosh, J.N.K. Rao, Small area estimation: an appraisal, Statist. Science 9 (1994) 55–76.

[12] W. James, C. Stein, Estimation with quadratic loss, Proc. Fourth Berkeley Symp. Math. Statist. Probab. 1 (1961) 361–379 University of California Press, Berkeley.

[13] T. Kubokawa, M.S. Srivastava, Improved empirical Bayes ridge regression estimators under multicollinearity, Com. Statist. Theory Methods 33 (2004) 1943–1973.

[14] T. Kubokawa, W.E. Strawderman, On minimaxity and admissibility of hierarchical Bayes estimators, J. Multivariate Anal. 98 (2007) 829–851.

[15] D. Pfeffermann, New important developments in small area estimation, Statist. Science 28 (2013) 40–68.

[16] N.N. Prasad, J.N.K. Rao, The estimation of the mean squared error of small-area estimators, J. Amer. Statist. Assoc. 85 (1990) 163–171.

[17] J.N.K. Rao, I. Molina, Small Area Estimation, 2nd ed. (2015) Wiley.

[18] N. Shinozaki, Y.T. Chang, Minimaxity of empirical Bayes estimators of the means of independent normal variables with unequal variances, Com. Statist. Theory Methods 22 (1993) 2147–2169.

[19] 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 University of California University, Berkeley.

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

[21] T. Yanagimoto, M. Yanagimoto, The use of marginal likelihood for a diagnostic test for the goodness of fit of the simple linear regression model, Technometrics 29 (1987) 95–101.