

Dominance properties of constrained Bayes and empirical Bayes estimators

TATSUYA KUBOKAWA¹ and WILLIAM E. STRAWDERMAN²

¹*Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.*

E-mail: tatsuya@e.u-tokyo.ac.jp

²*Department of Statistics, Rutgers University, Piscataway, NJ 08854-8019, USA.*

E-mail: straw@stat.rutgers.edu

This paper studies decision theoretic properties of benchmarked estimators which are of some importance in small area estimation problems. Benchmarking is intended to improve certain aggregate properties (such as study-wide averages) when model based estimates have been applied to individual small areas. We study decision-theoretic properties of such estimators by reducing the problem to one of studying these problems in a related derived problem. For certain such problems, we show that unconstrained solutions in the original (unbenchmarked) problem give unconstrained Bayes and improved estimators which automatically satisfy the benchmark constraint. Also, dominance properties of constrained empirical Bayes estimators are shown in the Fay–Herriot model, a frequently used model in small area estimation.

Keywords: admissibility; benchmark; constrained Bayes estimator; decision theory; dominance result; empirical Bayes; Fay–Herriot model; minimaxity; multivariate normal distribution; quadratic loss function; risk function; small area estimation

1. Introduction

This paper studies decision theoretic properties of benchmarked estimators which are of some importance in small area estimation problems. Benchmarking is intended to improve certain aggregate properties (such as study-wide averages) when empirical Bayes estimates have been applied to individual small areas. For example, model based small area estimates are often such that the average of a particular estimate over all areas may differ substantially from the average derived from a direct estimate. The reader is referred to the articles of Datta *et al.* [6] for an extended discussion of the background and desirability of benchmarking. Also see Frey and Cressie [10], Ghosh [11] and Pfeiffermann and Tiller [16] for related issues. For good accounts of small area estimation, see Battese, Harter and Fuller [1], Prasad and Rao [17], Ghosh and Rao [12], Rao [18] and Datta, Rao and Smith [8].

A useful method for benchmarking is the constrained Bayes and empirical Bayes estimator suggested by Ghosh [11]. Since the constrained Bayes estimator is not a real Bayesian procedure, its decision-theoretic properties like admissibility and minimaxity are interesting questions, though little has been known about such properties. Another query is whether there exists a prior distribution which results in the (unconstrained) real Bayes estimator satisfying the constraint. This paper will address these problems in a decision-theoretic framework.

In Section 2, we begin by explaining the empirical Bayes estimators of small-area means and their benchmarking in the Fay–Herriot area-level model, and give a motivation as well as

the setup of the problem. To investigate basic decision-theoretic properties of the constrained estimator, we decompose the risk function into two pieces; one depends on the risk of the unconstrained estimator in a related problem and one depends on the given means and the benchmark constraint but not the estimator in question. Admissibility considerations and sometimes minimaxity are then reduced to the study of these properties in a related problem. Section 2.3 studies prior distributions in the original problem that result in Bayes estimators which automatically satisfy the benchmark constraint. In fact, we clarify a condition on such prior distributions and gives examples. Such prior distributions and the resulting Bayes estimators enable us to study admissibility. The results in Section 2 are given without assuming normality of the underlying distribution.

Section 3 assumes the multivariate normal distribution, and provides more detailed properties for minimaxity, admissibility and inadmissibility of constrained Bayes estimators. In Section 3.2, we present a prior distribution such that the resulting (generalized) Bayes estimator satisfies the constraint and is also minimax. Admissibility and minimaxity of such unconstrained Bayes estimators are discussed based on some preliminary results given in the literature.

As indicated above, benchmarking is useful in the framework of small area estimation. The Fay–Herriot model is one that is often utilized in small area estimation problems. In Section 4, we consider this model and investigate conditions under which a constrained empirical Bayes estimator improves on the constrained uniform-prior generalized Bayes estimator, namely the constrained direct estimator. Since the Fay–Herriot model has heteroscedastic variances and employs covariates as regressors, establishing minimaxity of the constrained empirical Bayes estimator, while somewhat challenging, seems to be potentially useful. We also consider a prior distribution which results in an unconstrained empirical Bayes estimator satisfying the constraint and minimaxity. These constrained and unconstrained empirical Bayes estimators are investigated in terms of their risk performances by simulation as well as in terms of the conditions for their improvement or minimaxity. Finally, some concluding remarks are given in Section 5.

2. The constrained problem and the dominance property

2.1. The area-level model and the setup of the problem

The Fay–Herriot model has been used as an area-level model in small-area estimation. Let y_1, \dots, y_k be the direct estimators of the k small-area means μ_1, \dots, μ_k . The direct estimator may be taken to be a crude estimator like a sample mean over the small area. This is modeled as

$$y_i = \mu_i + \varepsilon_i, \quad i = 1, \dots, k, \quad (2.1)$$

where $\varepsilon_1, \dots, \varepsilon_k$ are independently distributed as $\varepsilon_i \sim \mathcal{N}(0, d_i)$. While the values of y_i 's are reported from government agencies, the values of the variances d_i 's, are usually not available, and we need to get the values by estimation from past data or other methods. In the framework of small area estimation, the d_i 's are treated as known constants. Small area refers to a small geographical area or a group for which little information is obtained from the sample survey, and the direct estimator based only on the data from a given small area is likely to be unreliable

because only a few observations are available from the small area. Also, y_i is more unreliable for larger d_i . To increase the precision of the estimate, relevant supplementary information such as data from other related small areas or data on covariates is used through Bayesian models. Fay and Herriot [9] suggested a Bayesian model for μ_i in (2.1) with prior distribution of μ_i given by

$$\mu_i \sim \mathcal{N}(\mathbf{x}'_i \boldsymbol{\beta}, \lambda), \quad i = 1, \dots, k, \tag{2.2}$$

where \mathbf{x}_i is a p -variate known vector including covariates, $\boldsymbol{\beta}$ is a p -variate unknown vector and λ is an unknown variance. The resulting empirical Bayes estimator of μ_i is

$$\widehat{\mu}_i^{\text{EB}} = \mathbf{x}'_i \widehat{\boldsymbol{\beta}} + \frac{\widehat{\lambda}}{\widehat{\lambda} + d_i} (y_i - \mathbf{x}'_i \widehat{\boldsymbol{\beta}}),$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\lambda}$ are suitable estimators of $\boldsymbol{\beta}$ and λ . For larger d_i , $\widehat{\mu}_i^{\text{EB}}$ can shrink y_i more toward the estimator $\mathbf{x}'_i \widehat{\boldsymbol{\beta}}$, so that it is expected that $\widehat{\mu}_i^{\text{EB}}$ has a higher precision than y_i . To measure the uncertainty of $\widehat{\mu}_i^{\text{EB}}$, Prasad and Rao [17], Datta and Lahiri [7] and Datta, Rao and Smith [8] derived a second-order approximation of the mean squared error (MSE) of $\widehat{\mu}_i^{\text{EB}}$ for large k under the unconditional model of (2.1) and (2.2), namely, $y_i \sim \mathcal{N}(\mathbf{x}'_i \boldsymbol{\beta}, \lambda + d_i)$. Since the second-order approximation of the MSE is smaller in a large parameter space than that of the direct estimator, $\widehat{\mu}_i^{\text{EB}}$ has been used practically. However, it is not guaranteed analytically that $\widehat{\mu}_i^{\text{EB}}$ has a uniformly smaller MSE than the direct estimator in terms of minimizing the second-order approximation, much less the exact MSE. This point will be demonstrated in Section 4.1.

We can consider the uncertainty of an estimator $\widehat{\mu}_i$ through the two kinds of MSE: the conditional MSE $E[(\widehat{\mu}_i - \mu_i)^2 | \mu_i]$ given μ_i and the unconditional MSE $E[(\widehat{\mu}_i - \mu_i)^2]$. The unconditional MSE is measured based on the unconditional (marginal) distribution of (2.1) and (2.2), and it is interpreted as a Bayesian measure from a Bayesian perspective. The conditional MSE is a measure supported by a frequentist, and it is stronger since it does not assume a distribution for μ_i . In the framework of the conditional MSE, it is known that y_i is admissible in the estimation of the individual mean μ_i , namely, $\widehat{\mu}_i^{\text{EB}}$ does not improve on y_i uniformly in terms of the conditional MSE. However, in simultaneous estimation of the small area means $\mu_i, i = 1, \dots, k$, $\widehat{\mu}_i^{\text{EB}}$ improves on the direct estimator y_i for $k \geq 3$ due to the Stein effect (Stein [22]). Thus, the framework of simultaneous estimation can justify the improvement of $\widehat{\mu}_i^{\text{EB}}$ theoretically.

In this paper, we consider simultaneous estimation of the small-area means. It is convenient to handle the problem in matricial form. Let $\mathbf{y} = (y_1, \dots, y_k)'$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)'$ and $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_k)'$. Then, the model (2.1) is written as

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}, \tag{2.3}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{D})$ for $\mathbf{D} = \text{diag}(d_1, \dots, d_k)$, the $k \times k$ diagonal matrix. When we estimate $\boldsymbol{\mu}$ by $\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{\mu}}(\mathbf{y}) = (\widehat{\mu}_1, \dots, \widehat{\mu}_k)'$ based on \mathbf{y} , the estimator is evaluated in terms of the conditional risk function given $\boldsymbol{\mu}$,

$$R(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) = E[L(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}; \mathbf{Q}) | \boldsymbol{\mu}],$$

relative to weighted squared error loss

$$L(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}; \mathbf{Q}) = (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \mathbf{Q} (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}), \tag{2.4}$$

where \mathbf{Q} is a positive definite matrix. In a decision-theoretic framework, the set of direct estimators \mathbf{y} is minimax, but inadmissible by the so-called Stein effect for $k \geq 3$, namely, there exist shrinkage or empirical Bayes estimators which have uniformly smaller risks than \mathbf{y} for large k . Since a goal in small area estimation is the derivation of estimators having high precisions, desirable estimators should satisfy at least the requirement that they have uniformly smaller risks than \mathbf{y} . This corresponds to the derivation of estimators which are minimax or improve on \mathbf{y} in terms of $R(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})$. It is noted that if an estimator improves on \mathbf{y} in terms of the conditional risk $R(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})$, then it improves on \mathbf{y} relative to the unconditional risk

$$R^U(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) = E^\pi [E[L(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}; \mathbf{Q})|\boldsymbol{\mu}]],$$

where $\pi(\boldsymbol{\mu})$ is a distribution of $\boldsymbol{\mu}$. The unconditional risk is treated for the Fay–Herriot model in Section 4.

As indicated in Section 1, a drawback of the empirical Bayes estimator $\hat{\mu}_i^{\text{EB}}$'s is that the weighted sum $\sum_{i=1}^k w_i \hat{\mu}_i^{\text{EB}}$ is not equal to $\sum_{i=1}^k w_i y_i$, which, for example, corresponds to the total sample mean over the whole area, where w_i 's are nonnegative constants. In the literature, several methods have been proposed in order to benchmark an estimator $\hat{\boldsymbol{\mu}}$ so as to satisfy the constraint $\sum_{i=1}^k w_i \hat{\mu}_i = \sum_{i=1}^k w_i y_i$. Of these, Ghosh [11] suggested the constrained Bayes estimator to satisfy the constraint. In this paper, we consider the general constraint given by

$$\mathbf{W}'\hat{\boldsymbol{\mu}} = \mathbf{t}(\mathbf{y}), \tag{2.5}$$

where \mathbf{W} is a $k \times m$ matrix with rank m , $m < k$, and $\mathbf{t} = \mathbf{t}(\mathbf{y})$ is a function from \mathbf{R}^k to \mathbf{R}^m . Typical examples of $\mathbf{t}(\mathbf{y})$ are $t(\mathbf{y}) = \sum_{i=1}^k w_i y_i$ and $t(\mathbf{y}) = t_0$, a constant. Denote the class of benchmarked estimators by

$$\Gamma_B = \{\hat{\boldsymbol{\mu}} \in \Gamma | \mathbf{W}'\hat{\boldsymbol{\mu}} = \mathbf{t}(\mathbf{y})\},$$

where Γ is the class of estimators with second moments given by $\Gamma = \{\hat{\boldsymbol{\mu}} | E[\hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\mu}}|\boldsymbol{\mu}] < \infty\}$. When a prior distribution π is assumed for $\boldsymbol{\mu}$, the constrained Bayes estimator is defined as the estimator $\hat{\boldsymbol{\mu}}$ which minimizes the posterior risk function $E^\pi[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\mathbf{Q}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})|\mathbf{y}]$ subject to $\hat{\boldsymbol{\mu}} \in \Gamma_B$, where $E^\pi[\cdot|\mathbf{y}]$ denotes a posterior expectation given \mathbf{y} . Noting that

$$E^\pi[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\mathbf{Q}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})|\mathbf{y}] = E^\pi[(\hat{\boldsymbol{\mu}}^B - \boldsymbol{\mu})'\mathbf{Q}(\hat{\boldsymbol{\mu}}^B - \boldsymbol{\mu})|\mathbf{y}] + (\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^B)'\mathbf{Q}(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}^B)$$

for the Bayes estimator $\hat{\boldsymbol{\mu}}^B = E^\pi[\boldsymbol{\mu}|\mathbf{y}]$, Datta *et al.* [6] showed that the constrained Bayes estimator is given by

$$\hat{\boldsymbol{\mu}}^{\text{CB}} = \hat{\boldsymbol{\mu}}^B + \mathbf{Q}^{-1}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\{\mathbf{t}(\mathbf{y}) - \mathbf{W}'\hat{\boldsymbol{\mu}}^B\},$$

Motivated by the constrained Bayes estimator, we can construct the following constrained estimator based on any given estimator $\hat{\boldsymbol{\mu}}$:

$$\hat{\boldsymbol{\mu}}^C(\hat{\boldsymbol{\mu}}, \mathbf{t}) = \hat{\boldsymbol{\mu}} + \mathbf{Q}^{-1}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\{\mathbf{t}(\mathbf{y}) - \mathbf{W}'\hat{\boldsymbol{\mu}}\}, \tag{2.6}$$

and denote the class by

$$\Gamma_C = \{\hat{\boldsymbol{\mu}}^C(\hat{\boldsymbol{\mu}}, \mathbf{t}) | \hat{\boldsymbol{\mu}} \in \Gamma\}.$$

It is seen that

$$\Gamma_C \subset \Gamma_B \subset \Gamma.$$

Since \mathbf{y} is the generalized Bayes estimator of $\boldsymbol{\mu}$ against the uniform prior, the constrained generalized Bayes estimator against the uniform prior is expressed as

$$\widehat{\boldsymbol{\mu}}^{CM}(\mathbf{t}) = \mathbf{y} + \mathbf{Q}^{-1}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\{\mathbf{t}(\mathbf{y}) - \mathbf{W}'\mathbf{y}\}. \tag{2.7}$$

It is noted that the direct estimator \mathbf{y} satisfies the constraint when the constraint is that $\mathbf{t}(\mathbf{y}) = \mathbf{W}'\mathbf{y}$.

Since the constrained Bayes estimator is not necessarily the Bayes estimator among all estimators in Γ , we have several interesting questions from a decision-theoretic perspective. For example, are the properties of minimaxity and inadmissibility of \mathbf{y} inherited by the constrained estimator $\widehat{\boldsymbol{\mu}}^{CM}$? Can one construct an empirical Bayes estimator improving on \mathbf{y} or $\widehat{\boldsymbol{\mu}}^{CM}$? Such issues have not been studied in the literature to our knowledge. The aim of this paper is to investigate such decision-theoretic properties for the constrained estimators.

2.2. Basic properties of a constrained estimator

In this subsection, we investigate basic properties of minimaxity and admissibility of the constrained estimator under the constraint (2.5) in the model (2.3), where normality of $\boldsymbol{\epsilon}$ is not assumed in this and the next subsections. We begin by decomposing the risk function, which will be useful for investigating the basic properties. Let

$$\mathbf{P}_W = \mathbf{Q}^{-1}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\mathbf{W}'.$$

Then the constrained estimator (2.6) is expressed as

$$\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \mathbf{t}) = (\mathbf{I} - \mathbf{P}_W)\widehat{\boldsymbol{\mu}} + \mathbf{Q}^{-1}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\mathbf{t}(\mathbf{y}). \tag{2.8}$$

To evaluate the risk of $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \mathbf{t})$, note that $\mathbf{W}'(\mathbf{I} - \mathbf{P}_W) = \mathbf{0}$,

$$\begin{aligned} \widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \mathbf{t}) - \boldsymbol{\mu} &= (\mathbf{I} - \mathbf{P}_W)(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) + \mathbf{Q}^{-1}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\{\mathbf{t}(\mathbf{y}) - \mathbf{W}'\boldsymbol{\mu}\}, \\ (\mathbf{I} - \mathbf{P}_W)'\mathbf{Q}(\mathbf{I} - \mathbf{P}_W) &= \mathbf{Q} - \mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\mathbf{W}' = \mathbf{Q}(\mathbf{I} - \mathbf{P}_W). \end{aligned}$$

Then the conditional risk function of $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \mathbf{t})$ relative to the loss (2.4) can be decomposed into two parts as given in the following lemma.

Lemma 2.1. *Assume that $\widehat{\boldsymbol{\mu}} \in \Gamma$. It follows that the conditional risk function of $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \mathbf{t})$ relative to the loss $L(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}; \mathbf{Q})$ is expressed as*

$$R(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \mathbf{t})) = R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) + R_2(\boldsymbol{\mu}, \mathbf{t}), \tag{2.9}$$

where $R_1(\mu, \hat{\mu}) = E[(\hat{\mu} - \mu)'Q(\mathbf{I} - \mathbf{P}_W)(\hat{\mu} - \mu)|\mu]$ and

$$R_2(\mu, \mathbf{t}) = E[(\mathbf{t}(\mathbf{y}) - \mathbf{W}'\mu)'(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}(\mathbf{t}(\mathbf{y}) - \mathbf{W}'\mu)|\mu].$$

Since $\mathbf{t}(\mathbf{y})$ is a given function and $R_2(\mu, \mathbf{t})$ does not depend on the estimator $\hat{\mu}$, the problem of finding improved estimators (in the original benchmark problem) can be reduced to that of finding superior estimators $\hat{\mu}$ in terms of the risk function $R_1(\mu, \hat{\mu})$ relative to the loss function $L(\mu, \hat{\mu}; \mathbf{Q}(\mathbf{I} - \mathbf{P}_W))$.

Proposition 2.1. For two estimators $\hat{\mu}_1$ and $\hat{\mu}_2$ in Γ , and the corresponding constrained estimators $\hat{\mu}^C(\hat{\mu}_1, \mathbf{t})$ and $\hat{\mu}^C(\hat{\mu}_2, \mathbf{t})$ in Γ_C , $\hat{\mu}^C(\hat{\mu}_1, \mathbf{t})$ dominates $\hat{\mu}^C(\hat{\mu}_2, \mathbf{t})$ relative to the loss $L(\mu, \hat{\mu}; \mathbf{Q})$ if and only if $\hat{\mu}_1$ dominates $\hat{\mu}_2$ relative to the loss $L(\mu, \hat{\mu}; \mathbf{Q}(\mathbf{I} - \mathbf{P}_W))$.

This proposition implies the following proposition concerning admissibility.

Proposition 2.2. Assume that $\hat{\mu} \in \Gamma$. Then the constrained estimator $\hat{\mu}^C(\hat{\mu}, \mathbf{t})$ is admissible in Γ_C in terms of the risk $R(\mu, \hat{\mu}^C)$ if and only if $\hat{\mu}$ is admissible in Γ in terms of the risk $R_1(\mu, \hat{\mu})$.

The above propositions show that dominance properties and admissibility of a constrained estimator $\hat{\mu}^C(\hat{\mu}, \mathbf{t})$ can be reduced to those of the estimator $\hat{\mu}$ in terms of the risk $R_1(\mu, \hat{\mu})$.

Concerning minimaxity, on the other hand, it is seen that the estimator $\hat{\mu}^C(\hat{\mu}^*, \mathbf{t})$ is minimax within the class Γ_C if and only if $\inf_{\hat{\mu} \in \Gamma_C} \sup_{\mu} R(\mu, \hat{\mu}) = \sup_{\mu} R(\mu, \hat{\mu}^*)$, or

$$\inf_{\hat{\mu} \in \Gamma} \sup_{\mu} \{R_1(\mu, \hat{\mu}) + R_2(\mu, \mathbf{t})\} = \sup_{\mu} \{R_1(\mu, \hat{\mu}^*) + R_2(\mu, \mathbf{t})\}.$$

This condition is satisfied if there exists a sequence of prior distributions $\{\pi_n(\mu)\}_{n=1,2,\dots}$ such that

$$\sup_{\mu} \{R_1(\mu, \hat{\mu}^*) + R_2(\mu, \mathbf{t})\} \leq \limsup_{n \rightarrow \infty} \int \{R_1(\mu, \hat{\mu}_n) + R_2(\mu, \mathbf{t})\} \pi_n(\mu) d\mu, \tag{2.10}$$

where $\hat{\mu}_n$ is the corresponding Bayes estimator relative to the risk $\int R(\mu, \hat{\mu})\pi_n(\mu) d\mu$. This condition follows from Theorem 6.5.2 in Zacks [23].

Proposition 2.3. If $\hat{\mu}^*$ satisfies the condition (2.10), then the estimator $\hat{\mu}^C(\hat{\mu}^*, \mathbf{t})$ is minimax within Γ_C .

In particular, under the following condition, the minimaxity problem for the conditional risk $R(\mu, \hat{\mu}^C)$ reduces to that of the risk $R_1(\mu, \hat{\mu})$.

(A1) Assume that $R_2(\mu, \hat{\mu})$ does not depend on the unknown μ .

Proposition 2.4. Assume the condition (A1). Then the constrained estimator $\hat{\mu}^C(\hat{\mu}^*, \mathbf{t})$ is minimax within Γ_C if and only if $\hat{\mu}^*$ is minimax in terms of the risk $R_1(\mu, \hat{\mu})$ in Γ .

Condition (A1) is satisfied for two typical examples of $\mathbf{t}(\mathbf{y})$:

Case 1: $\mathbf{t}(\mathbf{y}) = \mathbf{W}'\mathbf{y}$. In this case, it typically happens that $R_2(\boldsymbol{\mu}, \mathbf{t})$ is independent of $\boldsymbol{\mu}$ under the distributional assumption of a location family, and the condition (A1) holds.

Case 2: $\mathbf{t}(\mathbf{y}) = \mathbf{t}_0$, a constant. In this case, we need to restrict the space of $\boldsymbol{\mu}$ to $\{\boldsymbol{\mu} | \mathbf{W}'\boldsymbol{\mu} = \mathbf{t}_0\}$. Then it is clear that $R_2(\boldsymbol{\mu}, \mathbf{t}_0) = 0$ on the restricted space.

2.3. Unconstrained Bayes estimators satisfying the constraint

In the previous subsections, we studied shrinkage estimators induced from the constrained Bayes estimator and investigated their decision-theoretic properties within the class of constrained estimators. In some cases, however, we can derive constrained Bayes estimators without direct consideration of the constraint. In this subsection, we find a condition on prior distributions such that the resulting unconstrained generalized Bayes estimators satisfy the constraint automatically, where the normality of $\boldsymbol{\epsilon}$ is not assumed.

Assume a prior distribution π for $\boldsymbol{\mu}$. According to the expression in (2.8), we decompose $\boldsymbol{\mu}$ as $\boldsymbol{\mu} = (\mathbf{I} - \mathbf{P}_W)\boldsymbol{\mu} + \mathbf{P}_W\boldsymbol{\mu}$, which implies that the Bayes estimator of $\hat{\boldsymbol{\mu}}^B$ can be expressed as

$$\hat{\boldsymbol{\mu}}^B = (\mathbf{I} - \mathbf{P}_W)E^\pi[\boldsymbol{\mu} | \mathbf{y}] + \mathbf{P}_WE^\pi[\boldsymbol{\mu} | \mathbf{y}].$$

Comparing this expression and the constrained estimator (2.8), we can see that the unconstrained Bayes estimator $\hat{\boldsymbol{\mu}}^B$ belongs to the class Γ_C if the prior distribution satisfies the equation

$$\mathbf{P}_WE^\pi[\boldsymbol{\mu} | \mathbf{y}] = \mathbf{Q}^{-1}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\mathbf{t}(\mathbf{y}).$$

It follows from the definition of \mathbf{P}_W that this equality is simplified as

$$\mathbf{W}'E^\pi[\boldsymbol{\mu} | \mathbf{y}] = \mathbf{t}(\mathbf{y}). \tag{2.11}$$

Since the condition (2.11) means that the posterior expectation of $\mathbf{W}'\boldsymbol{\mu}$ is $\mathbf{t}(\mathbf{y})$, the following transformation is convenient for investigating prior distributions satisfying (2.11). Let \mathbf{H} be a $k \times k$ orthogonal matrix such that

$$\mathbf{H}\mathbf{Q}^{-1/2}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\mathbf{W}'\mathbf{Q}^{-1/2}\mathbf{H}' = \begin{pmatrix} \mathbf{0}_{k-m} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{pmatrix}. \tag{2.12}$$

Let $\mathbf{H}' = (\mathbf{H}'_1, \mathbf{H}'_2)$ for the $k \times (k - m)$ matrix \mathbf{H}_1 . Also, let $\boldsymbol{\xi} = \mathbf{H}\mathbf{Q}^{1/2}\boldsymbol{\mu}$ and $\boldsymbol{\xi}_i = \mathbf{H}_i\mathbf{Q}^{1/2}\boldsymbol{\mu}$ for $i = 1, 2$. Then, $\mathbf{P}_W = \mathbf{Q}^{-1/2}\mathbf{H}'_2\mathbf{H}_2\mathbf{Q}^{1/2}$ and $\mathbf{I} - \mathbf{P}_W = \mathbf{Q}^{-1/2}\mathbf{H}'_1\mathbf{H}_1\mathbf{Q}^{1/2}$. It is noted that $\mathbf{H}_1\mathbf{Q}^{-1/2}\mathbf{W} = \mathbf{0}$, since $\mathbf{H}_1\mathbf{Q}^{-1/2}\mathbf{W} = \mathbf{H}_1\mathbf{H}'_1\mathbf{H}_1\mathbf{Q}^{-1/2}\mathbf{W} = \mathbf{Q}^{1/2}(\mathbf{I} - \mathbf{P}_W)\mathbf{Q}^{-1}\mathbf{W} = \mathbf{0}$. Thus, $\boldsymbol{\mu}$ and the constrained estimator (2.8) are written as

$$\begin{aligned} \boldsymbol{\mu} &= \mathbf{Q}^{-1/2}\mathbf{H}\mathbf{H}'\mathbf{Q}^{1/2}\boldsymbol{\mu} = \mathbf{Q}^{-1/2}\mathbf{H}'_1\boldsymbol{\xi}_1 + \mathbf{Q}^{-1/2}\mathbf{H}'_2\boldsymbol{\xi}_2, \\ \hat{\boldsymbol{\mu}}^C(\hat{\boldsymbol{\mu}}, \mathbf{t}) &= \mathbf{Q}^{-1/2}\mathbf{H}'_1\mathbf{H}_1\mathbf{Q}^{1/2}\hat{\boldsymbol{\mu}} + \mathbf{Q}^{-1}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\mathbf{t}(\mathbf{y}), \end{aligned} \tag{2.13}$$

which shows that the unconstrained Bayes estimator belongs to Γ_C if

$$\mathbf{H}'_2 E^\pi [\xi_2 | \mathbf{y}] = \mathbf{Q}^{-1/2} \mathbf{W} (\mathbf{W}' \mathbf{Q}^{-1} \mathbf{W})^{-1} \mathbf{t}(\mathbf{y}). \quad (2.14)$$

Noting that $\mathbf{Q}^{-1/2} \mathbf{W} = (\mathbf{H}'_1 \mathbf{H}_1 + \mathbf{H}'_2 \mathbf{H}_2) \mathbf{Q}^{-1/2} \mathbf{W} = \mathbf{H}'_2 \mathbf{H}_2 \mathbf{Q}^{-1/2} \mathbf{W}$, we see that the equation (2.14) holds if

$$E^\pi [\xi_2 | \mathbf{y}] = \mathbf{H}_2 \mathbf{Q}^{-1/2} \mathbf{W} (\mathbf{W}' \mathbf{Q}^{-1} \mathbf{W})^{-1} \mathbf{t}(\mathbf{y}).$$

For example, consider the case that the constraint is given by $\mathbf{t}(\mathbf{y}) = \mathbf{W}' \mathbf{Q}^{-1/2} \mathbf{s}(\mathbf{y})$ for a k -variate vector $\mathbf{s}(\mathbf{y})$ of functions of \mathbf{y} . In this case, we have that

$$\mathbf{H}_2 \mathbf{Q}^{-1/2} \mathbf{W} (\mathbf{W}' \mathbf{Q}^{-1} \mathbf{W})^{-1} \mathbf{W}' \mathbf{Q}^{-1/2} \mathbf{H}' \mathbf{H} \mathbf{s}(\mathbf{y}) = \mathbf{H}_2 \mathbf{s}(\mathbf{y}),$$

so that the condition (2.14) may be simplified as $E^\pi [\xi_2 | \mathbf{y}] = \mathbf{H}_2 \mathbf{s}(\mathbf{y})$. Thus, we summarize the condition in the following.

(A2) Assume that $\mathbf{W}' E^\pi [\boldsymbol{\mu} | \mathbf{y}] = \mathbf{t}(\mathbf{y})$, or that $E^\pi [\xi_2 | \mathbf{y}] = \mathbf{H}_2 \mathbf{Q}^{-1/2} \mathbf{W} (\mathbf{W}' \mathbf{Q}^{-1} \mathbf{W})^{-1} \mathbf{t}(\mathbf{y})$. The latter condition is simplified as $E^\pi [\xi_2 | \mathbf{y}] = \mathbf{H}_2 \mathbf{s}(\mathbf{y})$ when $\mathbf{t}(\mathbf{y}) = \mathbf{W}' \mathbf{Q}^{-1/2} \mathbf{s}(\mathbf{y})$.

Proposition 2.5. *The unconstrained Bayes estimators belong to the class Γ_C , namely they automatically satisfy the constraint $\mathbf{W}' \hat{\boldsymbol{\mu}} = \mathbf{t}(\mathbf{y})$ if the posterior expectation $E^\pi [\boldsymbol{\mu} | \mathbf{y}]$ or $E^\pi [\xi_2 | \mathbf{y}]$ satisfies the condition (A2).*

Case 1: $\mathbf{t}(\mathbf{y}) = \mathbf{W}' \mathbf{y}$. In this case, the condition (A2) is $\mathbf{W}' E^\pi [\boldsymbol{\mu} | \mathbf{y}] = \mathbf{W}' \mathbf{y}$ or $E^\pi [\xi_2 | \mathbf{y}] = \mathbf{H}_2 \mathbf{Q}^{1/2} \mathbf{y}$ for $\xi_2 = \mathbf{H}_2 \mathbf{Q}^{1/2} \boldsymbol{\mu}$. As explained in the next section, it suffices that we assume the uniform prior for ξ_2 under normality of $\boldsymbol{\epsilon}$.

Case 2: $\mathbf{t}(\mathbf{y}) = \mathbf{t}_0$, a constant. In this case, the condition (A2) is $E^\pi [\xi_2 | \mathbf{y}] = \mathbf{H}_2 \mathbf{Q}^{-1/2} \mathbf{W} \times (\mathbf{W}' \mathbf{Q}^{-1} \mathbf{W})^{-1} \mathbf{t}_0$, which suggests that ξ_2 should take a point mass at $\xi_2 = \mathbf{H}_2 \mathbf{Q}^{-1/2} \mathbf{W} \times (\mathbf{W}' \mathbf{Q}^{-1} \mathbf{W})^{-1} \mathbf{t}_0$. Since $\mathbf{W}' \mathbf{Q}^{-1/2} \mathbf{H}'_1 = \mathbf{0}$, it is verified that this restriction satisfies $\mathbf{W}' \boldsymbol{\mu} = \mathbf{t}_0$.

3. Properties under normality and conditional risk

In this section, we further investigate minimaxity and admissibility properties for the benchmark problem in the model (2.3), where normality of $\boldsymbol{\epsilon}$ is assumed.

3.1. Constrained Bayes estimator

We begin by deriving the canonical form of the model (2.3) with \mathbf{y} having a multivariate normal distribution $\mathcal{N}_k(\boldsymbol{\mu}, \mathbf{D})$. For the matrix \mathbf{H} defined by (2.12), let $\mathbf{z}_i = \mathbf{H}_i \mathbf{Q}^{1/2} \mathbf{y}$ and $\mathbf{V}_{ij} = \mathbf{H}_i \mathbf{Q}^{1/2} \mathbf{D} \mathbf{Q}^{1/2} \mathbf{H}'_j$ for $i, j = 1, 2$. Then, $\mathbf{z} = (\mathbf{z}'_1, \mathbf{z}'_2)'$ is distributed as

$$\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} \sim \mathcal{N}_k \left(\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \right), \quad (3.1)$$

where $\xi_i = \mathbf{H}_i \mathbf{Q}^{1/2} \boldsymbol{\mu}$, $i = 1, 2$, for \mathbf{H}_i defined in (2.12). The problem of finding a constrained Bayes estimator may be expressed as the minimization of $E^{\pi}[(\widehat{\xi}_1 - \xi_1)'(\widehat{\xi}_1 - \xi_1)|\mathbf{z}]$ subject to $\mathbf{W}'\mathbf{Q}^{-1/2}\mathbf{H}'_2\widehat{\xi}_2 = \mathbf{t}(\mathbf{y})$, since $\mathbf{W}'\mathbf{Q}^{-1/2}\mathbf{H}'_1 = \mathbf{0}$. The constrained estimators given in (2.6) and (2.7) are rewritten as

$$\begin{aligned} \widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \mathbf{t}) &= \mathbf{Q}^{-1/2}\mathbf{H}'_1\widehat{\xi}_1 + \mathbf{Q}^{-1}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\mathbf{t}(\mathbf{y}) \equiv \widehat{\boldsymbol{\mu}}^{C*}(\widehat{\xi}_1, \mathbf{t}), \\ \widehat{\boldsymbol{\mu}}^{\text{CM}}(\mathbf{t}) &= \mathbf{Q}^{-1/2}\mathbf{H}'_1\mathbf{z}_1 + \mathbf{Q}^{-1}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\mathbf{t}(\mathbf{y}) \equiv \widehat{\boldsymbol{\mu}}^{C*}(\mathbf{z}_1, \mathbf{t}). \end{aligned}$$

For $\widehat{\xi}_1 = \mathbf{H}_1\mathbf{Q}^{1/2}\widehat{\boldsymbol{\mu}}$, and the conditional risk $R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}})$ given in (2.9) is written as

$$R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) = E[\|\widehat{\xi}_1 - \xi_1\|^2|\xi_1] = R^*(\xi_1, \widehat{\xi}_1),$$

where $\|\widehat{\xi}_1 - \xi_1\|^2 = (\widehat{\xi}_1 - \xi_1)'(\widehat{\xi}_1 - \xi_1)$. Hence from Proposition 2.2, we get the following proposition.

Proposition 3.1. *If $\widehat{\xi}_1$ is admissible in terms of $R^*(\xi_1, \widehat{\xi}_1)$, then $\widehat{\boldsymbol{\mu}}$ is admissible within the class Γ_C . In particular, if $\widehat{\xi}_1$ is the Bayes estimator for a proper prior on ξ_1 , then $\widehat{\boldsymbol{\mu}}$ is admissible within Γ_C . If $\widehat{\xi}_1$ is inadmissible in terms of the risk $R^*(\xi_1, \widehat{\xi}_1)$, then $\widehat{\boldsymbol{\mu}}$ is inadmissible.*

Also, from Propositions 2.4 and 3.1 and the well-known results of James and Stein [13] and Brown [5], the next proposition follows.

Proposition 3.2. *The constrained generalized Bayes estimator $\widehat{\boldsymbol{\mu}}^{\text{CM}}(\mathbf{t})$ for the uniform prior has the following decision-theoretic properties:*

- (1) $\widehat{\boldsymbol{\mu}}^{\text{CM}}(\mathbf{t})$ is minimax within Γ_C under the condition (A1).
- (2) $\widehat{\boldsymbol{\mu}}^{\text{CM}}(\mathbf{t})$ is admissible within Γ_C when $k - m$ is one or two.
- (3) $\widehat{\boldsymbol{\mu}}^{\text{CM}}(\mathbf{t})$ is inadmissible within Γ_C when $k - m \geq 3$.

Proposition 3.2(3) implies that there exist shrinkage estimators like empirical Bayes estimators which improve on $\widehat{\boldsymbol{\mu}}^{\text{CM}}(\mathbf{t})$ for large k . One of such improved estimators is given in Section 4. Noting that $\mathbf{z}_1 \sim \mathcal{N}_{k-m}(\xi_1, \mathbf{V}_{11})$, from the result in Berger [2], we can get an admissible and minimax estimator, denoted by $\widehat{\xi}_1^{\text{GB}}(\mathbf{z}_1, \mathbf{V}_{11})$, based on $(\mathbf{z}_1, \mathbf{V}_{11})$ relative to the risk $R^*(\xi_1, \widehat{\xi}_1)$. This leads to the constrained generalized Bayes estimator

$$\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}^{\text{GB}}, \mathbf{t}) = \mathbf{Q}^{-1/2}\mathbf{H}'_1\widehat{\xi}_1^{\text{GB}}(\mathbf{z}_1, \mathbf{V}_{11}) + \mathbf{Q}^{-1}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\mathbf{t}(\mathbf{y}),$$

which is admissible within the class Γ_C and improves on $\widehat{\boldsymbol{\mu}}^{\text{CM}}(\mathbf{t})$ when $k - m \geq 3$.

3.2. Unconstrained Bayes estimators

We now construct unconstrained Bayes estimators satisfying the constraint automatically in the two cases $\mathbf{t}(\mathbf{y}) = \mathbf{W}'\mathbf{y}$ and $\mathbf{t}(\mathbf{y}) = \mathbf{t}_0$, a constant. To this end, the following decomposition is

useful:

$$\begin{pmatrix} \mathbf{z}_3 \\ \mathbf{z}_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \xi_3 \\ \xi_2 \end{pmatrix}, \begin{pmatrix} \mathbf{V}_{11.2} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{22} \end{pmatrix} \right), \quad (3.2)$$

where $\mathbf{z}_3 = \mathbf{z}_1 - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{z}_2$, $\xi_3 = \xi_1 - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\xi_2$ and $\mathbf{V}_{11.2} = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$. Note that $\mathbf{z}_1 = \mathbf{z}_3 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{z}_2$ and that \mathbf{z}_3 is independent of $\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{z}_2$.

Case 1: $\mathbf{t}(\mathbf{y}) = \mathbf{W}'\mathbf{y}$. Consider the decomposition (3.2). Note that from (2.13), $\boldsymbol{\mu}$ is written as

$$\boldsymbol{\mu} = \mathbf{Q}^{-1/2}\mathbf{H}'_1\xi_3 + \mathbf{Q}^{-1/2}(\mathbf{H}'_2 + \mathbf{H}'_1\mathbf{V}_{12}\mathbf{V}_{22}^{-1})\xi_2.$$

Assume a prior distribution $\pi(\xi_3)$ for ξ_3 and the uniform prior $\pi(\xi_2) = 1$ for ξ_2 . Then, the resulting generalized Bayes estimator is

$$\begin{aligned} \hat{\boldsymbol{\mu}}^{\text{GB1}} &= \mathbf{Q}^{-1/2}\mathbf{H}'_1\hat{\xi}_3^{\text{GB}}(\mathbf{z}_3, \mathbf{V}_{11.2}) + \mathbf{Q}^{-1/2}(\mathbf{H}'_2 + \mathbf{H}'_1\mathbf{V}_{12}\mathbf{V}_{22}^{-1})\mathbf{z}_2 \\ &= \mathbf{Q}^{-1/2}\mathbf{H}'_1\{\hat{\xi}_3^{\text{GB}}(\mathbf{z}_3, \mathbf{V}_{11.2}) + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{z}_2\} + \mathbf{Q}^{-1/2}\mathbf{H}'_2\mathbf{z}_2, \end{aligned} \quad (3.3)$$

where $\hat{\xi}_3^{\text{GB}} = \hat{\xi}_3^{\text{GB}}(\mathbf{z}_3, \mathbf{V}_{11.2})$ is the generalized Bayes estimator of ξ_3 which can be constructed via the model $\mathbf{z}_3|\xi_3 \sim \mathcal{N}_{k-m}(\xi_3, \mathbf{V}_{11.2})$ and $\xi_3 \sim \pi(\xi_3)$. Note that the conditional risk $R(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}^{\text{GB1}}) = E[(\hat{\boldsymbol{\mu}}^{\text{GB1}} - \boldsymbol{\mu})'\mathbf{Q}(\hat{\boldsymbol{\mu}}^{\text{GB1}} - \boldsymbol{\mu})|\boldsymbol{\mu}]$ is evaluated as

$$\begin{aligned} &E\left[\{\hat{\xi}_3^{\text{GB}} - \xi_3 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{z}_2 - \xi_2)\}'\mathbf{H}_1\mathbf{H}'_1\{\hat{\xi}_3^{\text{GB}} - \xi_3 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{z}_2 - \xi_2)\}\right] \\ &\quad + E\left[(\mathbf{z}_2 - \xi_2)'\mathbf{H}_2\mathbf{H}'_2(\mathbf{z}_2 - \xi_2)\right] \\ &= E\left[\|\hat{\xi}_3^{\text{GB}} - \xi_3\|^2\right] + E\left[(\mathbf{z}_2 - \xi_2)'\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{z}_2 - \xi_2)\right] \\ &\quad + E\left[\|\mathbf{z}_2 - \xi_2\|^2\right] \\ &= R^*(\xi_3, \hat{\xi}_3^{\text{GB}}) + \text{tr}[\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}] + \text{tr}[\mathbf{V}_{22}], \end{aligned} \quad (3.4)$$

where $R^*(\xi_3, \hat{\xi}_3) = E[\|\hat{\xi}_3 - \xi_3\|^2|\xi_3]$. If $R^*(\xi_3, \hat{\xi}_3^{\text{GB}}) \leq R^*(\xi_3, \mathbf{z}_3)$, then $R(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}^{\text{GB1}}) \leq \text{tr}[\mathbf{V}_{11} + \mathbf{V}_{22}] = \text{tr}[\mathbf{DQ}]$, since $R^*(\xi_3, \mathbf{z}_3) = \text{tr}[\mathbf{V}_{11.2}]$. Since $\text{tr}[\mathbf{DQ}]$ is the minimax risk, the unconstrained Bayes estimator $\hat{\boldsymbol{\mu}}^{\text{GB1}}$ is minimax in Γ . Noting that $\mathbf{z}_3 \sim \mathcal{N}_{k-m}(\xi_3, \mathbf{V}_{11.2})$, from the result in Berger [2], we can get an admissible and minimax estimator based on $(\mathbf{z}_3, \mathbf{V}_{11.2})$ relative to the risk $R^*(\xi_3, \hat{\xi}_3)$.

Proposition 3.3. Assume the uniform prior for ξ_2 . Then the generalized Bayes estimator $\hat{\boldsymbol{\mu}}^{\text{GB1}}$ satisfies the constraint, namely $\hat{\boldsymbol{\mu}}^{\text{GB1}} \in \Gamma_C$.

- (1) If $R^*(\xi_3, \hat{\xi}_3^{\text{GB}}) \leq R^*(\xi_3, \mathbf{z}_3)$, then $\hat{\boldsymbol{\mu}}^{\text{GB1}}$ is minimax in Γ .
- (2) If $\hat{\xi}_3^{\text{GB}}$ is admissible in terms of the risk $R^*(\xi_3, \cdot)$, then $\hat{\boldsymbol{\mu}}^{\text{GB1}}$ is admissible within the constrained class Γ_C .
- (3) When $m \geq 3$, $\hat{\boldsymbol{\mu}}^{\text{GB1}}$ is not admissible in the unconstrained problem even if $\hat{\xi}_3^{\text{GB}}$ is admissible in terms of the risk $R^*(\xi_3, \cdot)$.

Case 2: $\mathbf{t}(\mathbf{y}) = \mathbf{t}_0$, a constant. Assume that $\mathbf{W}'\mathbf{Q}^{-1/2}\mathbf{H}'_2$ is non-singular. Since $\mathbf{W}'\boldsymbol{\mu} = \mathbf{W}'\mathbf{Q}^{-1/2}\mathbf{H}'_2\xi_2 = \mathbf{t}_0$, we can define ξ_0 by $\xi_0 = (\mathbf{W}'\mathbf{Q}^{-1/2}\mathbf{H}'_2)^{-1}\mathbf{t}_0$. Let $\mathbf{z}_4 = \mathbf{z}_1 - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{z}_2 - \xi_0)$ and $\xi_4 = \xi_1 - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\xi_2 - \xi_0)$. Then from the decomposition (3.2), the joint distribution of $(\mathbf{z}_4, \mathbf{z}_2)$ follows that

$$\begin{pmatrix} \mathbf{z}_4 \\ \mathbf{z}_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \xi_4 \\ \xi_2 \end{pmatrix}, \begin{pmatrix} \mathbf{V}_{11.2} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{22} \end{pmatrix} \right). \tag{3.5}$$

Assume a prior distribution $\pi(\xi_4)$ for ξ_4 and $P^\pi[\xi_2 = \xi_0] = 1$ for ξ_2 . Then from (3.3), the resulting generalized Bayes estimator is

$$\hat{\boldsymbol{\mu}}^{\text{GB2}} = \mathbf{Q}^{-1/2}\mathbf{H}'_1 \{ \hat{\xi}_4^{\text{GB}}(\mathbf{z}_4, \mathbf{V}_{11.2}) + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\xi_0 \} + \mathbf{Q}^{-1/2}\mathbf{H}'_2\xi_0, \tag{3.6}$$

where $\hat{\xi}_4^{\text{GB}} = \hat{\xi}_4^{\text{GB}}(\mathbf{z}_4, \mathbf{V}_{11.2})$ is the generalized Bayes estimator of ξ_4 which can be constructed via the model $\mathbf{z}_4|\xi_4 \sim \mathcal{N}_{k-m}(\xi_4, \mathbf{V}_{11.2})$ and $\xi_4 \sim \pi(\xi_4)$. It also follows from (3.4) that the risk function of $\hat{\boldsymbol{\mu}}^{\text{GB2}}$ is

$$R(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}^{\text{GB2}}) = R^*(\xi_4, \hat{\xi}_4^{\text{GB}}) + (\xi_0 - \xi_2)' \{ \mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{V}_{12}\mathbf{V}_{22}^{-1} + \mathbf{I}_m \} (\xi_0 - \xi_2),$$

so that the admissibility of $\hat{\boldsymbol{\mu}}^{\text{GB2}}$ is inherited from that of $\hat{\xi}_4^{\text{GB}}$. If the space of $\boldsymbol{\mu}$ is restricted to $\{\boldsymbol{\mu} | \mathbf{W}'\boldsymbol{\mu} = \mathbf{t}_0\}$, and if $R^*(\xi_4, \hat{\xi}_4^{\text{GB}}) \leq R^*(\xi_4, \mathbf{z}_4)$, then $R(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}^{\text{GB2}}) \leq \text{tr}[\mathbf{V}_{11.2}]$. Since $\text{tr}[\mathbf{V}_{11.2}]$ is the minimax risk under the restriction, the unconstrained Bayes estimator $\hat{\boldsymbol{\mu}}^{\text{GB2}}$ is minimax in Γ when $\boldsymbol{\mu}$ is restricted.

Proposition 3.4. *Assume the point mass prior for ξ_2 . Then the generalized Bayes estimator $\hat{\boldsymbol{\mu}}^{\text{GB2}}$ satisfies the constraint, namely $\hat{\boldsymbol{\mu}}^{\text{GB2}} \in \Gamma_C$.*

- (1) *If the estimator $\hat{\xi}_4^{\text{GB}}$ of ξ_4 is admissible in terms of the risk $R^*(\xi_4, \hat{\xi}_4^{\text{GB}})$, then $\hat{\boldsymbol{\mu}}^{\text{GB2}}$ is admissible in Γ (and also Γ_C).*
- (2) *If $R^*(\xi_4, \hat{\xi}_4^{\text{GB}}) \leq R^*(\xi_4, \mathbf{z}_4)$, then $\hat{\boldsymbol{\mu}}^{\text{GB2}}$ is minimax within the class Γ_C . Further, it is minimax in Γ when $\boldsymbol{\mu}$ is restricted to $\mathbf{W}'\boldsymbol{\mu} = \mathbf{t}_0$ or $\xi_2 = \xi_0$.*

4. Benchmarking in the Fay–Herriot model

As mentioned in the introduction and as explained in Datta *et al.* [6] benchmarking is useful in the framework of small area estimation. The Fay–Herriot model is often utilized in such problems. In this section, we develop a constrained empirical Bayes estimator for this model and investigate the dominance properties.

4.1. Constrained empirical Bayes estimator

The Fay–Herriot model given in (2.1) and (2.2) can be described in matricial form as

$$\mathbf{y} | \boldsymbol{\mu} \sim \mathcal{N}_k(\boldsymbol{\mu}, \mathbf{D}), \quad \mathbf{D} = \text{diag}(d_1, \dots, d_k), \quad \boldsymbol{\mu} \sim \mathcal{N}_k(\mathbf{X}\boldsymbol{\beta}, \lambda\mathbf{I}),$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_k)'$ is a $k \times p$ matrix of explanatory variables with rank p , $\boldsymbol{\beta}$ is a $p \times 1$ unknown vector of regression coefficients and λ is an unknown scalar. Suppose that $d_1 \geq \dots \geq d_k$ without any loss of generality. Consider estimation of $\boldsymbol{\mu}$ in terms of the conditional risk $R(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) = E[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \mathbf{Q}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) | \boldsymbol{\mu}]$ and the unconditional risk $R^U(\pi, \hat{\boldsymbol{\mu}}) = E[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \mathbf{Q}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})]$ where π denotes the distribution of $\boldsymbol{\mu}$. The Bayes estimator (under the assumption of known $\boldsymbol{\beta}$ and λ) is given by

$$\hat{\boldsymbol{\mu}}^B = \mathbf{X}\boldsymbol{\beta} + (\mathbf{D}/\lambda + \mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y} - \mathbf{D}(\mathbf{D} + \lambda\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

For estimation of λ , several estimators are known including the Prasad–Rao estimator given by Prasad and Rao [17], the Fay–Herriot estimator suggested by Fay and Herriot [9], the maximum likelihood estimator (MLE) and the restricted maximum likelihood estimator (REML). For the MLE and REML, see Searle, Casella and McCulloch [19] and Kubokawa [14], for example. Denoting an estimator of λ by $\hat{\lambda}$, we get the empirical Bayes estimator $\hat{\boldsymbol{\mu}}^{EB}(\hat{\lambda}) = \mathbf{y} - \mathbf{D}(\mathbf{D} + \hat{\lambda}\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\hat{\lambda}))$, where $\hat{\boldsymbol{\beta}}(\hat{\lambda}) = \{\mathbf{X}'\mathbf{V}(\hat{\lambda})^{-1}\mathbf{X}\}^{-1}\mathbf{X}'\mathbf{V}(\hat{\lambda})^{-1}\mathbf{y}$ for $\mathbf{V}(\lambda) = \mathbf{D} + \lambda\mathbf{I}$. The empirical Bayes estimator is called the empirical best linear unbiased predictor (EBLUP) in the framework of the linear mixed model, namely the unconditional model (4.1). Define $\mathbf{A}(\lambda)$ by

$$\mathbf{A}(\lambda) = \mathbf{V}(\lambda)^{-1} - \mathbf{V}(\lambda)^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}(\lambda)^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}(\lambda)^{-1}.$$

Then, the empirical Bayes estimator can be rewritten as

$$\hat{\boldsymbol{\mu}}^{EB}(\hat{\lambda}) = \mathbf{y} - \mathbf{D}\mathbf{A}(\hat{\lambda})\mathbf{y}. \tag{4.1}$$

Now consider the benchmark constraint $\mathbf{W}'\hat{\boldsymbol{\mu}} = \mathbf{t}(\mathbf{y})$. The constrained empirical Bayes estimator (CEB) based on $\hat{\boldsymbol{\mu}}^{EB}(\hat{\lambda})$ (as constructed in 2.8) is given by

$$\hat{\boldsymbol{\mu}}^{CEB}(\hat{\lambda}, \mathbf{t}) = (\mathbf{I} - \mathbf{P}_W)\hat{\boldsymbol{\mu}}^{EB}(\hat{\lambda}) + \mathbf{Q}^{-1}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\mathbf{t}(\mathbf{y}). \tag{4.2}$$

Concerning the estimation of λ , we here treat the Fay–Herriot estimator $\hat{\lambda}$ given by $\hat{\lambda} = \max\{\lambda^*, 0\}$ where λ^* is the solution of the equation

$$\mathbf{y}'\mathbf{A}(\lambda^*)\mathbf{y} = k - p. \tag{4.3}$$

1. *Conditional risk.* A sufficient condition for $\hat{\boldsymbol{\mu}}^{CEB}(\hat{\lambda}, \mathbf{t})$ to improve on $\hat{\boldsymbol{\mu}}^{CM}(\mathbf{t})$ in terms of the conditional risk is given in the following proposition which will be proved in the [Appendix](#).

Proposition 4.1. *The constrained empirical Bayes estimator $\hat{\boldsymbol{\mu}}^{CEB}(\hat{\lambda}, \mathbf{t})$ with $\hat{\lambda}$ given in (4.3) improves on $\hat{\boldsymbol{\mu}}^{CM}(\mathbf{t})$ given in (2.7) in terms of the conditional risk if the following inequality holds:*

$$\min_{\lambda > 0} \left\{ \frac{\text{tr}[\mathbf{D}\mathbf{Q}_W\mathbf{D}\mathbf{A}(\lambda)]}{\text{Ch}_{\max}(\mathbf{D}\mathbf{Q}_W\mathbf{D}\mathbf{A}(\lambda))} \right\} \geq \frac{k - p}{2} + 2, \tag{4.4}$$

where $\mathbf{Q}_W = \mathbf{Q} - \mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\mathbf{W}'$, and $\text{Ch}_{\max}(\mathbf{C})$ denotes the maximum eigenvalue of the matrix \mathbf{C} . If the constraint is given by $\mathbf{t}(\mathbf{y}) = \mathbf{W}'\mathbf{y}$, then the estimator $\hat{\boldsymbol{\mu}}^{CEB}(\hat{\lambda}, \mathbf{t})$ is minimax under the condition (4.4).

To derive explicit sufficient conditions, it is noted that

$$\begin{aligned} \text{tr}[\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)] &= \text{tr}[\mathbf{DQ}_W\mathbf{D}(\mathbf{D} + \lambda\mathbf{I})^{-1}] \\ &\quad - \text{tr}[\{\mathbf{X}'(\mathbf{D} + \lambda\mathbf{I})^{-1}\mathbf{X}\}^{-1}\mathbf{X}'(\mathbf{D} + \lambda\mathbf{I})^{-1}\mathbf{DQ}_W\mathbf{D}(\mathbf{D} + \lambda\mathbf{I})^{-1}\mathbf{X}] \\ &\geq \frac{1}{d_1 + \lambda} \text{tr}[\mathbf{D}^2\mathbf{Q}_W] - \frac{d_1 p}{d_1 + \lambda} \text{Ch}_{\max}(\mathbf{DQ}_W), \end{aligned}$$

$$\text{Ch}_{\max}(\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)) \leq \text{Ch}_{\max}(\mathbf{DQ}_W\mathbf{D}(\mathbf{D} + \lambda\mathbf{I})^{-1}) \leq \frac{d_1}{d_1 + \hat{\lambda}} \text{Ch}_{\max}(\mathbf{DQ}_W),$$

where $d_1 \geq \dots \geq d_k$. Then,

$$\text{tr}[\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)] / \text{Ch}_{\max}(\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)) \geq \text{tr}[\mathbf{D}^2\mathbf{Q}_W] / \{d_1 \text{Ch}_{\max}(\mathbf{DQ}_W)\} - p.$$

Similarly, it is observed that

$$\begin{aligned} \text{tr}[\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)] &\geq \frac{d_k}{d_k + \lambda} \text{tr}[\mathbf{DQ}_W] - \frac{p}{d_k + \lambda} \text{Ch}_{\max}(\mathbf{D}^2\mathbf{Q}_W), \\ \text{Ch}_{\max}(\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)) &\leq \frac{1}{d_k + \hat{\lambda}} \text{Ch}_{\max}(\mathbf{D}^2\mathbf{Q}_W), \end{aligned}$$

which implied that $\text{tr}[\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)] / \text{Ch}_{\max}(\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)) \geq d_k \text{tr}[\mathbf{DQ}_W] / \text{Ch}_{\max}(\mathbf{D}^2\mathbf{Q}_W) - p$. These provide the following sufficient condition.

Proposition 4.2. *The constrained empirical Bayes estimator $\hat{\boldsymbol{\mu}}^{\text{CEB}}(\hat{\lambda}, \mathbf{t})$ improves on $\hat{\boldsymbol{\mu}}^{\text{CM}}(\mathbf{t})$ in terms of the conditional risk if the following condition holds:*

$$\max \left\{ \frac{\text{tr}[\mathbf{D}^2\mathbf{Q}_W]}{d_1 \text{Ch}_{\max}(\mathbf{DQ}_W)}, \frac{d_k \text{tr}[\mathbf{DQ}_W]}{\text{Ch}_{\max}(\mathbf{D}^2\mathbf{Q}_W)} \right\} \geq p + 2 + \frac{k - p}{2}. \tag{4.5}$$

When $d_1 = \dots = d_k$ and $\mathbf{Q} = \mathbf{I}_m$, the condition (4.5) is written as $k - p \geq 2(m + 2)$, and improvement is guaranteed for large k . However, those sufficient conditions for the improvement are restrictive in the case of different d_i 's with large d_1 and small d_k .

2. *Unconditional risk.* We next investigate the dominance property relative to the unconditional risk. Let $\Delta^U = R^U(\pi, \hat{\boldsymbol{\mu}}^{\text{CEB}}(\hat{\lambda}, \mathbf{t})) - R^U(\pi, \hat{\boldsymbol{\mu}}^{\text{CM}}(\mathbf{t}))$. Since it is hard to evaluate Δ^U exactly, we shall approximate Δ^U / k up to $O(k^{-3/2})$ for large k .

Proposition 4.3. *Assume that the elements of \mathbf{X} and \mathbf{W} are uniformly bounded and $\mathbf{X}'\mathbf{V}^{-1}(\lambda) \times \mathbf{X}/k$ is positive definite and converges to a positive definite matrix. Assume also that d_i 's are bounded above and bounded away from zero. Then, Δ^U / k is approximated as $\Delta^U / k = \Delta_{\text{APR}}(\lambda) / k + O(k^{-3/2})$, where*

$$\begin{aligned} \Delta_{\text{APR}}(\lambda) &= -\text{tr}[\mathbf{DQ}_W\mathbf{D}\mathbf{V}^{-1}(\lambda)] + \text{tr}[(\mathbf{X}'\mathbf{V}^{-1}(\lambda)\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{V}^{-1}(\lambda)\mathbf{X}] \\ &\quad + \text{tr}[\mathbf{DQ}_W\mathbf{D}\mathbf{V}^{-3}(\lambda)] \frac{2k}{(\text{tr}[\mathbf{V}^{-1}(\lambda)])^2}. \end{aligned} \tag{4.6}$$

A necessary condition for $\Delta_{\text{APR}}(\lambda) \leq 0$ is given by

$$\text{tr}[\mathbf{D}\mathbf{Q}\mathbf{W}] \geq \text{tr}[(\mathbf{X}'\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}\mathbf{W}\mathbf{X}] + \text{tr}[\mathbf{Q}\mathbf{W}\mathbf{D}^{-1}] \frac{2k}{(\text{tr}[\mathbf{D}^{-1}])^2}. \quad (4.7)$$

A sufficient condition for $\Delta_{\text{APR}}(\lambda) \leq 0$ is that

$$\min_{\lambda > 0} \left\{ \frac{\text{tr}[\mathbf{D}\mathbf{Q}\mathbf{W}\mathbf{D}\mathbf{V}^{-1}(\lambda)]}{\text{Ch}_{\max}(\mathbf{D}\mathbf{Q}\mathbf{W}\mathbf{D}\mathbf{V}^{-1}(\lambda))} \right\} \geq p + 2 \frac{k \text{tr}[\mathbf{D}^{-2}]}{(\text{tr}[\mathbf{D}^{-1}])^2}. \quad (4.8)$$

The proof is given in the [Appendix](#). The approximation (4.6) was derived by Datta, Rao and Smith [8]. When $\Delta_{\text{APR}}(\lambda) \leq 0$ for any $\lambda > 0$, it is said that $\hat{\boldsymbol{\mu}}^{\text{CEB}}(\hat{\lambda}, \mathbf{t})$ improves on $\hat{\boldsymbol{\mu}}^{\text{CM}}(\mathbf{t})$ in terms of the second-order approximation of the unconditional risk. Using the same arguments as in (4.5), it follows from (4.8) that the inequality $\Delta_{\text{APR}}(\lambda) \leq 0$ holds if

$$\max \left\{ \frac{\text{tr}[\mathbf{D}^2\mathbf{Q}\mathbf{W}]}{d_1 \text{Ch}_{\max}(\mathbf{D}\mathbf{Q}\mathbf{W})}, \frac{d_k \text{tr}[\mathbf{D}\mathbf{Q}\mathbf{W}]}{\text{Ch}_{\max}(\mathbf{D}^2\mathbf{Q}\mathbf{W})} \right\} \geq p + 2 \frac{k \text{tr}[\mathbf{D}^{-2}]}{(\text{tr}[\mathbf{D}^{-1}])^2}. \quad (4.9)$$

The necessary condition (4.7) is useful in the sense that if the condition (4.7) is violated, then $\hat{\boldsymbol{\mu}}^{\text{CEB}}(\hat{\lambda}, \mathbf{t})$ does not improve on $\hat{\boldsymbol{\mu}}^{\text{CM}}(\mathbf{t})$ in terms of the second-order approximation of the unconditional risk. This means that $\hat{\boldsymbol{\mu}}^{\text{CEB}}(\hat{\lambda}, \mathbf{t})$ should satisfy the condition (4.7) at least.

Remark 4.1. Propositions 4.1, 4.2 and 4.3 give us the conditions for the improvement by the constrained empirical Bayes estimator (4.2). By replacing $\mathbf{Q}\mathbf{W}$ with \mathbf{Q} , these propositions can provide the conditions under which the empirical Bayes estimator given in (4.1) improves on \mathbf{y} .

4.2. Unconstrained empirical Bayes estimator satisfying constraints

In this subsection, we set up a prior distribution which results in an unconstrained empirical Bayes and minimax estimator satisfying the constraint in the Fay–Herriot model with heteroscedastic variances and covariates as regressors.

Case 1: $\mathbf{t}(\mathbf{y}) = \mathbf{W}'\mathbf{y}$. Recall the arguments as in Case 1 of Section 3.2. Since $\boldsymbol{\xi}_3 = (\mathbf{H}_1 - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{H}_2)\mathbf{Q}^{1/2}\boldsymbol{\mu}$ and we set up the linear regression structure $\mathbf{X}\boldsymbol{\beta}$ for $\boldsymbol{\mu}$, it may be reasonable to assume the prior distribution $\boldsymbol{\xi}_3|\lambda \sim \mathcal{N}_{k-m}(\mathbf{X}_3\boldsymbol{\beta}, \lambda\mathbf{I}_{k-m})$ for $\boldsymbol{\xi}_3$ and to assume the uniform prior for $\boldsymbol{\xi}_2$, where $\mathbf{X}_3 = (\mathbf{H}_1 - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{H}_2)\mathbf{Q}^{1/2}\mathbf{X}$, which is assumed to be of rank p . Combining the contents in Sections 3.2 and 4.1, we get the empirical Bayes estimator given by

$$\hat{\boldsymbol{\mu}}^{\text{EB1}} = \mathbf{Q}^{-1/2}\mathbf{H}'_1\{\hat{\boldsymbol{\xi}}_3^{\text{EB}}(\mathbf{z}_3) + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{z}_2\} + \mathbf{Q}^{-1/2}\mathbf{H}'_2\mathbf{z}_2.$$

Here the empirical Bayes estimator $\hat{\boldsymbol{\xi}}_3^{\text{EB}}(\mathbf{z}_3)$ is given as follows: Note that $\mathbf{z}_3|\boldsymbol{\xi}_3 \sim \mathcal{N}_{k-m}(\boldsymbol{\xi}_3, \mathbf{V}_{11.2})$ and $\boldsymbol{\xi}_3 \sim \mathcal{N}_{k-m}(\mathbf{X}_3\boldsymbol{\beta}, \lambda\mathbf{I})$. According to the arguments in Section 4.1, we estimate λ by $\hat{\lambda} = \max\{\lambda^*, 0\}$, where λ^* is the solution of the equation $\mathbf{z}'_3\mathbf{A}_3(\lambda^*)\mathbf{z}_3 = k - m - p$ for $\mathbf{A}_3(\lambda) = \mathbf{V}_3^{-1} - \mathbf{X}_3(\mathbf{X}'_3\mathbf{V}_3^{-1}\mathbf{X}_3)^{-1}\mathbf{X}'_3\mathbf{V}_3^{-1}$ for $\mathbf{V}_3 = \mathbf{V}_{11.2} + \lambda\mathbf{I}$. Then, the empirical Bayes estimator of $\boldsymbol{\xi}_3$ is written by

$$\hat{\boldsymbol{\xi}}_3^{\text{EB}}(\mathbf{z}_3) = \mathbf{z}_3 - \mathbf{V}_{11.2}(\mathbf{V}_{11.2} + \hat{\lambda}\mathbf{I}_{k-m})^{-1}\{\mathbf{z}_3 - \mathbf{X}_3\hat{\boldsymbol{\beta}}_3(\hat{\lambda})\} \quad (4.10)$$

for $\widehat{\boldsymbol{\beta}}_3(\lambda) = (\mathbf{X}'_3 \mathbf{V}_3^{-1} \mathbf{X}_3)^{-1} \mathbf{X}'_3 \mathbf{V}_3^{-1} \mathbf{z}_3$.

Clearly, $\widehat{\boldsymbol{\mu}}^{\text{EB1}}$ satisfies the constraint, namely, $\mathbf{W}'\widehat{\boldsymbol{\mu}}^{\text{EB1}} = \mathbf{W}'\mathbf{y}$. Since \mathbf{Q}_W and \mathbf{D} in Section 4.1 correspond to \mathbf{I}_{k-m} and $\mathbf{V}_{11.2}$, respectively. The dominance results for $\widehat{\boldsymbol{\mu}}^{\text{EB1}}$ follow from Propositions 4.2 and 4.3.

Proposition 4.4. *The unconstrained empirical Bayes estimator $\widehat{\boldsymbol{\mu}}^{\text{EB1}}$ satisfies the constraint $\mathbf{W}'\widehat{\boldsymbol{\mu}}^{\text{EB1}} = \mathbf{W}'\mathbf{y}$. It is also minimax in Γ in terms of the conditional risk if*

$$\max \left\{ \frac{\text{tr}[\mathbf{V}_{11.2}^2]}{\{\text{Ch}_{\max}(\mathbf{V}_{11.2})\}^2}, \frac{\text{Ch}_{\min}(\mathbf{V}_{11.2})}{\text{Ch}_{\max}(\mathbf{V}_{11.2})} \text{tr}[\mathbf{V}_{11.2}] \right\} \geq p + 2 + \frac{k-p}{2}. \tag{4.11}$$

In the sense of the second-order approximation relative to the unconditional risk, a sufficient condition for $\widehat{\boldsymbol{\mu}}^{\text{EB1}}$ to improve on \mathbf{y} is

$$\max \left\{ \frac{\text{tr}[\mathbf{V}_{11.2}^2]}{\{\text{Ch}_{\max}(\mathbf{V}_{11.2})\}^2}, \frac{\text{Ch}_{\min}(\mathbf{V}_{11.2})}{\text{Ch}_{\max}(\mathbf{V}_{11.2})} \text{tr}[\mathbf{V}_{11.2}] \right\} \geq p + 2 \frac{k \text{tr}[\mathbf{V}_{11.2}^{-2}]}{(\text{tr}[\mathbf{V}_{11.2}^{-1}])^2}, \tag{4.12}$$

and a necessary condition for the improvement is given by

$$\text{tr}[\mathbf{V}_{11.2}] \geq \text{tr}[(\mathbf{X}'_3 \mathbf{V}_{11.2}^{-1} \mathbf{X}_3)^{-1} \mathbf{X}'_3 \mathbf{X}_3] + 2k / \text{tr}[\mathbf{V}_{11.2}^{-1}]. \tag{4.13}$$

Case 2: $\mathbf{t}(\mathbf{y}) = \mathbf{t}_0$. Recall the arguments as in Case 2 of Section 3.2. If we assume the prior distribution that $\boldsymbol{\xi}_2 = \boldsymbol{\xi}_0$ with probability one for $\boldsymbol{\xi}_0 = (\mathbf{W}'\mathbf{Q}^{-1/2}\mathbf{H}'_2)^{-1}\mathbf{t}_0$, it is seen that $\boldsymbol{\xi}_4 = \boldsymbol{\xi}_1$. Since $\boldsymbol{\xi}_1 = \mathbf{H}_1\mathbf{Q}^{1/2}\boldsymbol{\mu}$, it is reasonable to assume the prior distribution $\boldsymbol{\xi}_1|\lambda \sim \mathcal{N}_{k-m}(\mathbf{H}_1\mathbf{Q}^{1/2}\mathbf{X}\boldsymbol{\beta}, \lambda\mathbf{I}_{k-m})$ for $\boldsymbol{\xi}_1$. Then from (3.6), the generalized Bayes estimator is given by

$$\widehat{\boldsymbol{\mu}}^{\text{EB2}} = \mathbf{Q}^{-1/2}\mathbf{H}'_1\widehat{\boldsymbol{\xi}}_1^{\text{EB}}(\mathbf{z}_4) + \mathbf{Q}^{-1/2}\mathbf{H}'_2\boldsymbol{\xi}_0, \tag{4.14}$$

where $\widehat{\boldsymbol{\xi}}_1^{\text{EB}}(\mathbf{z}_4)$ has the same form as $\widehat{\boldsymbol{\xi}}_3^{\text{EB}}(\mathbf{z}_3)$ given in (4.10) except replacing \mathbf{z}_3 and \mathbf{X}_3 with \mathbf{z}_4 and $\mathbf{H}_1\mathbf{Q}^{1/2}\mathbf{X}$, respectively. It is assumed that $\mathbf{H}_1\mathbf{Q}^{1/2}\mathbf{X}$ is of rank p .

Clearly, $\widehat{\boldsymbol{\mu}}^{\text{EB2}}$ satisfies the constraint, namely, $\mathbf{W}'\widehat{\boldsymbol{\mu}}^{\text{EB2}} = \mathbf{t}_0$. The improvement of $\widehat{\boldsymbol{\mu}}^{\text{EB2}}$ follows from Propositions 4.1 and 4.4. When $\boldsymbol{\mu}$ is restricted to $\mathbf{W}'\boldsymbol{\mu} = \mathbf{t}_0$ or $\boldsymbol{\xi}_2 = \boldsymbol{\xi}_0$, $\widehat{\boldsymbol{\mu}}^{\text{EB2}}$ is minimax in Γ from Proposition 3.4.

Proposition 4.5. *The unconstrained empirical Bayes estimator $\widehat{\boldsymbol{\mu}}^{\text{EB2}}$ satisfies the constraint $\mathbf{W}'\widehat{\boldsymbol{\mu}}^{\text{EB2}} = \mathbf{t}_0$ and dominates the estimator $\mathbf{Q}^{-1/2}\mathbf{H}'_1\mathbf{z}_4 + \mathbf{Q}^{-1/2}\mathbf{H}'_2\boldsymbol{\xi}_0$ in terms of the conditional risk if the condition (4.11) holds. This implies the minimaxity of $\widehat{\boldsymbol{\mu}}^{\text{EB2}}$ within Γ_C . When $\boldsymbol{\mu}$ is restricted on $\mathbf{W}'\boldsymbol{\mu} = \mathbf{t}_0$, $\widehat{\boldsymbol{\mu}}^{\text{EB2}}$ is minimax in Γ under the condition (4.11).*

In the sense of the second-order approximation relative to the unconditional risk, a sufficient condition for $\widehat{\boldsymbol{\mu}}^{\text{EB2}}$ to improve on \mathbf{y} is given by (4.12), and a necessary condition for the improvement is given by

$$\text{tr}[\mathbf{V}_{11.2}] \geq \text{tr}[(\mathbf{X}'_4 \mathbf{V}_{11.2}^{-1} \mathbf{X}_4)^{-1} \mathbf{X}'_4 \mathbf{X}_4] + 2k / \text{tr}[\mathbf{V}_{11.2}^{-1}] \tag{4.15}$$

for $\mathbf{X}_4 = \mathbf{H}_1\mathbf{Q}^{1/2}\mathbf{X}$.

4.3. Simulation study

We investigate the unconditional risk behaviors of the constrained estimators by simulation. We consider the Fay–Herriot model (4.1) with $k = 15$, $\lambda = 1$ and four d_i -patterns: (a) 0.5, 0.5, 0.4, 0.3, 0.3; (b) 0.7, 0.6, 0.5, 0.4, 0.3; (c) 2.0, 0.6, 0.5, 0.4, 0.2; (d) 4.0, 0.6, 0.5, 0.4, 0.1. Patterns (b)–(d) are treated by Datta, Rao and Smith [8], and pattern (a) is less variable in d_i -values, while pattern (d) has larger variability. There are five groups G_1, \dots, G_5 and three small areas in each group. The sampling variances d_i are the same for areas within the same group. For the matrix of covariates \mathbf{X} , the column vectors of \mathbf{X}' are generated as random vectors from $\mathcal{N}_p(\mathbf{0}, (1 - 0.2)\mathbf{I} + 0.2\mathbf{j}_k\mathbf{j}'_k)$ where \mathbf{j}_k is the k -dimensional vector with all the elements ones. Each element of $\boldsymbol{\beta}$ is generated as $1 + 4u$ where $u \sim U(0, 1)$, the uniform distribution on $(0, 1)$.

In this simulation, we treat the case that $\mathbf{W} = \mathbf{D}^{-1}\mathbf{j}_k$, $t_0 = 3\mathbf{W}'\mathbf{X}\mathbf{j}_k$ and $\mathbf{Q} = \mathbf{I}$, \mathbf{D}^{-1} for $m = 1$ and $p = 2$. We compare the unconditional risks $R^U(\pi, \hat{\boldsymbol{\mu}})$ for the five estimators of $\boldsymbol{\mu}$: the crude estimator \mathbf{y} , the empirical Bayes estimator EB given in (4.1), the constrained empirical Bayes estimator CB in (4.2), the unconstrained empirical Bayes estimator UC1 in (4.10) for Case 1 and the unconstrained empirical Bayes estimator UC2 in (4.14) for Case 2, where Case 1 and Case 2 denote the constraints $\mathbf{t}(\mathbf{y}) = \mathbf{W}'\mathbf{y}$ and $\mathbf{t}(\mathbf{y}) = \mathbf{t}_0$, respectively. The unconditional risks of these estimators are computed as average values based on 10,000 simulation runs, and those values are reported in Table 1, where Case 2* treats the unconditional risks for $\boldsymbol{\mu}$ restricted to $\mathbf{W}'\boldsymbol{\mu} = t_0$. It is noted that \mathbf{y} and EB do not satisfy the constraints. The values of the column of \mathbf{y} correspond to the minimax risks for Case 1 and Case 2*, and it is revealed that EB, CB, UC1 and UC2 have smaller risks than \mathbf{y} . For Case 1, the risks of the estimators CB and UC1 with the constraints are slightly larger than those of EB. It is interesting to note that the difference between Case 2 and Case 2* supports Proposition 2.4, namely, CB and UC2 improve on \mathbf{y} when $\boldsymbol{\mu}$ is restricted to $\mathbf{W}'\boldsymbol{\mu} = t_0$, while their maximum risks are beyond the risks of \mathbf{y} without the restriction.

We next investigate whether the conditions for the improvement derived in Sections 4.1 and 4.2 are satisfied or not. Table 2 reports this investigation where + is marked if the condition is satisfied, otherwise, – is marked. For improvement by CB, the sufficient condition relative to the conditional risk is (4.5), denoted by SR, and the sufficient and necessary conditions in terms

Table 1. Values of unconditional risks of the constrained estimators for $\lambda = 1$

Q	d_i	y	EB	Case 1		Case 2		Case 2*	
				CB	UC1	CB	UC2	CB	UC2
Q = I	(a)	6.00	4.76	4.84	4.88	9.24	9.40	4.48	4.64
	(b)	7.51	5.53	5.63	5.70	9.68	9.90	5.22	5.45
	(c)	11.05	6.40	6.45	6.65	7.22	7.41	6.15	6.34
	(d)	16.88	6.60	6.61	7.26	14.60	19.92	6.47	11.79
Q = D ⁻¹	(a)	14.93	11.92	12.02	12.28	38.88	39.14	11.02	11.28
	(b)	14.99	11.49	11.72	11.87	13.17	13.31	10.74	10.89
	(c)	14.99	10.90	11.05	11.76	13.47	14.17	10.09	10.80
	(d)	14.99	10.57	10.68	12.07	26.91	28.30	9.70	11.09

Table 2. Whether the conditions for the improvement are satisfied or not? When the condition is satisfied, + is marked, and otherwise, - is marked

Q	d_i	EB			CB			UC1			UC2		
		SR	SR ^U	NR ^U	SR	SR ^U	NR ^U	SR	SR ^U	NR ^U	SR	SR ^U	NR ^U
Q = I	(a)	-	+	+	-	+	+	-	+	+	-	+	+
	(b)	-	+	+	-	+	+	-	+	+	-	+	+
	(c)	-	-	+	-	-	+	-	-	+	-	-	+
	(d)	-	-	+	-	-	+	-	-	+	-	-	+
Q = D⁻¹	(a)	+	+	+	-	+	+	+	+	+	+	+	+
	(b)	-	+	+	-	+	+	+	+	+	+	+	+
	(c)	-	-	+	-	-	+	+	+	+	+	+	+
	(d)	-	-	+	-	-	+	+	+	+	+	+	+

of the second-order approximation relative to the unconditional risk are given by (4.9) and (4.7), respectively, denoted by SR^U and NR^U. As noted in Remark 4.1, similar conditions for EB can be given by SR, SR^U and NR^U by replacing **Q_W** with **Q**. SR and SR^U are given by (4.11) and (4.12) for the improvement by UC1 and UC2. The necessary conditions NR^U for UC1 and UC2 are given by (4.13) and (4.15), respectively. As seen from Table 2, the sufficient conditions SR under the conditional risks for EB and CB are very restrictive in both cases of **Q = I** and **Q = D⁻¹**, and AR for UC1 and UC2 are also restrictive for **Q = I**. That is, the conditions SR are not satisfied in most cases. It should be noted that this does not imply that those estimators do not improve on **y**, because the necessary conditions NR^U are always satisfied. For the estimators UC1 and UC2, all the conditions for the improvement are satisfied relative to the loss $(\hat{\mu} - \mu)' \mathbf{D}^{-1} (\hat{\mu} - \mu)$ for **Q = D⁻¹**.

5. Concluding remarks

Benchmarking has been recognized as an important issue in small area problems, and constrained Bayesian estimators have been studied in the literature. However, little has been known about decision-theoretic properties such as admissibility and minimaxity for constrained generalized Bayes estimators. In this paper, we have clarified admissibility, minimaxity and dominance properties of constrained estimators by decomposing the conditional risk function into two pieces: one depends on the estimator, but the other does not depend on the estimator. In the context of a multivariate normal population, we have provided a canonical form, which allows us to establish admissibility and inadmissibility of the constrained uniform-prior generalized Bayes estimator. We have also derived a condition on the prior distribution such that the resulting unconstrained generalized Bayes estimator automatically satisfies the constraint. Finally, we have provided constrained empirical Bayes and improved estimators in the Fay–Herriot model.

Although a constrained empirical Bayes estimator is treated in Section 4, it is not admissible. To develop admissible and minimax estimators, we would need to consider hierarchical prior

distributions and to investigate admissibility and minimaxity of the resulting hierarchical generalized Bayes estimators. Berger and Robert [3], Berger and Strawderman [4] and Kubokawa and Strawderman [15] have studied the admissibility and minimaxity of hierarchical Bayes estimators. The extension of their results to the setup of this paper seems a reasonable goal and is one that we plan to study.

Appendix: Proofs

Proof of Proposition 4.1. We first prove Proposition 4.1 which give us the sufficient condition for the constrained empirical Bayes estimator $\hat{\mu}^{CEB}(\hat{\lambda}, \mathbf{t})$ to improve on the constrained uniform-prior generalized Bayes estimator $\hat{\mu}^{CM}(\mathbf{t})$. The arguments as in Shinozaki and Chang ([20,21]) are useful for the proof. The conditional risk difference of the two estimators is written as

$$\begin{aligned} \Delta &= E[(\hat{\mu}^{CEB}(\hat{\lambda}, \mathbf{t}) - \mu)' \mathbf{Q}(\hat{\mu}^{CEB}(\hat{\lambda}, \mathbf{t}) - \mu) | \mu] - E[(\hat{\mu}^{CM}(\mathbf{t}) - \mu)' \mathbf{Q}(\hat{\mu}^{CM}(\mathbf{t}) - \mu) | \mu] \\ &= E[(\hat{\mu}^{EB}(\hat{\lambda}) - \mu)' \mathbf{Q}_W(\hat{\mu}^{EB}(\hat{\lambda}) - \mu) | \mu] - E[(\mathbf{y} - \mu)' \mathbf{Q}_W(\mathbf{y} - \mu) | \mu], \end{aligned}$$

where $\mathbf{Q}_W = \mathbf{Q} - \mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}\mathbf{W}'$. It is noted that \mathbf{Q}_W is of rank $k - m$ and that $E[(\mathbf{y} - \mu)' \mathbf{Q}_W(\mathbf{y} - \mu) | \mu] = \text{tr}[\mathbf{D}\mathbf{Q}_W] = \text{tr}[\mathbf{D}\mathbf{Q}] - \text{tr}[\mathbf{W}'\mathbf{D}\mathbf{W}(\mathbf{W}'\mathbf{Q}^{-1}\mathbf{W})^{-1}]$. The risk difference is written as

$$\Delta = -2E[(\mathbf{y} - \mu)' \mathbf{Q}_W \mathbf{D} \mathbf{A}(\hat{\lambda}) \mathbf{y} | \mu] + E[\mathbf{y}' \mathbf{A}(\hat{\lambda}) \mathbf{D} \mathbf{Q}_W \mathbf{D} \mathbf{A}(\hat{\lambda}) \mathbf{y} | \mu]. \tag{A.1}$$

Using the Stein identity given in Stein [22], we can rewrite the cross product term as

$$E[(\mathbf{y} - \mu)' \mathbf{Q}_W \mathbf{D} \mathbf{A}(\hat{\lambda}) \mathbf{y} | \mu] = E[\nabla' \{ \mathbf{D} \mathbf{Q}_W \mathbf{D} \mathbf{A}(\hat{\lambda}) \mathbf{y} \} | \mu].$$

Let $\mathbf{G}(\hat{\lambda}) = (g_{ij}(\hat{\lambda})) = \mathbf{D} \mathbf{Q}_W \mathbf{D} \mathbf{A}(\hat{\lambda})$. Then

$$\begin{aligned} \nabla' \{ \mathbf{G}(\hat{\lambda}) \mathbf{y} \} &= \sum_{i,j} \frac{\partial}{\partial y_i} \{ g_{ij}(\hat{\lambda}) y_j \} \\ &= \sum_i g_{ii}(\hat{\lambda}) + \sum_{i,j} y_j \left\{ \frac{d}{d\lambda} g_{ij}(\lambda) \Big|_{\lambda=\hat{\lambda}} \right\} \frac{\partial \hat{\lambda}}{\partial y_i} \\ &= \text{tr}[\mathbf{D} \mathbf{Q}_W \mathbf{D} \mathbf{A}(\hat{\lambda})] + \mathbf{y}' \left\{ \frac{d}{d\lambda} \mathbf{A}(\lambda) \Big|_{\lambda=\hat{\lambda}} \right\} \mathbf{D} \mathbf{Q}_W \mathbf{D} (\nabla \hat{\lambda}), \end{aligned}$$

since $g_{ij}(\hat{\lambda})$ depends on \mathbf{y} through $\hat{\lambda}$. Differentiating $\mathbf{A}(\lambda)$ with respect to λ for $\mathbf{A}(\lambda)$ given in (4.1), we can see that

$$\frac{d}{d\lambda} \mathbf{A}(\lambda) = -\mathbf{A}^2(\lambda), \tag{A.2}$$

which can be used to get the expression $\Delta = E[\widehat{\Delta}|\boldsymbol{\mu}]$ where

$$\begin{aligned} \widehat{\Delta}(\hat{\lambda}) &= -2 \operatorname{tr}[\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})] + 2\mathbf{y}'\mathbf{A}^2(\hat{\lambda})\mathbf{DQ}_w \mathbf{D}(\nabla \hat{\lambda}) \\ &\quad + \mathbf{y}'\mathbf{A}(\hat{\lambda})\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})\mathbf{y} \end{aligned}$$

for $\nabla = (\partial/\partial y_1, \dots, \partial/\partial y_k)'$.

Differentiating $\mathbf{y}'\mathbf{A}(\hat{\lambda})\mathbf{y} = k - p$ with respect to \mathbf{y} and using the implicit function theorem, we get the equation $2\mathbf{A}(\hat{\lambda})\mathbf{y} - \mathbf{y}'\mathbf{A}^2(\hat{\lambda})\mathbf{y}\nabla \hat{\lambda} = \mathbf{0}$ in the case of $0 < \hat{\lambda}$, or

$$\nabla \hat{\lambda} = \frac{2}{\mathbf{y}'\mathbf{A}^2(\hat{\lambda})\mathbf{y}}\mathbf{A}(\hat{\lambda})\mathbf{y}I(0 < \hat{\lambda}).$$

Thus, $\widehat{\Delta}$ is expressed as

$$\begin{aligned} \widehat{\Delta}(\hat{\lambda}) &= -2 \operatorname{tr}[\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})] + 4 \frac{\mathbf{y}'\mathbf{A}^2(\hat{\lambda})\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})\mathbf{y}}{\mathbf{y}'\mathbf{A}^2(\hat{\lambda})\mathbf{y}} I(0 < \hat{\lambda}) \\ &\quad + \mathbf{y}'\mathbf{A}(\hat{\lambda})\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})\mathbf{y}, \end{aligned}$$

where $I(A)$ is the indicator function such that $I(A) = 1$ if A is true, and otherwise, $I(A) = 0$. It is observed that

$$\begin{aligned} \mathbf{y}'\mathbf{A}(\hat{\lambda})\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})\mathbf{y} &\leq \mathbf{y}'\mathbf{A}(\hat{\lambda})\mathbf{y} \sup_{\mathbf{x}} \left\{ \frac{\mathbf{x}'\mathbf{A}(\hat{\lambda})\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})\mathbf{x}}{\mathbf{x}'\mathbf{A}(\hat{\lambda})\mathbf{x}} \right\} \\ &\leq (k - p) \times \operatorname{Ch}_{\max}(\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})), \\ \frac{\mathbf{y}'\mathbf{A}^2(\hat{\lambda})\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})\mathbf{y}}{\mathbf{y}'\mathbf{A}^2(\hat{\lambda})\mathbf{y}} I(\hat{\lambda} > 0) &\leq \sup_{\mathbf{x}} \left\{ \frac{\mathbf{x}'\mathbf{A}^2(\hat{\lambda})\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})\mathbf{x}}{\mathbf{x}'\mathbf{A}^2(\hat{\lambda})\mathbf{x}} \right\} \\ &= \operatorname{Ch}_{\max}(\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})). \end{aligned}$$

Hence,

$$\widehat{\Delta}(\hat{\lambda}) \leq -2 \operatorname{tr}[\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})] + (k - p + 4) \operatorname{Ch}_{\max}(\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})), \tag{A.3}$$

which proves Proposition 4.1. □

Proof of Proposition 4.3. We next prove Proposition 4.3. The unconditional risk difference can be written from (A.1) as

$$\begin{aligned} \Delta^U &= R^U(\pi, \widehat{\boldsymbol{\mu}}^{\text{CEB}}(\hat{\lambda}, \mathbf{t})) - R^U(\pi, \widehat{\boldsymbol{\mu}}^{\text{CM}}(\mathbf{t})) \\ &= -2E[(\mathbf{y} - E[\boldsymbol{\mu}|\mathbf{y}])'\mathbf{Q}_w \mathbf{DA}(\hat{\lambda})\mathbf{y}] + E[\mathbf{y}'\mathbf{A}(\hat{\lambda})\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})\mathbf{y}]. \end{aligned}$$

Noting that $E[\boldsymbol{\mu}|\mathbf{y}] = \mathbf{y} - \mathbf{DV}(\lambda)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ and $\mathbf{A}(\hat{\lambda})\mathbf{X} = \mathbf{0}$, we see that

$$\Delta^U = -2E[\mathbf{u}'\mathbf{V}(\lambda)^{-1}\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})\mathbf{u}] + E[\mathbf{u}'\mathbf{A}(\hat{\lambda})\mathbf{DQ}_w \mathbf{DA}(\hat{\lambda})\mathbf{u}], \tag{A.4}$$

where \mathbf{u} is a random variable having $\mathcal{N}'_k(\mathbf{0}, \mathbf{V}(\lambda))$. We shall derive the second order approximation of Δ^U/k up to $O(k^{-1})$. To this end, $\mathbf{A}(\hat{\lambda})$ is approximated by the Taylor series expansion as

$$\mathbf{A}(\hat{\lambda}) = \mathbf{A}(\lambda) + \mathbf{A}^{(1)}(\lambda)(\hat{\lambda} - \lambda) + 2^{-1}\mathbf{A}^{(2)}(\lambda)(\hat{\lambda} - \lambda)^2 + [O(k^{-3/2})]_{k \times k},$$

where $\mathbf{A}^{(i)}(\lambda) = \partial^i \mathbf{A}(\lambda) / \partial \lambda^i$, $i = 1, 2$, and $[O(k^{-3/2})]_{k \times k}$ means that all elements of the matrix are of $O(k^{-3/2})$. Then

$$\begin{aligned} & E[\mathbf{u}'\mathbf{A}(\hat{\lambda})\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\hat{\lambda})\mathbf{u}] \\ &= E[\mathbf{u}'\mathbf{A}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)\mathbf{u} + 2\mathbf{u}'\mathbf{A}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{A}^{(1)}(\lambda)\mathbf{u}(\hat{\lambda} - \lambda) \\ &\quad + \mathbf{u}'\mathbf{A}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{A}^{(2)}(\lambda)\mathbf{u}(\hat{\lambda} - \lambda)^2 + \mathbf{u}'\mathbf{A}^{(1)}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{A}^{(1)}(\lambda)\mathbf{u}(\hat{\lambda} - \lambda)^2] + O(k^{-1/2}) \\ &= \text{tr}[\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)] - \text{tr}[(\mathbf{X}'\mathbf{V}^{-1}(\lambda)\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)\mathbf{X}] \\ &\quad + E[2\mathbf{u}'\mathbf{V}^{-1}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{A}^{(1)}(\lambda)\mathbf{u}(\hat{\lambda} - \lambda) + \mathbf{u}'\mathbf{V}^{-1}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{A}^{(2)}(\lambda)\mathbf{u}(\hat{\lambda} - \lambda)^2 \\ &\quad + \mathbf{u}'\mathbf{V}^{-2}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{V}^{-2}(\lambda)\mathbf{u}(\hat{\lambda} - \lambda)^2] + O(k^{-1/2}), \end{aligned}$$

since $\mathbf{u}'\mathbf{A}(\lambda)\mathbf{C}\mathbf{u} = \mathbf{u}'\mathbf{V}^{-1}(\lambda)\mathbf{C}\mathbf{u} - \mathbf{u}'\mathbf{V}^{-1}(\lambda)\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}(\lambda)\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\lambda)\mathbf{C}\mathbf{u} = \mathbf{u}'\mathbf{V}^{-1}(\lambda)\mathbf{C}\mathbf{u} + O_p(1)$ for a matrix $\mathbf{C} = [O(1)]_{k \times k}$, and $\mathbf{A}^{(1)}(\lambda) = -\mathbf{A}^2(\lambda)$. Similarly,

$$\begin{aligned} & -2E[\mathbf{u}'\mathbf{V}(\lambda)^{-1}\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\hat{\lambda})\mathbf{u}] \\ &= -2\text{tr}[\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)] - 2E[\mathbf{u}'\mathbf{V}^{-1}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{A}^{(1)}(\lambda)\mathbf{u}(\hat{\lambda} - \lambda)] \\ &\quad - E[\mathbf{u}'\mathbf{V}^{-1}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{A}^{(2)}(\lambda)\mathbf{u}(\hat{\lambda} - \lambda)^2] + O(k^{-1/2}). \end{aligned}$$

Since $E[\mathbf{u}'\mathbf{V}^{-2}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{V}^{-2}(\lambda)\mathbf{u}(\hat{\lambda} - \lambda)^2] = E[\mathbf{u}'\mathbf{V}^{-2}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{V}^{-2}(\lambda)\mathbf{u}] \text{Var}(\hat{\lambda}) + O(k^{-1/2})$ and $\mathbf{A}(\lambda)\mathbf{X} = \mathbf{0}$, it follows that

$$\Delta^U = -\text{tr}[\mathbf{DQ}_W\mathbf{D}\mathbf{A}(\lambda)] + \text{tr}[\mathbf{V}^{-3}(\lambda)\mathbf{DQ}_W\mathbf{D}] \text{Var}(\hat{\lambda}) + O(k^{-1/2}).$$

It is noted that $\text{Var}(\hat{\lambda}) = 2k/(\text{tr}[\mathbf{V}^{-1}(\lambda)])^2 + O(k^{-1/2})$ from Datta, Rao and Smith [8]. Hence, Δ^U/k can be approximated as $\Delta^U/k = \Delta_{\text{APR}}(\lambda)/k + O(k^{-3/2})$, where $\Delta_{\text{APR}}(\lambda)$ is given in (4.6). A necessary condition for $\Delta_{\text{APR}}(\lambda) \leq 0$ is that $\Delta_{\text{APR}}(0) \leq 0$, which is given in (4.7). To derive a sufficient condition, note that $\text{tr}[(\mathbf{X}'\mathbf{V}^{-1}(\lambda)\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\lambda)\mathbf{DQ}_W\mathbf{D}\mathbf{V}^{-1}(\lambda)\mathbf{X}] \leq p \text{Ch}_{\max}(\mathbf{DQ}_W\mathbf{D}\mathbf{V}^{-1}(\lambda))$ and that $\text{tr}[\mathbf{DQ}_W\mathbf{D}\mathbf{V}^{-3}(\lambda)] \leq \text{Ch}_{\max}(\mathbf{DQ}_W\mathbf{D}\mathbf{V}^{-1}(\lambda)) \text{tr}[\mathbf{V}^{-2}(\lambda)]$. By making the differentiation, it can be verified that $\text{tr}[\mathbf{V}^{-2}(\lambda)]/(\text{tr}[\mathbf{V}^{-1}(\lambda)])^2$ is decreasing in λ , so that $\text{tr}[\mathbf{V}^{-2}(\lambda)]/(\text{tr}[\mathbf{V}^{-1}(\lambda)])^2 \leq \text{tr}[\mathbf{D}^{-2}]/(\text{tr}[\mathbf{D}^{-1}])^2$. Thus,

$$\Delta_{\text{APR}}(\lambda) \leq -\text{tr}[\mathbf{DQ}_W\mathbf{D}\mathbf{V}^{-1}(\lambda)] + \text{Ch}_{\max}(\mathbf{DQ}_W\mathbf{D}\mathbf{V}^{-1}(\lambda)) \{p + 2k \text{tr}[\mathbf{D}^{-2}]/(\text{tr}[\mathbf{D}^{-1}])^2\},$$

which is expressed as (4.8). Therefore, we get Proposition 4.3. \square

Acknowledgements

We are grateful to the Editor, the Associate Editor and the reviewers for their valuable comments and helpful suggestions. This research was supported in part by Grant-in-Aid for Scientific Research # 21540114 and # 23243039 from Japan Society for the Promotion of Science. This work was partially supported by a grant from the Simons Foundation (# 209035 to William Strawderman).

References

- [1] Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83** 28–36.
- [2] Berger, J.O. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* **4** 223–226. [MR0397940](#)
- [3] Berger, J.O. and Robert, C. (1990). Subjective hierarchical Bayes estimation of a multivariate normal mean: On the frequentist interface. *Ann. Statist.* **18** 617–651. [MR1056330](#)
- [4] Berger, J.O. and Strawderman, W.E. (1996). Choice of hierarchical priors: Admissibility in estimation of normal means. *Ann. Statist.* **24** 931–951. [MR1401831](#)
- [5] Brown, L.D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903. [MR0286209](#)
- [6] Datta, G.S., Ghosh, M., Steorts, R. and Maples, J. (2011). Bayesian benchmarking with applications to small area estimation. *TEST* **20** 574–588. [MR2864715](#)
- [7] Datta, G.S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statist. Sinica* **10** 613–627. [MR1769758](#)
- [8] Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika* **92** 183–196. [MR2158619](#)
- [9] Fay, R.E. III and Herriot, R.A. (1979). Estimates of income for small places: An application of James–Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277. [MR0548019](#)
- [10] Frey, J. and Cressie, N. (2003). Some results on constrained Bayes estimators. *Statist. Probab. Lett.* **65** 389–399. [MR2039883](#)
- [11] Ghosh, M. (1992). Constrained Bayes estimation with applications. *J. Amer. Statist. Assoc.* **87** 533–540. [MR1173817](#)
- [12] Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statist. Sci.* **9** 55–93. With comments and a rejoinder by the authors. [MR1278679](#)
- [13] James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.* **1** 361–379. Berkeley, CA: Univ. California Press. [MR0133191](#)
- [14] Kubokawa, T. (2011). On measuring uncertainty of small area estimators with higher order accuracy. *J. Japan Statist. Soc.* **41** 93–119.
- [15] Kubokawa, T. and Strawderman, W.E. (2007). On minimaxity and admissibility of hierarchical Bayes estimators. *J. Multivariate Anal.* **98** 829–851. [MR2322131](#)
- [16] Pfeffermann, D. and Tiller, R. (2006). Small-area estimation with state-space models subject to benchmark constraints. *J. Amer. Statist. Assoc.* **101** 1387–1397. [MR2307572](#)
- [17] Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.* **85** 163–171. [MR1137362](#)
- [18] Rao, J.N.K. (2003). *Small Area Estimation*. *Wiley Series in Survey Methodology*. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons]. With a foreword by Graham Kalton. [MR1953089](#)

- [19] Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components*. *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. New York: Wiley. [MR1190470](#)
- [20] Shinozaki, N. and Chang, Y.T. (1993). Minimaxity of empirical Bayes estimators of the means of independent normal variables with unequal variances. *Comm. Statist. Theory Methods* **22** 2147–2169. [MR1238161](#)
- [21] Shinozaki, N. and Chang, Y.T. (1996). Minimaxity of empirical Bayes estimators shrinking toward the grand mean when variances are unequal. *Comm. Statist. Theory Methods* **25** 183–199. [MR1378954](#)
- [22] Stein, C.M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. [MR0630098](#)
- [23] Zacks, S. (1971). *The Theory of Statistical Inference*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. [MR0420923](#)

Received July 2011 and revised February 2012