CIRJE-F-809

Admissibility and Minimaxity of Benchmarked Shrinkage Estimators

Tatsuya Kubokawa
University of Tokyo

William E. Strawderman
Rutgers University

July 2011

Admissibility and Minimaxity of Benchmarked Shrinkage Estimators

Tatsuya Kubokawa*and William E. Strawderman†
University of Tokyo and *Rutgers University*

July 20, 2011

Abstract

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 admissibility and minimaxity 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, minimax or admissible estimators which automatically satisfy the benchmark constraint. We illustrate the results with several examples. Also, minimaxity of a benchmarked empirical Bayes estimator is shown in the Fay-Herriot model, a frequently used model in small area estimation.

MSC 2010: primary 62C15, 62C20; secondary 62C12, 62F10, 62F15

Key words and phrases: 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 model based 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

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

†Department of Statistics, Rutgers University, Piscataway, New Jersey, 08854-8019, E-mail: straw@stat.rutgers.edu

may differ substantially from the average derived from a direct estimate. The reader is referred to the articles of Datta, Ghosh, Steorts and Maples (2011) for an extended discussion of the background and desirability of benchmarking. Also see Frey and Cressie (2003), Ghosh (1992) and Pfeffermann and Tiller (2006) for the related issues. For good accounts of small area estimation, see Battese, Harter and Fuller (1988), Prasad and Rao (1990), Ghosh and Rao (1994), Rao (2003) and Datta, Rao and Smith (2005)

We study admissibility and minimaxity properties of benchmarked estimators in the context of a multivariate normal population by reducing the problem to one of studying these properties in a related derived problem. For certain benchmark constraints we develop estimators that improve on the benchmarked version of the usual (UMVUE or Generalized Bayes with respect to the uniform prior) estimator and are minimax and/or admissible in the benchmark problem. Also, for certain such problems we show that unconstrained solutions in the original (unbenchmarked) problem give unconstrained Bayes, minimax or admissible estimators which automatically satisfy the benchmark constraint. We illustrate the results with several examples.

Section 2 gives the general setup of the problem and presents the benchmarked version of a general estimator, as well as a decomposition of the risk of such an estimator into two pieces; one which depends on the risk of the unbenchmarked estimator in a related problem and one which depends on the parameter and the benchmark constraint but not the estimator in question. Admissibility considerations and sometimes minimaxity are then reduced to study of these properties in a related problem. Some preliminary admissibility and minimax results are also given in this section.

Section 3 is devoted to developing improved shrinkage benchmark estimators in the multivariate normal case. We present a canonical form useful for studying the problem and give two alternate decompositions which lead to several different classes of shrinkage benchmarked estimators. We give conditions under which these classes are minimax and/or admissible in the benchmarked problem by studying the related derived problem in the canonical setting.

Section 4 studies prior distributions in the original problem that result in estimators which automatically satisfy the benchmark constraint. The development is motivated by a simple example that illustrates the utility of placing a uniform prior on a portion of the parameter space so that the resulting generalized Bayes shrinkage estimators satisfy the constraint. A more general development is also given for certain benchmark constraints.

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 5, we consider this model and investigate minimaxity of a constrained emprical Bayes 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.

Finally, some concluding remarks are given in Section 6.

2 General Setup of the Constrained Problem

Let $\boldsymbol{X} = (X_1, \ldots, X_k)'$ be a $k \times 1$ random vector and consider estimation of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ by an estimator $\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{\mu}}(\boldsymbol{X}) = (\widehat{\mu}_1, \ldots, \widehat{\mu}_k)'$. Let $\boldsymbol{W}$ be a $k \times m$ matrix with rank m, $m < k$, and let $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{X})$ be a function from $\boldsymbol{R}^k$ to $\boldsymbol{R}^m$.The constraint we consider is to restrict the estimator $\widehat{\boldsymbol{\mu}}$ to satisfy the benchmark condition $\boldsymbol{W}'\widehat{\boldsymbol{\mu}} = \boldsymbol{t}(\boldsymbol{X})$, namely, $\widehat{\boldsymbol{\mu}} \in \Gamma_0$ for

$$\Gamma_0 = \{\widehat{\boldsymbol{\mu}} \in \Gamma \mid \boldsymbol{W}'\widehat{\boldsymbol{\mu}} = \boldsymbol{t}(\boldsymbol{X})\}.$$

We will restrict attention throughout to estimators in Γ, the class of estimators with second moments, i.e.,

$$\Gamma = \{\widehat{\boldsymbol{\mu}} \mid E[\widehat{\boldsymbol{\mu}}'\widehat{\boldsymbol{\mu}}] < \infty].$$

Typical examples of $\boldsymbol{t}(\boldsymbol{X})$ are $t(\boldsymbol{X}) = \overline{X}$ for $\overline{X} = k^{-1}\sum_{j=1}^{k} X_j$ and $t(\boldsymbol{X}) = t_0$, a constant, both of which are cases where $\boldsymbol{t}(\boldsymbol{X})$ are scalar valued functions.

Let the quadratic loss function be given by $L(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}; \boldsymbol{Q}) = (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})$ for a known $k \times k$ positive definite matrix $\boldsymbol{Q}$. The risk function of $\widehat{\boldsymbol{\mu}}$ is denoted by

$$R(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) = E[(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})].$$

A benchmarked Bayes estimator $\widehat{\boldsymbol{\mu}}^{BM}$ is defined as the estimator $\widehat{\boldsymbol{\mu}}$ which minimizes the posterior risk function $E[(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})|\boldsymbol{X}]$ subject to $\widehat{\boldsymbol{\mu}} \in \Gamma_0$, where $E[\cdot|\boldsymbol{X}]$ denotes a posterior expectation given $\boldsymbol{X}$. Noting that

$$E[(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})|\boldsymbol{X}] = E[(\widehat{\boldsymbol{\mu}}^B - \boldsymbol{\mu})'\boldsymbol{Q}(\widehat{\boldsymbol{\mu}}^B - \boldsymbol{\mu})|\boldsymbol{X}] + (\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\mu}}^B)'\boldsymbol{Q}(\widehat{\boldsymbol{\mu}} - \widehat{\boldsymbol{\mu}}^B)$$

for the Bayes estimator $\widehat{\boldsymbol{\mu}}^B = E[\boldsymbol{\mu}|\boldsymbol{X}]$, Datta, Ghosh, Steort and Maples (2011, Test) showed that the constrained Bayes estimator is given by

$$\widehat{\boldsymbol{\mu}}^B + \boldsymbol{Q}^{-1}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\{\boldsymbol{t}(\boldsymbol{X}) - \boldsymbol{W}'\widehat{\boldsymbol{\mu}}^B\}, \tag{2.1}$$

which can be expressed as

$$(\boldsymbol{I} - \boldsymbol{P_W})\widehat{\boldsymbol{\mu}}^B + \boldsymbol{Q}^{-1}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{t}(\boldsymbol{X}),$$

where

$$\boldsymbol{P_W} = \boldsymbol{Q}^{-1}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{W}'.$$

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

$$\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t}) = (\boldsymbol{I} - \boldsymbol{P_W})\widehat{\boldsymbol{\mu}} + \boldsymbol{Q}^{-1}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{t}(\boldsymbol{X}), \tag{2.2}$$

and denote the class by

$$\Gamma_1 = \{\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t}) \mid \widehat{\boldsymbol{\mu}} \in \Gamma\}.$$

It is seen that

$$\Gamma_1 \subset \Gamma_0 \subset \Gamma.$$

To evaluate the risk of the estimator $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t})$, note that

$$\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t}) - \boldsymbol{\mu} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{W}})(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) + \boldsymbol{Q}^{-1}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\{\boldsymbol{t}(\boldsymbol{X}) - \boldsymbol{W}'\boldsymbol{\mu}\}.$$

Also note that $\boldsymbol{W}'(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{W}}) = \boldsymbol{0}$ and

$$(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{W}})'\boldsymbol{Q}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{W}}) = \boldsymbol{Q} - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{W}' = \boldsymbol{Q}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{W}}).$$

Then the risk function of $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t})$ relative to the loss $L(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}; \boldsymbol{Q})$ 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 risk function of $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t})$ relative to the loss $L(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}; \boldsymbol{Q})$ is expressed as*

$$R(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t})) = R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) + R_2(\boldsymbol{\mu}, \boldsymbol{t}), \tag{2.3}$$

where

$$\begin{aligned} R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) =& E[(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})'\boldsymbol{Q}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{W}})(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})], \\ R_2(\boldsymbol{\mu}, \boldsymbol{t}) =& E[(\boldsymbol{t}(\boldsymbol{X}) - \boldsymbol{W}'\boldsymbol{\mu})'(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}(\boldsymbol{t}(\boldsymbol{X}) - \boldsymbol{W}'\boldsymbol{\mu})]. \end{aligned} \tag{2.4}$$

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

Proposition 2.1 *For two estimators $\widehat{\boldsymbol{\mu}}_1$ and $\widehat{\boldsymbol{\mu}}_2$ in Γ, and the corresponding benchmarked estimators $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}_1, \boldsymbol{t})$ and $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}_2, \boldsymbol{t})$ in Γ_1, $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}_1, \boldsymbol{t})$ dominates $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}_2, \boldsymbol{t})$ relative to the loss $L(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}; \boldsymbol{Q})$ if and only if $\widehat{\boldsymbol{\mu}}_1$ dominates $\widehat{\boldsymbol{\mu}}_2$ relative to the loss $L(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}; \boldsymbol{Q}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{W}}))$.*

This proposition implies the following proposition concerning admissibility.

Proposition 2.2 *Assume that $\widehat{\boldsymbol{\mu}} \in \Gamma$. Then the estimator $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t})$ is admissible in Γ_1 in terms of the risk $R(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}^C)$ if and only if $\widehat{\boldsymbol{\mu}}$ is admissible in Γ in terms of the risk $R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}})$.*

The above propositions show that dominance properties and admissibility of an estimator $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t})$ can be reduced to those of the estimator $\widehat{\boldsymbol{\mu}}$ in terrms of the risk $R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}})$.

Concerning minimaxity, on the other hand, it is seen that the estimator $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}^*, \boldsymbol{t})$ is minimax if and only if

$$\begin{aligned} \inf_{\widehat{\boldsymbol{\mu}} \in \Gamma_1} \sup_{\boldsymbol{\mu}} R(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) =& \inf_{\widehat{\boldsymbol{\mu}} \in \Gamma} \sup_{\boldsymbol{\mu}} \Big\{ R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) + R_2(\boldsymbol{\mu}, \boldsymbol{t}) \Big\} \\ =& \sup_{\boldsymbol{\mu}} \Big\{ R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}^*) + R_2(\boldsymbol{\mu}, \boldsymbol{t}) \Big\}. \end{aligned}$$

This equality holds if there exists a sequence $\{\boldsymbol{\mu}_n\}_{n=1,2,\cdots}$ such that

$$\liminf_{n\to\infty} \Big\{ R_1(\boldsymbol{\mu}_n, \widehat{\boldsymbol{\mu}}^*) + R_2(\boldsymbol{\mu}_n, \boldsymbol{t}) \Big\} = \sup_{\boldsymbol{\mu}} \Big\{ R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}^*) + R_2(\boldsymbol{\mu}, \boldsymbol{t}) \Big\}. \tag{2.5}$$

Proposition 2.3 *If* $\widehat{\boldsymbol{\mu}}^*$ *satisfies the condition* (2.5), *then the estimator* $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}^*, \boldsymbol{t})$ *is minimax within* Γ_1.

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

(A1) Assume that $R_2(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}})$ does not depend on the unknown parameters.

Proposition 2.4 *Assume the condition* (A1). *Then the estimator* $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}^*, \boldsymbol{t})$ *is minimax within* Γ_1 *if and only if* $\widehat{\boldsymbol{\mu}}^*$ *is minimax in terms of the risk* $R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}})$ *in* Γ.

Condition (A1) is often satisfied for two typical examples of $\boldsymbol{t}(\boldsymbol{X})$:

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

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

3 Constrained Shrinkage Estimators under Normality

In this section, we restrict the estimators to the constrained estimators (2.2) or the class Γ_1, and investigate the admissibility within the class (2.2) when the underlying distribution is normal.

3.1 A canonical form of the problem

We hereafter assume that $\boldsymbol{X}$ has a multivariate normal distribution

$$\boldsymbol{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{3.1}$$

for a known $k \times k$ matrix $\boldsymbol{\Sigma}$. Under this normality assumption, we study admissibility and minimaxity in the simultaneous estimation of $\boldsymbol{\mu}$ subject to the benchmark constraint $\boldsymbol{W}'\widehat{\boldsymbol{\mu}} = \boldsymbol{t}(\boldsymbol{X})$.

Assuming the uniform prior for $\boldsymbol{\mu}$, namely, $\pi(\boldsymbol{\mu}) = 1$, the benchmarked generalized Bayes estimator within the class Γ_0 is

$$\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t}) = \widehat{\boldsymbol{\mu}}^C(\boldsymbol{X}, \boldsymbol{t}) = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{W}})\boldsymbol{X} + \boldsymbol{Q}^{-1}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{t}(\boldsymbol{X}), \tag{3.2}$$

since the generalized Bayes estimator of $\boldsymbol{\mu}$ in Γ is just $\boldsymbol{X}$. Propositions 2.2 and 2.3 imply that the decision-theoretic properties of the benchmarked generalized Bayes estimator come from those of $\boldsymbol{X}$ in terms of the risk $R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}})$. Let $\boldsymbol{H}$ be a $k \times k$ orthogonal matrix such that

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

Let $\boldsymbol{H}' = (\boldsymbol{H}_1', \boldsymbol{H}_2')$ for the $k \times (k-m)$ matrix $\boldsymbol{H}_1$. Also, let $\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{Q}^{1/2}\boldsymbol{X}$, $\boldsymbol{\xi} = \boldsymbol{H}\boldsymbol{Q}^{1/2}\boldsymbol{\mu}$,

$$\boldsymbol{Y}_i = \boldsymbol{H}_i\boldsymbol{Q}^{1/2}\boldsymbol{X} \quad \text{and} \quad \boldsymbol{\xi}_i = \boldsymbol{H}_i\boldsymbol{Q}^{1/2}\boldsymbol{\mu}$$

for $i = 1, 2$. Then, $\boldsymbol{\mu}$ is expressed as

$$\boldsymbol{\mu} = \boldsymbol{Q}^{-1/2}\boldsymbol{H}_1'\boldsymbol{\xi}_1 + \boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'\boldsymbol{\xi}_2, \tag{3.3}$$

and the distributions of $\boldsymbol{Y}$ is written as

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim \mathcal{N}_k\left(\begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{pmatrix}\right), \tag{3.4}$$

where

$$\boldsymbol{V}_{ij} = \boldsymbol{H}_i\boldsymbol{Q}^{1/2}\boldsymbol{\Sigma}\boldsymbol{Q}^{1/2}\boldsymbol{H}_j'$$

for $i, j = 1, 2$. The problem of finding a benchmarked generalized Bayes estimator may be expressed as the minimization of $E[(\widehat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_1)'(\widehat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_1)|\boldsymbol{Y}]$ subject to $\boldsymbol{W}'\boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'\widehat{\boldsymbol{\xi}}_2 = \boldsymbol{t}(\boldsymbol{X})$, since $\boldsymbol{W}'\boldsymbol{Q}^{-1/2}\boldsymbol{H}_1' = \boldsymbol{0}$. Note that

$$\boldsymbol{W}'\boldsymbol{\mu} = \boldsymbol{W}'\boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'\boldsymbol{\xi}_2.$$

The estimators $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t})$ and $\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t})$ can be rewritten as

$$\begin{aligned} \widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t}) =& \boldsymbol{Q}^{-1/2}\boldsymbol{H}_1'\widehat{\boldsymbol{\xi}}_1 + \boldsymbol{Q}^{-1}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{t}(\boldsymbol{X}) \equiv \widehat{\boldsymbol{\mu}}^{C*}(\widehat{\boldsymbol{\xi}}_1, \boldsymbol{t}), \\ \widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t}) =& \boldsymbol{Q}^{-1/2}\boldsymbol{H}_1'\boldsymbol{Y}_1 + \boldsymbol{Q}^{-1}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{t}(\boldsymbol{X}) \equiv \widehat{\boldsymbol{\mu}}^{C*}(\boldsymbol{Y}_1, \boldsymbol{t}). \end{aligned} \tag{3.5}$$

for $\widehat{\boldsymbol{\xi}}_1 = \boldsymbol{H}_1\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{\mu}}$, and the risks $R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}})$ and $R_1(\boldsymbol{\mu}, \boldsymbol{X})$ can be expressed as

$$\begin{aligned} R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}}) =& E[(\widehat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_1)'(\widehat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_1)], \\ R_1(\boldsymbol{\mu}, \boldsymbol{X}) =& E[(\boldsymbol{Y}_1 - \boldsymbol{\xi}_1)'(\boldsymbol{Y}_1 - \boldsymbol{\xi}_1)]. \end{aligned} \tag{3.6}$$

3.2 Admissibility and inadmissibility results

To investigate admissibility of the constrained estimator $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t})$ within the constrained class Γ_1, the following decomposition from (3.4) is helpful:

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 - \boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1}\boldsymbol{Y}_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 - \boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1}\boldsymbol{\xi}_1 \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_{22.1} \end{pmatrix}\right). \tag{3.7}$$

Let $\boldsymbol{\xi}_3 = \boldsymbol{\xi}_2 - \boldsymbol{V}_{21}\boldsymbol{V}_{11}^{-1}\boldsymbol{\xi}_1$. As long as a proper prior is taken for $\boldsymbol{\xi}_3$, the admissibility of $\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t})$ depends on that of $\widehat{\boldsymbol{\xi}}_1 = \boldsymbol{H}_1\boldsymbol{Q}^{1/2}\widehat{\boldsymbol{\mu}}$ in terms of the risk $R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}})$.

Proposition 3.1 *If $\widehat{\boldsymbol{\xi}}_1$ is admissible in terms of $R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}})$, then $\widehat{\boldsymbol{\mu}}$ is admissible within the class Γ_1. In particular, if $\widehat{\boldsymbol{\xi}}_1$ is the Bayes estimator for a proper prior on $\boldsymbol{\xi}_1$, then $\widehat{\boldsymbol{\mu}}$ is admissible within Γ_1. If $\widehat{\boldsymbol{\xi}}_1$ is inadmissible in terms of the risk $R_1(\boldsymbol{\mu}, \widehat{\boldsymbol{\mu}})$, then $\widehat{\boldsymbol{\mu}}$ is inadmissible.*

This proposition along with well known results of James and Stein (1961), Brown (1971) and others implies the following result on the admissibility of $\boldsymbol{Y}_1$.

Proposition 3.2 *The benchmarked generalized Bayes estimator* $\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t})$ *under the uniform prior has the following decision-theoretic properties:*

(1) $\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t})$ *is minimax within* Γ_1 *under the condition* (2.5) *or* (A1).

(2) $\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t})$ *is admissible within* Γ_1 *when* $k-m$ *is one or two.*

(3) $\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t})$ *is inadmissible within* Γ_1 *when* $k-m \geq 3$.

The part (2) and (3) of Proposition 3.2 follows from James and Stein (1961), Brown (1971) and others.

When $k-m \geq 3$, the Stein effect leads to shrinkage estimators improving on the benchmarked generalized Bayes estimator $\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t})$ under the uniform prior. ¿From (3.7), we can derive improved estimators by shrinking $\boldsymbol{Y}_1$. For example, Berger (1976) proposed the shrinkage estimator

$$\widehat{\boldsymbol{\xi}}_1^B = \boldsymbol{Y}_1 - c(\boldsymbol{Y}_1'\boldsymbol{V}_{11}^{-2}\boldsymbol{Y}_1)^{-1}\boldsymbol{V}_{11}^{-1}\boldsymbol{Y}_1 \equiv \boldsymbol{g}(\boldsymbol{Y}_1, \boldsymbol{V}_{11}), \tag{3.8}$$

which dominates $\boldsymbol{Y}_1$ for $0 < c \leq 2(k-m-2)$. Shinozaki (1974) and Bock (1975) also gave other forms of improved estimators.

An admissible estimator for $k-m \geq 3$ can be derived from the result of Berger (1976). Assume the following prior distribution for $\boldsymbol{\xi}_1$:

$$\begin{aligned} \boldsymbol{\xi}_1|\gamma \sim& \mathcal{N}_{k-m}(\boldsymbol{0}, \{(\boldsymbol{V}_{11}-\gamma\boldsymbol{I})^{-1} - \boldsymbol{V}_{11}^{-1}\}^{-1}), \\ \gamma \sim& \gamma^{a/2-2}, \quad 0 < \gamma < \mathbf{ch}_{min}(\boldsymbol{V}_{11}), \end{aligned} \tag{3.9}$$

where a is a constant and $\mathbf{ch}_{min}(\boldsymbol{V}_{11})$ denotes the smallest eigenvalue of $\boldsymbol{V}_{11}$. This is an extension of the prior suggested in Strawderman (1971). For $\boldsymbol{\xi}_3$, it is possible to assume any proper prior distribution. Then, the generalized Bayes estimator of $\boldsymbol{\xi}_1$ is

$$\widehat{\boldsymbol{\xi}}_1^{GB}(\boldsymbol{Y}_1, \boldsymbol{V}_{11}) = \boldsymbol{Y}_1 - \frac{1}{\boldsymbol{Y}_1'\boldsymbol{V}_{11}^{-2}\boldsymbol{Y}_1}\psi_a^{SW}(\boldsymbol{Y}_1'\boldsymbol{V}_{11}^{-2}\boldsymbol{Y}_1)\boldsymbol{V}_{11}^{-1}\boldsymbol{Y}_1,$$

where

$$\psi_{a,k-m}^{SW}(w) = \int_0^w y^{(a+k-m)/2-1}e^{-y/2}\mathrm{d}y \Big/ \int_0^w y^{(a+k-m)/2-2}e^{-y/2}\mathrm{d}y. \tag{3.10}$$

Since the admissibility of $\widehat{\boldsymbol{\xi}}_1^{GB}$ was shown in Berger (1976), we have the following proposition.

Proposition 3.3 *The constrained generalized Bayes estimator*

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

is admissible and dominates $\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t})$ *within the class* Γ_1 *when* $k-m \geq 3$ *and* $0 \leq a \leq k-m-2$.

We can also provide other improved estimators through the following alternative decomposition:

$$\begin{pmatrix} \boldsymbol{Y}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\xi}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{\xi}_2 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_{11.2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_{22} \end{pmatrix} \right). \tag{3.11}$$

Note that $\boldsymbol{Y}_1 = (\boldsymbol{Y}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2) + \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2$ and that $\boldsymbol{Y}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2$ is independent of $\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2$. For $k - m \geq 3$, another improved shrinkage estimator given by

$$\widehat{\boldsymbol{\xi}}_1^{(2)} = \boldsymbol{g}(\boldsymbol{Y}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2, \boldsymbol{V}_{11.2}) + \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2,$$

where $\boldsymbol{g}(\cdot,\cdot)$ is defined in (3.8), since the risk function in terms of the loss $\|\widehat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_1\|^2 = (\widehat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_1)'(\widehat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_1)$ is

$$\begin{aligned} E[\|\widehat{\boldsymbol{\xi}}_1^{(2)} - \boldsymbol{\xi}_1\|^2] =& E[\|\boldsymbol{g}(\boldsymbol{Y}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2, \boldsymbol{V}_{11.2}) - (\boldsymbol{\xi}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{\xi}_2)\|^2] \\ &+ E\|\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{\xi}_2\|^2]. \end{aligned}$$

When $m - k \geq 3$ and $m \geq 3$, we also consider the shrinkage estimators

$$\begin{aligned} \widehat{\boldsymbol{\xi}}_1^{(3)} =& (\boldsymbol{Y}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2) + \boldsymbol{g}(\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2, \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21}), \\ \widehat{\boldsymbol{\xi}}_1^{(4)} =& \varepsilon\widehat{\boldsymbol{\xi}}_1^{(2)} + (1-\varepsilon)\widehat{\boldsymbol{\xi}}_1^{(3)}, \end{aligned}$$

where ε is a constant satifying $0 < \varepsilon < 1$.

Proposition 3.4 *The benchmarked generalized Bayes estimator $\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t})$ under the uniform prior can be improved by shrinkage estimators as follows:*

(1) *When $k - m \geq 3$, the shrinkage estimators $\widehat{\boldsymbol{\xi}}_1^{(1)} = \boldsymbol{g}(\boldsymbol{Y}_1, \boldsymbol{V}_{11})$ and $\widehat{\boldsymbol{\xi}}_1^{(2)} = \boldsymbol{g}(\boldsymbol{Y}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2, \boldsymbol{V}_{11.2}) + \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2$ dominate $\boldsymbol{Y}_1$ relative to the loss $\|\widehat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_1\|^2$.*

(2) *When $k - m \geq 3$ and $m \geq 3$, the shrinkage estimators $\widehat{\boldsymbol{\xi}}_1^{(3)} = (\boldsymbol{Y}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2) + \boldsymbol{g}(\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2, \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21})$ and $\widehat{\boldsymbol{\xi}}_1^{(4)} = \varepsilon\widehat{\boldsymbol{\xi}}_1^{(2)} + (1-\varepsilon)\widehat{\boldsymbol{\xi}}_1^{(3)}$ for $0 < \varepsilon < 1$ dominate $\boldsymbol{Y}_1$ relative to the loss $\|\widehat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_1\|^2$.*

4 Unconstrained generalized Bayes and minimax estimators satisfying the constraints

In the previous sections, 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 sub-section we find prior distributions such that the resulting unconstrained generalized Bayes estimators satisfy the constraints automatically and hence are also, therefore, the benchmarked generalized Bayes estimators.

4.1 An example

We begin with an illustrative example. Assume that random variables $X_1, \ldots, X_k$ are mutually independently distributed as $X_i \sim \mathcal{N}(\mu_i, 1)$. Let $\boldsymbol{X} = (X_1, \ldots, X_k)'$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$. Assume that an estimator $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_k)'$ satisfies the constraint $\sum_{i=1}^k \widehat{\mu}_i = \sum_{i=1}^k X_i$, namely, $\boldsymbol{j}'\widehat{\boldsymbol{\mu}} = \boldsymbol{j}'\boldsymbol{X}$ for $\boldsymbol{j} = (1, \ldots, 1)' \in \boldsymbol{R}^k$. Then, we consider estimation of $\boldsymbol{\mu}$ relative to the loss $\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2$. In the setting of Section 3.1, the model is $\boldsymbol{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{I}_k)$, where $\boldsymbol{Q} = \boldsymbol{\Sigma} = \boldsymbol{I}_k$.

Note that $\boldsymbol{\mu}$ may be expressed as

$$\begin{aligned}\boldsymbol{\mu} =&(\boldsymbol{I} - k^{-1}\boldsymbol{j}\boldsymbol{j}')\boldsymbol{\mu} + (\boldsymbol{j}'\boldsymbol{\mu}/k)\boldsymbol{j} \\ =&\boldsymbol{H}_1'\boldsymbol{\xi}_1 + \boldsymbol{h}_2'\xi_2,\end{aligned}$$

where $\boldsymbol{\xi}_1 = \boldsymbol{H}_1\boldsymbol{\mu}$, $\xi_2 = \boldsymbol{h}_2\boldsymbol{\mu}$, $\boldsymbol{H}_1\boldsymbol{j} = \boldsymbol{0}$ and $\boldsymbol{h}_2\boldsymbol{j} = \sqrt{k}$ for a $k \times k$ orthogonal matrix $\boldsymbol{H}' = (\boldsymbol{H}_1', \boldsymbol{h}_2')$, $\boldsymbol{H}_1$ being $(k-1) \times k$. Let $\boldsymbol{Y}_1 = \boldsymbol{H}_1\boldsymbol{X}$ and $y_2 = \boldsymbol{h}_2\boldsymbol{X}$. Then,

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ y_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\xi}_1 \\ \xi_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{I}_{k-1} & \boldsymbol{0} \\ \boldsymbol{0}' & 1 \end{pmatrix} \right).$$

Assume the following prior distribution:

$$\begin{aligned}\boldsymbol{\xi}_1|\gamma \sim&\mathcal{N}_{k-1}(\boldsymbol{0}, \gamma/(1-\gamma)\boldsymbol{I}_{k-1}), \\ \gamma \sim&\gamma^{a/2-2}, \quad 0 < \gamma < 1, \\ \xi_2 \sim&1.\end{aligned}$$

This type of prior distribution was suggested in Strawderman (1971). Then, the generalized Bayes estimator $\widehat{\boldsymbol{\xi}}_1^{GB}$ and $\hat{\xi}_2^{GB}$ of $\boldsymbol{\xi}_1$ and ξ_2 are given by

$$\begin{aligned}\widehat{\boldsymbol{\xi}}_1^{GB} =&\boldsymbol{Y}_1 - \|\boldsymbol{Y}_1\|^{-2}\psi_{a,k-1}^{SW}(\|\boldsymbol{Y}_1\|^2)\boldsymbol{Y}_1, \\ \hat{\xi}_2^{GB} =&y_2,\end{aligned}$$

where $\psi_{a,k-1}^{SW}(\cdot)$ is defined in (3.10).

Note that

$$\begin{aligned}\boldsymbol{H}_1'\boldsymbol{Y}_1 =&(\boldsymbol{I} - k^{-1}\boldsymbol{j}\boldsymbol{j}')\boldsymbol{X} = \boldsymbol{X} - \overline{X}\boldsymbol{j}, \\ \boldsymbol{h}_2'y_2 =&\overline{X}\boldsymbol{j}.\end{aligned}$$

We thus get the generalized Bayes estimator

$$\begin{aligned}\widehat{\boldsymbol{\mu}}^{GB} =&\boldsymbol{H}_1'\widehat{\boldsymbol{\xi}}_1^{GB} + \boldsymbol{h}_2'\hat{\xi}_2^{GB} \\ =&(\boldsymbol{X} - \overline{X}\boldsymbol{j}) - \frac{1}{\sum_{i=1}^k (X_i - \overline{X})^2}\psi_{a,k-1}^{SW}(\sum_{i=1}^k (X_i - \overline{X})^2)(\boldsymbol{X} - \overline{X}\boldsymbol{j}) + \overline{X}\boldsymbol{j} \\ =&\boldsymbol{X} - \frac{1}{\sum_{i=1}^k (X_i - \overline{X})^2}\psi_{a,k-1}^{SW}(\sum_{i=1}^k (X_i - \overline{X})^2)(\boldsymbol{X} - \overline{X}\boldsymbol{j}).\end{aligned}$$

This estimator shrinks X_i toward $\overline{X}$, and is often called a Lindley type estimator (Lindley, 1962).

It is clear that this estimator satisfies the constraint, namely,

$$\boldsymbol{j}'\widehat{\boldsymbol{\mu}}^{GB} = \boldsymbol{j}'\boldsymbol{X}.$$

Hence, although not derived as a constrained Bayes estimator, it satisfies the constraint automatically as shown above. The key to this phenomenon is that the prior distribution of ξ_2 is uniform.

4.2 Which priors automatically produce the constrained Bayes estimator?

In the previous subsection, we gave an example of unconstrained Bayes estimator which automatically satisfies the constraint. Now we give a general condition on the prior distribution such that the resulting generalized Bayes estimator possesses such a property. Suppose that the constraint is $\boldsymbol{W}'\widehat{\boldsymbol{\mu}} = \boldsymbol{t}(\boldsymbol{X})$. Note that from (3.3), $\boldsymbol{\mu}$ may be expressed as

$$\boldsymbol{\mu} = \boldsymbol{Q}^{-1/2}\boldsymbol{H}_1'\boldsymbol{\xi}_1 + \boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'\boldsymbol{\xi}_2,$$

so that the (unconstrained) Bayes or generalized Bayes estimator of $\boldsymbol{\mu}$ is decomposed as

$$\widehat{\boldsymbol{\mu}}^B = \boldsymbol{Q}^{-1/2}\boldsymbol{H}_1'E[\boldsymbol{\xi}_1|\boldsymbol{X}] + \boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'E[\boldsymbol{\xi}_2|\boldsymbol{X}],$$

where $E[\cdot|\boldsymbol{X}]$ denotes the posterior expectation. On the other hand, from (3.5), the constrained Bayes estimator must have the form

$$\widehat{\boldsymbol{\mu}}^C(\widehat{\boldsymbol{\mu}}, \boldsymbol{t}) = \boldsymbol{Q}^{-1/2}\boldsymbol{H}_1'\widehat{\boldsymbol{\xi}}_1 + \boldsymbol{Q}^{-1}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{t}(\boldsymbol{X}).$$

Hence, the unconstrained Bayes estimator $\widehat{\boldsymbol{\mu}}^B$ satisfies the constraint if the prior distribution satisfies the equation

$$\boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'E[\boldsymbol{\xi}_2|\boldsymbol{X}] = \boldsymbol{Q}^{-1}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{t}(\boldsymbol{X}),$$

or

$$E[\boldsymbol{\xi}_2|\boldsymbol{X}] = \boldsymbol{H}_2\boldsymbol{Q}^{-1/2}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{t}(\boldsymbol{X}). \tag{4.1}$$

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

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

so that the condition (4.1) may be simplified as

$$E[\boldsymbol{\xi}_2|\boldsymbol{X}] = \boldsymbol{H}_2\boldsymbol{s}(\boldsymbol{X}). \tag{4.2}$$

Proposition 4.1 *The unconstrained generalized Bayes estimators satisfy the constraint* $\boldsymbol{W}'\widehat{\boldsymbol{\mu}} = \boldsymbol{t}(\boldsymbol{X})$ *if the posterior expectation* $E[\boldsymbol{\xi}_2|\boldsymbol{X}]$ *satisfies the equation* (4.1) *or* (4.2).

When $\boldsymbol{t}(\boldsymbol{X}) = \boldsymbol{W}'\boldsymbol{X}$, the condition (4.2) is $E[\boldsymbol{\xi}_2|\boldsymbol{X}] = \boldsymbol{H}_2\boldsymbol{Q}^{1/2}\boldsymbol{X} = \boldsymbol{Y}_2$, for which it suffices that we assume the uniform prior for $\boldsymbol{\xi}_2$. When $\boldsymbol{t}(\boldsymbol{X}) = \boldsymbol{t}_0$, a constant, the condition (4.1) is $E[\boldsymbol{\xi}_2|\boldsymbol{X}] = \boldsymbol{H}_2\boldsymbol{Q}^{-1/2}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{t}_0$, which suggests that $\boldsymbol{\xi}_2$ should take a point mass at $\boldsymbol{\xi}_2 = \boldsymbol{H}_2\boldsymbol{Q}^{-1/2}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{t}_0$. Since $\boldsymbol{W}'\boldsymbol{Q}^{-1/2}\boldsymbol{H}_1' = \boldsymbol{0}$, this restriction is rewritten as

$$\begin{aligned}\boldsymbol{W}'\boldsymbol{\mu} =&\boldsymbol{W}'\boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'\boldsymbol{\xi}_2 = \boldsymbol{W}'\boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'\boldsymbol{H}_2\boldsymbol{Q}^{-1/2}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{t}_0\\ =&\boldsymbol{W}'\boldsymbol{Q}^{-1/2}(\boldsymbol{H}_1'\boldsymbol{H}_1 + \boldsymbol{H}_2'\boldsymbol{H}_2)\boldsymbol{Q}^{-1/2}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{Q}^{-1}\boldsymbol{W})^{-1}\boldsymbol{t}_0 = \boldsymbol{t}_0.\end{aligned}$$

These two cases are explained in the following subsections.

4.3 Case of $\boldsymbol{t}(\boldsymbol{X}) = \boldsymbol{W}'\boldsymbol{X}$

Consider the decomposition (3.11) and put $\boldsymbol{\xi}_4 = \boldsymbol{\xi}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{\xi}_2$. Then,

$$\begin{aligned}\boldsymbol{\mu} =&\boldsymbol{Q}^{-1/2}\boldsymbol{H}_1'\boldsymbol{\xi}_1 + \boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'\boldsymbol{\xi}_2\\ =&\boldsymbol{Q}^{-1/2}\boldsymbol{H}_1'\boldsymbol{\xi}_4 + \boldsymbol{Q}^{-1/2}(\boldsymbol{H}_2' + \boldsymbol{H}_1'\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1})\boldsymbol{\xi}_2.\end{aligned}$$

Assume the following prior distribution for $(\boldsymbol{\xi}_4, \boldsymbol{\xi}_2)$:

$$\begin{aligned}\boldsymbol{\xi}_4|\gamma \sim&\mathcal{N}_{k-m}(\boldsymbol{0}, \{(\boldsymbol{V}_{11.2} - \gamma\boldsymbol{I})^{-1} - \boldsymbol{V}_{11.2}^{-1}\}^{-1}),\\ \gamma \sim&\gamma^{a/2-2}, \quad 0 < \gamma < \mathbf{ch}_{min}(\boldsymbol{V}_{11.2}),\\ \boldsymbol{\xi}_2 \sim&1,\end{aligned} \tag{4.3}$$

where $\mathbf{ch}_{min}(\boldsymbol{V}_{11.2})$ denotes the smallest eigenvalue of $\boldsymbol{V}_{11.2}$. The prior for $\boldsymbol{\xi}_4$ was treated by Berger (1976) (and was already utilized in section (3.2)) and the prior for $\boldsymbol{\xi}_2$ is uniform. The resulting generalized Bayes estimator is

$$\begin{aligned}\widehat{\boldsymbol{\mu}}^{GB1} =&\boldsymbol{Q}^{-1/2}\boldsymbol{H}_1'\widehat{\boldsymbol{\xi}}_4^{GB}(\boldsymbol{Y}_4) + \boldsymbol{Q}^{-1/2}(\boldsymbol{H}_2' + \boldsymbol{H}_1'\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1})\boldsymbol{Y}_2\\ =&\boldsymbol{Q}^{-1/2}\boldsymbol{H}_1'\{\widehat{\boldsymbol{\xi}}_4^{GB}(\boldsymbol{Y}_4) + \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2\} + \boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'\boldsymbol{Y}_2,\end{aligned} \tag{4.4}$$

where $\boldsymbol{Y}_4 = \boldsymbol{Y}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2$ and

$$\widehat{\boldsymbol{\xi}}_4^{GB}(\boldsymbol{Y}_4) = \boldsymbol{Y}_4 - \frac{1}{\boldsymbol{Y}_4'\boldsymbol{V}_{11.2}^{-2}\boldsymbol{Y}_4}\psi_a^{SW}(\boldsymbol{Y}_4'\boldsymbol{V}_{11.2}^{-2}\boldsymbol{Y}_4)\boldsymbol{V}_{11.2}^{-1}\boldsymbol{Y}_4,$$

for $\psi_{a,k-m}^{SW}(w)$ given in (3.10). Note that

$$\begin{aligned}E[\|\widehat{\boldsymbol{\mu}}^{GB1} - \boldsymbol{\mu}\|^2] =&E[(\widehat{\boldsymbol{\xi}}_4^{GB} + \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2 - \boldsymbol{\xi}_4)'\boldsymbol{H}_1\boldsymbol{Q}^{-1}\boldsymbol{H}_1'(\widehat{\boldsymbol{\xi}}_4^{GB} + \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2 - \boldsymbol{\xi}_4)]\\ &+ E[(\boldsymbol{Y}_2 - \boldsymbol{\xi}_2)'\boldsymbol{H}_2\boldsymbol{Q}^{-1}\boldsymbol{H}_2'(\boldsymbol{Y}_2 - \boldsymbol{\xi}_2)]\\ =&E[(\widehat{\boldsymbol{\xi}}_4^{GB} - \boldsymbol{\xi}_1)'\boldsymbol{H}_1\boldsymbol{Q}^{-1}\boldsymbol{H}_1'(\widehat{\boldsymbol{\xi}}_4^{GB} - \boldsymbol{\xi}_1)]\\ &+ E[(\boldsymbol{Y}_2 - \boldsymbol{\xi}_2)'\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21}\boldsymbol{H}_1\boldsymbol{Q}^{-1}\boldsymbol{H}_1'\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}(\boldsymbol{Y}_2 - \boldsymbol{\xi}_2)]\\ &+ E[(\boldsymbol{Y}_2 - \boldsymbol{\xi}_2)'\boldsymbol{H}_2\boldsymbol{Q}^{-1}\boldsymbol{H}_2'(\boldsymbol{Y}_2 - \boldsymbol{\xi}_2)].\end{aligned}$$

Note also (as in section (3.2)) that, provided $0 \le a \le k-m-2$, $\widehat{\boldsymbol{\xi}}_4^{GB}$ is admissible in terms of the risk $E[(\widehat{\boldsymbol{\xi}}_4^{GB} - \boldsymbol{\xi}_1)'\boldsymbol{H}_1\boldsymbol{Q}^{-1}\boldsymbol{H}_1'(\widehat{\boldsymbol{\xi}}_4^{GB} - \boldsymbol{\xi}_1)]$. When $m \ge 3$, however, $\widehat{\boldsymbol{\mu}}^{GB1}$ is not admissible in the unconstrained problem. Combining Propositions 2.3 and 3.4 and the result of Berger (1976), we get the following proposition.

Proposition 4.2 *The generalized Bayes estimator* $\widehat{\boldsymbol{\mu}}^{GB1}$ *satisfies the constraint, namely* $\widehat{\boldsymbol{\mu}}^{GB1} \in \Gamma_0$. *If* $k-m \ge 3$ *and* $0 \le a \le k-m-2$, *then* $\widehat{\boldsymbol{\mu}}^{GB1}$ *is admissible and minimax within the constrained class* Γ_1. *When* $m \ge 3$, *it is not admissible in the unconstrained problem.*

4.4 Case of $\boldsymbol{t}(\boldsymbol{X}) = \boldsymbol{t}_0$ or $\boldsymbol{W}'\boldsymbol{\mu} = \boldsymbol{t}_0$

Since $\boldsymbol{W}'\boldsymbol{\mu} = \boldsymbol{W}'\boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'\boldsymbol{\xi}_2 = \boldsymbol{t}_0$, a constant, and $\boldsymbol{W}'\boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'$ is non-singular, we can define $\boldsymbol{\xi}_0$ by $\boldsymbol{\xi}_0 = (\boldsymbol{W}'\boldsymbol{Q}^{-1/2}\boldsymbol{H}_2')^{-1}\boldsymbol{t}_0$. In the decomposition (3.11), we assume the following prior distribution:

$$\begin{aligned}\boldsymbol{\xi}_1|\gamma \sim& \mathcal{N}_{k-m}(\boldsymbol{0}, \{(\boldsymbol{V}_{11.2} - \gamma\boldsymbol{I})^{-1} - \boldsymbol{V}_{11.2}^{-1}\}^{-1}),\\ \gamma \sim& \gamma^{a/2-2}, \quad 0 < \gamma < \mathbf{ch}_{min}(\boldsymbol{V}_{11.2}),\\ \boldsymbol{\xi}_2 =& \boldsymbol{\xi}_0 \quad \text{with probability one.}\end{aligned} \tag{4.5}$$

Let $\boldsymbol{Y}_5 = \boldsymbol{Y}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}(\boldsymbol{Y}_2 - \boldsymbol{\xi}_0)$. Since the decomposition (3.11) under the constraint $\boldsymbol{\xi}_2 = \boldsymbol{\xi}_0$ is expressed as

$$\begin{pmatrix} \boldsymbol{Y}_5 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_{11.2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_{22} \end{pmatrix}\right), \tag{4.6}$$

the generalized Bayes estimator is given by

$$\widehat{\boldsymbol{\mu}}^{GB2} = \boldsymbol{Q}^{-1/2}\boldsymbol{H}_1'\widehat{\boldsymbol{\xi}}_1^{GB}(\boldsymbol{Y}_5) + \boldsymbol{Q}^{-1/2}\boldsymbol{H}_2'\boldsymbol{\xi}_0,$$

where $\widehat{\boldsymbol{\xi}}_1^{GB}(\boldsymbol{Y}_5)$ has the same form as $\widehat{\boldsymbol{\xi}}_4^{GB}(\boldsymbol{Y}_4)$ except replacing $\boldsymbol{Y}_4$ with $\boldsymbol{Y}_5$, namely, $\widehat{\boldsymbol{\xi}}_1^{GB}(\boldsymbol{Y}_5) = \widehat{\boldsymbol{\xi}}_4^{GB}(\boldsymbol{Y}_5)$. Clearly, $\widehat{\boldsymbol{\mu}}^{GB2}$ satisfies the constraint $\boldsymbol{W}'\widehat{\boldsymbol{\mu}}^{GB2} = \boldsymbol{t}_0$. Also the risk is expressed as

$$\begin{aligned}E[\|\widehat{\boldsymbol{\mu}}^{GB2} - \boldsymbol{\mu}\|^2] =& E[(\widehat{\boldsymbol{\xi}}_1^{GB} - \boldsymbol{\xi}_1)'\boldsymbol{H}_1\boldsymbol{Q}^{-1}\boldsymbol{H}_1'(\widehat{\boldsymbol{\xi}}_1^{GB} - \boldsymbol{\xi}_1)]\\ &+ (\boldsymbol{\xi}_0 - \boldsymbol{\xi}_2)'\boldsymbol{H}_2\boldsymbol{Q}^{-1}\boldsymbol{H}_2'(\boldsymbol{\xi}_0 - \boldsymbol{\xi}_2),\end{aligned}$$

so that the admissibility of $\widehat{\boldsymbol{\mu}}^{GB2}$ is inherited from that of $\widehat{\boldsymbol{\xi}}_1^{GB}$.

Proposition 4.3 *The generalized Bayes estimator* $\widehat{\boldsymbol{\mu}}^{GB2}$ *satisfies the constraint, namely* $\widehat{\boldsymbol{\mu}}^{GB2} \in \Gamma_0$. *If* $k-m \ge 3$ *and* $0 \le a \le k-m-2$, *then* $\widehat{\boldsymbol{\mu}}^{GB2}$ *is admissible and improves on* $\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t}_0)$ *within the whole class* Γ (*and also* Γ_1).

Alternatively, we may consider the constrained parameter space problem where $\boldsymbol{\xi}_2 = \boldsymbol{\xi}_0$. In this case the above development gives the same generalized Bayes estimator, but the risk function simplifies to

$$E[\|\widehat{\boldsymbol{\mu}}^{GB2} - \boldsymbol{\mu}\|^2] = E[(\widehat{\boldsymbol{\xi}}_1^{GB} - \boldsymbol{\xi}_1)'\boldsymbol{H}_1\boldsymbol{Q}^{-1}\boldsymbol{H}_1'(\widehat{\boldsymbol{\xi}}_1^{GB} - \boldsymbol{\xi}_1)]$$

Additionally, the above proposition remains true but in addition both estimators are minimax.

5 Benchmarking in the Fay-Herriot Model

As mentioned in the introduction and as explained in Datta *et al.* (2011) 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 its minimaxity. The Fay-Herriot model has heteroscedastic variances and covariates as regressors, so that establishing minimaxity of the constrained empirical Bayes estimator, while somewhat challenging, seems to be potentially useful.

5.1 Constrained empirical Bayes estimator

The Fay-Herriot model which we study is described as

$$\begin{aligned} \boldsymbol{X}|\boldsymbol{\mu} &\sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{D}), \quad \boldsymbol{D} = \mathrm{diag}\,(d_1, \ldots, d_k), \\ \boldsymbol{\mu} &\sim \mathcal{N}_k(\boldsymbol{Z\beta}, \lambda\boldsymbol{I}), \end{aligned} \tag{5.1}$$

where $\boldsymbol{Z}$ 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 unknown scalar. Suppose that $d_1 \geq \cdots \geq d_k$ without any loss of generality. Consider estimation of $\boldsymbol{\mu}$ relative to the loss $\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 = (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})'(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})$. Note that

$$\begin{aligned} &(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{D}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) + (\boldsymbol{\mu} - \boldsymbol{Z\beta})'(\boldsymbol{\mu} - \boldsymbol{Z\beta})/\lambda \\ &\quad = (\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}^B)'(\lambda^{-1}\boldsymbol{I} + \boldsymbol{D}^{-1})(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}^B) + (\boldsymbol{X} - \boldsymbol{Z\beta})'(\boldsymbol{D} + \lambda\boldsymbol{I})^{-1}(\boldsymbol{X} - \boldsymbol{Z\beta}), \end{aligned}$$

where $\widehat{\boldsymbol{\mu}}^B$ is the Bayes estimator (under the assumption of known $\boldsymbol{\beta}$ and λ) given by

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

For estimation of λ, several estimators are known including the Prasad-Rao estimator given by Prasad and Rao (1990), the Fay-Herriot estimator suggested by Fay and Herriot (1979), the maximum likelihood estimator (MLE) and the restricted maximum likelihood estimator (REML). For the MLE and REML, see Searle, Casella and McCulloch (1992) and Kubokawa (2011) for example. Denoting an estimator of λ by $\hat{\lambda}$, we get the empirical Bayes estimator

$$\widehat{\boldsymbol{\mu}}^{EB}(\hat{\lambda}) = \boldsymbol{X} - \boldsymbol{D}(\boldsymbol{D} + \hat{\lambda}\boldsymbol{I})^{-1}(\boldsymbol{X} - \boldsymbol{Z}\widehat{\boldsymbol{\beta}}(\hat{\lambda})),$$

where

$$\widehat{\boldsymbol{\beta}}(\hat{\lambda}) = \{\boldsymbol{Z}'\boldsymbol{V}(\hat{\lambda})^{-1}\boldsymbol{Z}\}^{-1}\boldsymbol{Z}'\boldsymbol{V}(\hat{\lambda})^{-1}\boldsymbol{X},$$

for $\boldsymbol{V}(\lambda) = \boldsymbol{D} + \lambda \boldsymbol{I}$. Define $\boldsymbol{A}(\lambda)$ by

$$\boldsymbol{A}(\lambda) = \boldsymbol{V}(\lambda)^{-1} - \boldsymbol{V}(\lambda)^{-1}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{V}(\lambda)^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}(\lambda)^{-1}. \tag{5.2}$$

Then, the empirical Bayes estimator can be rewritten as

$$\widehat{\boldsymbol{\mu}}^{EB}(\hat{\lambda}) = \boldsymbol{X} - \boldsymbol{D}\boldsymbol{A}(\hat{\lambda})\boldsymbol{X}. \tag{5.3}$$

Now consider the benchmark constraint $\boldsymbol{W}'\widehat{\boldsymbol{\mu}} = \boldsymbol{t}(\boldsymbol{X})$ for a $k \times m$ matrix $\boldsymbol{W}$ and a function $\boldsymbol{t}(\boldsymbol{X})$. Then the constrained empirical Bayes estimator (CEB) based on $\widehat{\boldsymbol{\mu}}^{EB}(\hat{\lambda})$ (as constructed in 2.2) is given by

$$\begin{aligned}\widehat{\boldsymbol{\mu}}^{CEB}(\hat{\lambda}, \boldsymbol{t}) =& \widehat{\boldsymbol{\mu}}^{C}(\widehat{\boldsymbol{\mu}}^{EB}(\hat{\lambda}), \boldsymbol{t}) = \widehat{\boldsymbol{\mu}}^{EB}(\hat{\lambda}) + \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\{\boldsymbol{t}(\boldsymbol{X}) - \boldsymbol{W}'\widehat{\boldsymbol{\mu}}^{EB}(\hat{\lambda})\}\\ =&\{\boldsymbol{I} - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\}\widehat{\boldsymbol{\mu}}^{EB}(\hat{\lambda}) + \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{t}(\boldsymbol{X}). \end{aligned} \tag{5.4}$$

5.2 Conditions for improvement

We now derive a condition under which the constrained empirical Bayes estimator $\widehat{\boldsymbol{\mu}}^{CEB}(\hat{\lambda}, \boldsymbol{t})$ improves on the constrained uniform-prior generalized Bayes estimator $\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t}) = \{\boldsymbol{I} - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\}\boldsymbol{X} + \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{t}(\boldsymbol{X})$. The risk difference of the two estimators is written as

$$\begin{aligned}\Delta =& E[(\widehat{\boldsymbol{\mu}}^{CEB}(\hat{\lambda}, \boldsymbol{t}) - \boldsymbol{\mu})'(\widehat{\boldsymbol{\mu}}^{CEB}(\hat{\lambda}, \boldsymbol{t}) - \boldsymbol{\mu})] - E[(\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t}) - \boldsymbol{\mu})'(\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t}) - \boldsymbol{\mu})]\\ =& E[(\widehat{\boldsymbol{\mu}}^{EB}(\hat{\lambda}) - \boldsymbol{\mu})'\boldsymbol{Q}_{\boldsymbol{W}}(\widehat{\boldsymbol{\mu}}^{EB}(\hat{\lambda}) - \boldsymbol{\mu})] - E[(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{Q}_{\boldsymbol{W}}(\boldsymbol{X} - \boldsymbol{\mu})],\end{aligned}$$

where $\boldsymbol{Q}_{\boldsymbol{W}} = \boldsymbol{I} - \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'$. It is noted that $\boldsymbol{Q}_{\boldsymbol{W}}$ is of rank $k - m$ and that $E[E[(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{Q}_{\boldsymbol{W}}(\boldsymbol{X} - \boldsymbol{\mu})] = \mathrm{tr}\,[\boldsymbol{D}\boldsymbol{Q}_{\boldsymbol{W}}] = \mathrm{tr}\,[\boldsymbol{D}] - \mathrm{tr}\,[\boldsymbol{W}'\boldsymbol{D}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}]$.

Lemma 5.1 *The risk difference Δ is expressed as $\Delta = E[\widehat{\Delta}]$, where*

$$\begin{aligned}\widehat{\Delta}(\hat{\lambda}) =& -2\mathrm{tr}\,[\boldsymbol{D}\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{A}(\hat{\lambda})] + 2\boldsymbol{X}'\boldsymbol{A}^2(\hat{\lambda})\boldsymbol{D}\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}(\boldsymbol{\nabla}\hat{\lambda})\\ &+ \boldsymbol{X}'\boldsymbol{A}(\hat{\lambda})\boldsymbol{D}\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{A}(\hat{\lambda})\boldsymbol{X},\end{aligned} \tag{5.5}$$

for $\boldsymbol{\nabla} = (\partial/\partial X_1, \ldots, \partial/\partial X_k)'$.

Proof. The risk difference is written as

$$\Delta = -2E[(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{A}(\hat{\lambda})\boldsymbol{X}] + E[\boldsymbol{X}'\boldsymbol{A}(\hat{\lambda})\boldsymbol{D}\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{A}(\hat{\lambda})\boldsymbol{X}].$$

Using the Stein identity given in Stein (1973, 81), we can rewrite the cross product term as

$$E[(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{A}(\hat{\lambda})\boldsymbol{X}] = E[\boldsymbol{\nabla}'\{\boldsymbol{D}\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{A}(\hat{\lambda})\boldsymbol{X}\}].$$

Let $\boldsymbol{G}(\hat{\lambda}) = (g_{ij}(\hat{\lambda})) = \boldsymbol{D}\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{A}(\hat{\lambda})$. Then,

$$\begin{aligned}\boldsymbol{\nabla}'\{\boldsymbol{G}(\hat{\lambda})\boldsymbol{X}\} =& \sum_{i,j} \frac{\partial}{\partial X_i}\{g_{ij}(\hat{\lambda})X_j\}\\ =& \sum_i g_{ii}(\hat{\lambda}) + \sum_{i,j} X_j \Big\{\frac{\mathrm{d}}{\mathrm{d}\lambda} g_{ij}(\lambda)\Big|_{\lambda=\hat{\lambda}}\Big\}\frac{\partial \hat{\lambda}}{\partial X_i}\\ =& \mathrm{tr}\,[\boldsymbol{D}\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{A}(\hat{\lambda})] + \boldsymbol{X}'\Big\{\frac{\mathrm{d}}{\mathrm{d}\lambda}\boldsymbol{A}(\lambda)\Big|_{\lambda=\hat{\lambda}}\Big\}\boldsymbol{D}\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}(\boldsymbol{\nabla}\hat{\lambda}),\end{aligned}$$

since $g_{ij}(\hat{\lambda})$ depends on $\boldsymbol{X}$ through $\hat{\lambda}$. Differentiating $\boldsymbol{A}(\lambda)$ with respect to λ for $\boldsymbol{A}(\lambda)$ given in (5.2), we can see that

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\boldsymbol{A}(\lambda) = -\boldsymbol{A}^2(\lambda), \tag{5.6}$$

which can be used to get the expression in (5.5). ■

To establish improvement of $\widehat{\boldsymbol{\mu}}^{CEB}(\hat{\lambda}, \boldsymbol{t})$ over $\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t})$, we need to find an estimator $\hat{\lambda}$ such that $\widehat{\Delta}(\hat{\lambda}) \leq 0$ for all $\boldsymbol{X}$. We here treat the estimator $\hat{\lambda}$ given by $\hat{\lambda} = \max\{\lambda_0, 0\}$ where λ_0 is the solution of the equation

$$\boldsymbol{X}'\boldsymbol{A}(\lambda_0)\boldsymbol{X} = c, \tag{5.7}$$

for a positive constant c. When $c = k-p$, this estimator was suggested by Fay and Herriot (1979). Differentiating $\boldsymbol{X}'\boldsymbol{A}(\hat{\lambda})\boldsymbol{X} = c$ with respect to $\boldsymbol{X}$ and using the implicit function theorem, we get the equation $2\boldsymbol{A}(\hat{\lambda})\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{A}^2(\hat{\lambda})\boldsymbol{X}\boldsymbol{\nabla}\hat{\lambda} = \boldsymbol{0}$ in the case of $\hat{\lambda} > 0$, or

$$\boldsymbol{\nabla}\hat{\lambda} = \frac{2}{\boldsymbol{X}'\boldsymbol{A}^2(\hat{\lambda})\boldsymbol{X}}\boldsymbol{A}(\hat{\lambda})\boldsymbol{X}I(\hat{\lambda} > 0).$$

Thus, $\widehat{\Delta}$ given in (5.5) is expressed as

$$\begin{aligned}\widehat{\Delta}(\hat{\lambda}) =& -2\mathrm{tr}\,[\boldsymbol{D}\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{A}(\hat{\lambda})] + 4\frac{\boldsymbol{X}'\boldsymbol{A}^2(\hat{\lambda})\boldsymbol{D}\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{A}(\hat{\lambda})\boldsymbol{X}}{\boldsymbol{X}'\boldsymbol{A}^2(\hat{\lambda})\boldsymbol{X}}I(\hat{\lambda} > 0)\\ &+ \boldsymbol{X}'\boldsymbol{A}(\hat{\lambda})\boldsymbol{D}\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{D}\boldsymbol{A}(\hat{\lambda})\boldsymbol{X},\end{aligned} \tag{5.8}$$

where $I(A)$ is the indicator function such that $I(A) = 1$ if A is true, and otherwise, $I(A) = 0$. Evaluating each term in (5.8), we get a condition for $\widehat{\Delta}(\hat{\lambda}) \leq 0$, which is given in the following proposition.

Proposition 5.1 *The constrained empirical Bayes estimator $\widehat{\boldsymbol{\mu}}^{CEB}(\hat{\lambda}, \boldsymbol{t})$ with $\hat{\lambda}$ given in* (5.7) *improves on $\widehat{\boldsymbol{\mu}}^{Cm}(\boldsymbol{t})$ if c satisfies the condition*

$$0 < c \leq 2\Big\{d_1^{-2}\mathrm{tr}\,[\boldsymbol{D}^2\boldsymbol{Q}_{\boldsymbol{W}}] - \mathrm{tr}\,[(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{Z}] - 2\Big\}. \tag{5.9}$$

If the constrant is given by $\boldsymbol{t}(\boldsymbol{X}) = \boldsymbol{W}'\boldsymbol{X}$, then the estimator $\widehat{\boldsymbol{\mu}}^{CEB}(\hat{\lambda}, \boldsymbol{t})$ is minimax under the condition (5.9)*.*

Since $\mathrm{tr}\,[(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{Z}] \leq p$, we get another sufficient condition

$$0 < c \leq 2\Big\{d_1^{-2}\mathrm{tr}\,[\boldsymbol{D}^2\boldsymbol{Q}_{\boldsymbol{W}}] - p - 2\Big\}, \tag{5.10}$$

which is given as $0 < c \leq 2\{k-m-p-2\}$ when $\boldsymbol{D} = \boldsymbol{I}$. Since the Fay-Herriot estimator corresponds to the case of $c = k - p$ in (5.7), the condition (5.9) for the Fay-Herriot estimator is

$$k - p \leq 2\Big\{d_1^{-2}\mathrm{tr}\,[\boldsymbol{D}^2\boldsymbol{Q}_{\boldsymbol{W}}] - \mathrm{tr}\,[(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{Z}] - 2\Big\},$$

so that the improvement depends on $\boldsymbol{D}$, $\boldsymbol{W}$ and $\boldsymbol{Z}$.

Proposition 5.1 is proved below by using the same arguments as in Shinozaki and Chang (1996) who treated slightly different empirical Bayes estimators and derived different conditions on their minimaxity in the case of $\boldsymbol{Q_W} = \boldsymbol{I}$. In the case of $\boldsymbol{Q_W} = \boldsymbol{I}$ and $\boldsymbol{Z} = \boldsymbol{0}$, the condition (5.10) is identical to that of Shinozaki and Chang (1993).

Proof of Proposition 5.1. We shall evaluate each term in (5.8) above. It is observed that

$$
\begin{aligned}
\boldsymbol{X}'\boldsymbol{A}(\hat\lambda)\boldsymbol{D}\boldsymbol{Q_W}\boldsymbol{D}\boldsymbol{A}(\hat\lambda)\boldsymbol{X} \leq & \boldsymbol{X}'\boldsymbol{A}(\hat\lambda)\boldsymbol{D}^2\boldsymbol{A}(\hat\lambda)\boldsymbol{X} \\
\leq & \boldsymbol{X}'\boldsymbol{A}(\hat\lambda)\boldsymbol{X}\sup_{\boldsymbol{x}}\Big\{\frac{\boldsymbol{x}'\boldsymbol{A}(\hat\lambda)\boldsymbol{D}^2\boldsymbol{A}(\hat\lambda)\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{A}(\hat\lambda)\boldsymbol{x}}\Big\} \\
= & c \times \mathbf{ch}_{max}(\boldsymbol{A}(\hat\lambda)\boldsymbol{D}^2) \leq c\frac{d_1^2}{d_1+\hat\lambda}.
\end{aligned}
$$

Also,

$$
\begin{aligned}
\frac{\boldsymbol{X}'\boldsymbol{A}^2(\hat\lambda)\boldsymbol{D}\boldsymbol{Q_W}\boldsymbol{D}\boldsymbol{A}(\hat\lambda)\boldsymbol{X}}{\boldsymbol{X}'\boldsymbol{A}^2(\hat\lambda)\boldsymbol{X}}I(\hat\lambda>0) \leq & \sup_{\boldsymbol{x}}\Big\{\frac{\boldsymbol{x}'\boldsymbol{A}^2(\hat\lambda)\boldsymbol{D}^2\boldsymbol{A}(\hat\lambda)\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{A}^2(\hat\lambda)\boldsymbol{x}}\Big\} \\
= & \mathbf{ch}_{max}(\boldsymbol{A}(\hat\lambda)\boldsymbol{D}^2) = \frac{d_1^2}{d_1+\hat\lambda}.
\end{aligned}
$$

For the first term, it can be seen that

$$
\begin{aligned}
\mathrm{tr}\,[\boldsymbol{D}\boldsymbol{Q_W}\boldsymbol{D}\boldsymbol{A}(\hat\lambda)] = & \mathrm{tr}\,[\boldsymbol{D}\boldsymbol{Q_W}\boldsymbol{D}(\boldsymbol{D}+\hat\lambda\boldsymbol{I})^{-1}] \\
& - \mathrm{tr}\,[\{\boldsymbol{Z}'(\boldsymbol{D}+\hat\lambda\boldsymbol{I})^{-1}\boldsymbol{Z}\}^{-1}\boldsymbol{Z}'(\boldsymbol{D}+\hat\lambda\boldsymbol{I})^{-1}\boldsymbol{D}\boldsymbol{Q_W}\boldsymbol{D}(\boldsymbol{D}+\hat\lambda\boldsymbol{I})^{-1}\boldsymbol{Z}] \\
\geq & \frac{1}{d_1+\hat\lambda}\mathrm{tr}\,[\boldsymbol{D}^2\boldsymbol{Q_W}] - (d_1+\hat\lambda)\frac{d_1^2}{(d_1+\hat\lambda)^2}\mathrm{tr}\,[(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Q_W}\boldsymbol{Z}].
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\widehat{\Delta}(\hat\lambda) \leq & -2\Big\{\frac{1}{d_1+\hat\lambda}\mathrm{tr}\,[\boldsymbol{D}^2\boldsymbol{Q_W}] - \frac{d_1^2}{d_1+\hat\lambda}\mathrm{tr}\,[(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Q_W}\boldsymbol{Z}] - 2\frac{d_1^2}{d_1+\hat\lambda}\Big\} \\
& + c\frac{d_1^2}{d_1+\hat\lambda},
\end{aligned}
$$

which is not positive if c satisfies the condition (5.10). ■

5.3 Unconstrained empirical Bayes estimator

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

[1] Case of $\boldsymbol{t}(\boldsymbol{X}) = \boldsymbol{W}'\boldsymbol{X}$. In this case, let $\boldsymbol{Y}_4 = \boldsymbol{Y}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2$ and $\boldsymbol{\xi}_4 = \boldsymbol{\xi}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{\xi}_2$. Then we use the same arguments as in Section 4.3. Consider the decomposition (3.11) and note that $\boldsymbol{Y}_4$ is independent of $\boldsymbol{Y}_2$ and that

$$\begin{aligned}\boldsymbol{\mu} =&\boldsymbol{H}_1'\boldsymbol{\xi}_1 + \boldsymbol{H}_2'\boldsymbol{\xi}_2\\ =&\boldsymbol{H}_1'\boldsymbol{\xi}_4 + (\boldsymbol{H}_2' + \boldsymbol{H}_1'\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1})\boldsymbol{\xi}_2.\end{aligned}$$

We set up the linear regression structure $\boldsymbol{Z\beta}$ for $\boldsymbol{\mu}$. Since $\boldsymbol{\xi}_4 = (\boldsymbol{H}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{H}_2)\boldsymbol{\mu}$, it may be reasonable to assume the following prior distribution:

$$\begin{aligned}\boldsymbol{\xi}_4|\lambda \sim&\mathcal{N}_{k-m}(\boldsymbol{Z}_4\boldsymbol{\beta}, \lambda\boldsymbol{I}_{k-m}),\\ \boldsymbol{\xi}_2 \sim&1,\end{aligned} \tag{5.11}$$

for $\boldsymbol{Z}_4 = (\boldsymbol{H}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{H}_2)\boldsymbol{Z}$.

Combining the contents in Sections 4.3, 5.1 and 5.2, we get the empirical Bayes estimator given by

$$\begin{aligned}\widehat{\boldsymbol{\mu}}^{EB1} =&\boldsymbol{H}_1'\widehat{\boldsymbol{\xi}}_4^{EB}(\boldsymbol{Y}_4) + (\boldsymbol{H}_2' + \boldsymbol{H}_1'\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1})\boldsymbol{Y}_2\\ =&\boldsymbol{H}_1'\{\widehat{\boldsymbol{\xi}}_4^{EB}(\boldsymbol{Y}_4) + \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{Y}_2\} + \boldsymbol{H}_2'\boldsymbol{Y}_2.\end{aligned}$$

Here the empirical Bayes estimator $\widehat{\boldsymbol{\xi}}_4^{EB}(\boldsymbol{Y}_4)$ is given as follows: Note that $\boldsymbol{Y}_4|\boldsymbol{\xi}_4 \sim \mathcal{N}_{k-m}(\boldsymbol{\xi}_4, \boldsymbol{V}_{11.2})$ and $\boldsymbol{\xi}_4 \sim \mathcal{N}_{k-m}(\boldsymbol{Z}_4\boldsymbol{\beta}, \lambda\boldsymbol{I})$. According to the arguments in Sections 5.1 and 5.2, we estimate λ by $\hat{\lambda} = \max\{\lambda_0, 0\}$, where λ_0 is the solution of the equation $\boldsymbol{Y}_4'\boldsymbol{A}_4(\lambda_0)\boldsymbol{Y}_4 = c$ for

$$\boldsymbol{A}_4(\lambda) = \boldsymbol{V}_4^{-1} - \boldsymbol{Z}_4(\boldsymbol{Z}_4'\boldsymbol{V}_4^{-1}\boldsymbol{Z}_4)^{-1}\boldsymbol{Z}_4'\boldsymbol{V}_4^{-1},$$

for $\boldsymbol{V}_4 = \boldsymbol{V}_{11.2} + \lambda\boldsymbol{I}$. Then, the empirical Bayes estimator is written by

$$\widehat{\boldsymbol{\xi}}_4^{EB}(\boldsymbol{Y}_4) = \boldsymbol{Y}_4 - \boldsymbol{V}_{11.2}(\boldsymbol{V}_{11.2} + \hat{\lambda}\boldsymbol{I}_{k-m})^{-1}\big\{\boldsymbol{Y}_4 - \boldsymbol{Z}_4\widehat{\boldsymbol{\beta}}_4(\hat{\lambda})\big\}, \tag{5.12}$$

for

$$\widehat{\boldsymbol{\beta}}_4(\lambda) = (\boldsymbol{Z}_4'\boldsymbol{V}_4^{-1}\boldsymbol{Z}_4)^{-1}\boldsymbol{Z}_4'\boldsymbol{V}_4^{-1}\boldsymbol{Y}_4.$$

Clearly, $\widehat{\boldsymbol{\mu}}^{EB1}$ satisfies the constraint, namely, $\boldsymbol{W}'\widehat{\boldsymbol{\mu}}^{EB1} = \boldsymbol{W}'\boldsymbol{X}$. The minimaxity of $\widehat{\boldsymbol{\mu}}^{EB1}$ follows from Proposition 5.1.

Proposition 5.2 *The unconstrained empirical Bayes estimator $\widehat{\boldsymbol{\mu}}^{EB1}$ satisfies the constraint $\boldsymbol{W}'\widehat{\boldsymbol{\mu}}^{EB1} = \boldsymbol{W}'\boldsymbol{X}$. It is also minimax if*

$$0 < c \le 2\Big\{\{\mathbf{ch}_{\mathbf{max}}(\boldsymbol{V}_{\mathbf{11.2}})\}^{-2}\mathrm{tr}\,[\boldsymbol{V}_{11.2}^2\boldsymbol{Q}_{\boldsymbol{W}}] - \mathrm{tr}\,[(\boldsymbol{Z}_4'\boldsymbol{Z}_4)^{-1}\boldsymbol{Z}_4'\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{Z}_4] - 2\Big\}. \tag{5.13}$$

[2] Case of $\boldsymbol{t}(\boldsymbol{X}) = \boldsymbol{t}_0$ or $\boldsymbol{W}'\boldsymbol{\mu} = \boldsymbol{t}_0$. In this case, we can derive a desired result by combining the arguments given above in the case of $\boldsymbol{t}(\boldsymbol{X}) = \boldsymbol{W}'\boldsymbol{X}$ and the contents given in Subsection 4.4. Assume the prior distribution

$$\begin{aligned}\boldsymbol{\xi}_1|\lambda \sim&\mathcal{N}_{k-m}(\boldsymbol{H}_1\boldsymbol{Z\beta}, \lambda\boldsymbol{I}_{k-m}),\\ \boldsymbol{\xi}_2 =&\boldsymbol{\xi}_0 \quad \text{with probability one},\end{aligned} \tag{5.14}$$

where $\boldsymbol{\xi}_0 = (\boldsymbol{W}'\boldsymbol{H}_2')^{-1}\boldsymbol{t}_0$. For $\boldsymbol{Y}_5 = \boldsymbol{Y}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}(\boldsymbol{Y}_2 - \boldsymbol{\xi}_0)$, the joint distribution of $(\boldsymbol{Y}_5, \boldsymbol{Y}_2)$ is given in (4.6), so that the empirical Bayes estimator is

$$\widehat{\boldsymbol{\mu}}^{EB2} = \boldsymbol{H}_1'\widehat{\boldsymbol{\xi}}_1^{EB}(\boldsymbol{Y}_5) + \boldsymbol{H}_2'\boldsymbol{\xi}_0,$$

where $\widehat{\boldsymbol{\xi}}_1^{EB}(\boldsymbol{Y}_5)$ has the same form as $\widehat{\boldsymbol{\xi}}_4^{EB}(\boldsymbol{Y}_4)$ given in (5.12) except replacing $\boldsymbol{Y}_4$ and $\boldsymbol{Z}_4$ with $\boldsymbol{Y}_5$ and $\boldsymbol{H}_1\boldsymbol{Z}$, respectively.

Clearly, $\widehat{\boldsymbol{\mu}}^{EB2}$ satisfies the constraint, namely, $\boldsymbol{W}'\widehat{\boldsymbol{\mu}}^{EB2} = \boldsymbol{t}_0$. The improvement of $\widehat{\boldsymbol{\mu}}^{EB2}$ follows from Propositions 5.1 and 5.2. When the parameter space is restricted on $\boldsymbol{W}'\boldsymbol{\mu} = \boldsymbol{t}_0$ or $\boldsymbol{\xi}_2 = \boldsymbol{\xi}_0$, $\widehat{\boldsymbol{\mu}}^{EB1}$ is minimax.

Proposition 5.3 *The unconstrained empirical Bayes estimator $\widehat{\boldsymbol{\mu}}^{EB2}$ satisfies the constraint $\boldsymbol{W}'\widehat{\boldsymbol{\mu}}^{EB2} = \boldsymbol{t}_0$. $\widehat{\boldsymbol{\mu}}^{EB2}$ dominates the estimator $\boldsymbol{H}_1'\boldsymbol{Y}_5 + \boldsymbol{H}_2'\boldsymbol{\xi}_0$ if*

$$0 < c \le 2\Big\{\{\mathbf{ch_{max}}(\boldsymbol{V}_{\mathbf{11.2}})\}^{-2}\mathrm{tr}\,[\boldsymbol{V}_{11.2}^2\boldsymbol{Q}_{\boldsymbol{W}}] - \mathrm{tr}\,[(\boldsymbol{Z}'\boldsymbol{H}_1'\boldsymbol{H}_1\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{H}_1'\boldsymbol{Q}_{\boldsymbol{W}}\boldsymbol{H}_1\boldsymbol{Z}] - 2\Big\}. \tag{5.15}$$

When the parameter space is restricted on $\boldsymbol{W}'\boldsymbol{\mu} = \boldsymbol{t}_0$, $\widehat{\boldsymbol{\mu}}^{EB2}$ is minimax under the condition (5.15).

6 Concluding remarks

Benchmarking has been recognized as important in small area problems, and constrained Bayesian estimators have been studied in the literature. ¿From a decision-theoretic point of view, however, little has been known about admissibility and minimaxity properties of constrained generalized Bayes estimators. In this paper, we have clarified admissibility, minimaxity and dominance properties of benchmarked estimators by decomposing the 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, and to provide admissible and minimax estimators in the constrained problem. 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 a constrained empirical Bayes and minimax estimator in the Fay-Herriot model.

An interesting, but unresolved problem is admissibility or inadmissibility of the generalized Bayes estimator $\widehat{\boldsymbol{\mu}}^{GB1}$ given in (4.4). As shown in Proposition 4.2, $\widehat{\boldsymbol{\mu}}^{GB1}$ is admissible and minimax within the constrained class Γ_1 if $k - m \ge 3$ and $0 \le a \le k - m - 2$. When we consider admissibility in the unconstrained problem, however, this estimator is not admissible if $m \ge 3$. We conjecture that the estimator $\widehat{\boldsymbol{\mu}}^{GB1}$ is admissible (in the unconstrained problem) in the case of $k - m \ge 3$ and $m = 1, 2$.

Although a constrained empirical Bayes estimator is treated in Section 5, 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 (1990), Berger and

Strawderman (1996) and Kubokawa and Strawderman (2007) 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.

Acknowledgments. 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. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.*, **4**, 223-226.

[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.

[4] Berger, J.O. and Strawderman, W.E. (1996). Choice of hierarchical priors: Admissibility in estimation of normal means. *Ann. Statist.*, **24**, 931-951.

[5] Bock, M.E. (1975). Minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.*, **3**, 209-218.

[6] Brown, L.D. (1971). Admissible estimators, recurrent diffusions, and insolvable boundary value problems. *Ann. Math. Statist.*, **42**, 855-904.

[7] Datta, G.S., Ghosh, M., Steorts, R., and Maples, J. (2011). Bayesian benchmarking with applications to small area estimation. *Test*, , to appear.

[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.

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

[10] Frey, J., and Cressie, N. (2003). Some results on constrained Bayes estimators. *Statist. Probab. Letters*, **65**, 389-399.

[11] Ghosh, M. (1992). Constrained Bayes estimation with applications. *J. American Statist. Assoc.*, **87**, 533-540.

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

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

[14] Kubokawa, T. (2011). On measuring uncertainty of small area estimators with higher order accuracy. *J. Japan Statist. Soc.*, to appear.

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

[16] Lindley, D.V. (1962). Discussion of "Confidence sets for the mean of a multivariate normal distribution", by C. Stein. *J. Roy. Statist. Soc., B*, **24**, 285-287.

[17] Pfeffermann, D., and Tiller, R. (2006). Small-area estimation with state-space models subject to benchmark constraints. *J. American Statist. Assoc.*, **101**, 1387-1397.

[18] 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.

[19] Rao, J.N.K. (2003). *Small Area Estimation*. Wiley, New Jersey.

[20] Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components*, Wiley, New York.

[21] Shinozaki, N. (1974). A note on estimating the mean vector of a multivariate normal distribution with general quadratic loss function. *Keio Engineering Reports*, 105-112.

[22] Shinozaki, N. and Chang, Y.-T. (1993). Minimaxity of empirical Bayes estimators of the means of independent normal variables with unequal variances. *Commun. Statist. - Theory Method*, **22**, 2147-2169.

[23] Shinozaki, N. and Chang, Y.-T. (1996). Minimaxity of empirical Bayes estimators shrinking toward the grand mean when variances are unequal. *Commun. Statist. - Theory Method*, **25**, 183-199.

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

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

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