CIRJE-F-187

Minimax Multivariate Empirical Bayes Estimators under Multicollinearity

M.S. Srivastava
University of Toronto

Tatsuya Kubokawa
University of Tokyo

December 2002; Revised in August 2009

Minimax Multivariate Empirical Bayes Estimators under Multicollinearity

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

January 13, 2004

In this paper we consider the problem of estimating the matrix of regression coefficients in a multivariate linear regression model in which the design matrix is near singular. Under the assumption of normality, we propose empirical Bayes ridge regression estimators with three types of shrinkage functions, that is, scalar, componentwise and matricial shrinkage. These proposed estimators are proved to be uniformly better than the least squares estimator, that is, minimax in terms of risk under the Strawderman's loss function. Through simulation and empirical studies, they are also shown to be useful in the multicollinearity cases.

Key words and phrases: Empirical Bayes estimator, ridge regression estimator, multicollinearity, multivariate linear regression model, multivariate normal distribution.

AMS subject classifications: Primary 62J07, 62C12, 62C20, Secondary 62J05, 62C10.

1 Introduction

Consider a multivariate linear regression model in which a vector $\boldsymbol{y}$ of p responses depends linearly on m independent variables $z_1, \ldots, z_m$ as

$$\boldsymbol{y} = \boldsymbol{\beta}^t \boldsymbol{z} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$, $\boldsymbol{z}^t = (z_1, \ldots, z_m)$ and $\boldsymbol{\beta}$ is an $m \times p$ matrix of unknown regression parameters. Writing

$$\boldsymbol{\beta}^t = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m) \quad \text{and} \quad \boldsymbol{\beta} = (\boldsymbol{\beta}_{(1)}, \ldots, \boldsymbol{\beta}_{(p)})$$

we find that $\boldsymbol{\beta}_i$ is the vector of regression coefficients associated with the independent variables z_i. With N independent observations on $\boldsymbol{y}$ and with the corresponding N values on $\boldsymbol{z}$ denoted by an $N \times m$ matrix $\boldsymbol{Z}$ of rank m, the regression model becomes

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{E},$$

*Department of Statistics, University of Toronto, 100 St George Street, Toronto, Ontario, CANADA M5S 3G3; E-Mail: srivasta@utstat.utstat.toronto.edu ; Fax: +1-416-978-5133

†Faculty of Economics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN; E-Mail: tatsuya@e.u-tokyo.ac.jp ; Fax: +81-3-5841-5521

where

$$\boldsymbol{Y} = (\boldsymbol{y}_{(1)}, \dots, \boldsymbol{y}_{(p)}) = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_N)^t \ : \ N \times p$$

and the N rows of $\boldsymbol{E}$ are i.i.d. $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. The least squares estimate of $\boldsymbol{\beta}_{(i)}$ is given by

$$\widehat{\boldsymbol{\beta}}_{(i)} = (\boldsymbol{Z}^t\boldsymbol{Z})^{-1}\boldsymbol{Z}^t\boldsymbol{y}_{(i)}, \quad i = 1, \dots, p$$

which can be written compactly as

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{Z}^t\boldsymbol{Z})^{-1}\boldsymbol{Z}^t\boldsymbol{Y}.$$

When some of the independent variables $z_1, \dots, z_m$ are highly correlated, the matrix $\boldsymbol{Z}^t\boldsymbol{Z}$ is near singular and the least squares estimator $\widehat{\boldsymbol{\beta}}$ becomes unstable. In such a situation, known as multicollinearity in the literature, the regression coefficient vector $\boldsymbol{\beta}_i$ corresponding to the highly correlated independent variable z_i is shrunken or pulled towards zero by using Stein-type estimators or ridge-regression type estimators proposed by Hoerl and Kennard (1970). However, because of simplicity and ease of computation since the least squares computing packages can also be used for ridge regression estimators (see Sen and Srivastava, 1990, p 257), the ridge-regression estimator is a popular procedure among practicing statisticians. The most commonly used ridge regression estimator is given by

$$(\boldsymbol{Z}^t\boldsymbol{Z} + \boldsymbol{K})^{-1}\boldsymbol{Z}^t\boldsymbol{Y}, \tag{1.1}$$

where $\boldsymbol{K}$ is an $m \times m$ matrix chosen on the basis of some criteria; $\boldsymbol{K}$ is also sometimes chosen as a diagonal matrix. Some authors, such as Breiman and Friedman (1997), however, apply ridge regression estimators to $\widehat{\boldsymbol{\beta}}_{(i)}$ separately for each of the p regressions for the p response variables, namely, they consider

$$\widehat{\boldsymbol{\beta}}_{(i)}(k_i) = (\boldsymbol{Z}^t\boldsymbol{Z} + k_i\boldsymbol{I})^{-1}\boldsymbol{Z}^t\boldsymbol{y}_{(i)}, \quad i = 1, \dots, p. \tag{1.2}$$

While both (1.1) and (1.2) shrinks the matrix regression coefficients $\boldsymbol{\beta}$, it is not clear if either of them shrinks $\widehat{\boldsymbol{\beta}}_i$ corresponding to the highly correlated variable $\boldsymbol{z}_i$.

In this paper we design the shrinkage in a manner that achieves the above mentioned goal of shrinking the 'culprit' $\widehat{\boldsymbol{\beta}}_i$ towards zero. In addition, we provide minimax estimators under an appropriate loss function of the regression parameters. Attempts in the past to obtain minimax adaptive ridge regression estimators of the matrix $\boldsymbol{K}$ in (1.1) have not been successful, see for example, Brown and Zidek (1980, 82). On the other hand, minimax estimators of Stein-type (shrinkage) have been proposed in the literature for regression parameters by Bilodeau and Kariya (1989), Konno (1990, 1991) and Srivastava and Solanky (2003). However, Srivastava and Solanky (2003) have shown that one of the estimators proposed by Konno (1991) is the best among the many shrinkage estimators available in the literature including the one proposed by Breiman and Friedman (1997) whose minimaxity is not known. Thus in our comparison we shall include Konno's estimator, defined in Section 4.

The organization of the paper is as follows: In Section 2, we reduce the problem to a canonical form and then propose empirical Bayes ridge regression estimators with three types of shrinkage functions, that is, scalar, componentwise and matricial shrinkage. In Section 3, these proposed estimators are proved to be uniformly better than the least squares estimator, that is, minimax in terms of risk under the Strawderman's loss function. In Section 4, we investigate risk-behaviors of the proposed esitmators, principal component regression estimators and Konno's estimator under the loss function $L_j(\omega, \boldsymbol{\delta}, (\boldsymbol{Z}^t\boldsymbol{Z})^j) = (\boldsymbol{\delta}-\boldsymbol{\beta})^t(\boldsymbol{Z}^t\boldsymbol{Z})^j(\boldsymbol{\delta}-\boldsymbol{\beta})$, $j = 0, 1, 2$. These procedures are also applied to the chemometrics data analyzed by Skagerberg, MacGregor and Kiparissides (1992) and compared through prediction error estimated via the leave-one-out cross-validation. Through these numerical and empirical studies, the minimax empirical Bayes ridge regression estimators are useful in the multicollinearity cases.

2 Minimax Empirical Bayes Ridge Regression Estimators

Following the notation of Srivastava and Khatri (1979, pp 54, 55), under the assumption of normality,

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_{m,p}(\boldsymbol{\beta}, (\boldsymbol{Z}^t\boldsymbol{Z})^{-1}, \boldsymbol{\Sigma}).$$

For obtaining minimax estimators of $\boldsymbol{\beta}$, we shall consider the loss function

$$L(\omega, \widetilde{\boldsymbol{\beta}}, (\boldsymbol{Z}^t\boldsymbol{Z})^2) = \mathrm{tr}\,(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^t(\boldsymbol{Z}^t\boldsymbol{Z})^2, \tag{2.1}$$

for any estimator $\widetilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and $\omega = (\boldsymbol{\beta}, \boldsymbol{\Sigma})$. This loss function was proposed by Strawderman (1978), and it is most appropriate for multicollinearity case.

Let $\boldsymbol{P}$ be an $m \times m$ orthogonal matrix such that $\boldsymbol{P}(\boldsymbol{Z}^t\boldsymbol{Z})^{-1}\boldsymbol{P}^t = \boldsymbol{D} = \mathrm{diag}\,(d_1, \ldots, d_m)$ for $d_1 \geq \ldots \geq d_m > 0$. Then, with

$$\boldsymbol{X} = \boldsymbol{P}\widehat{\boldsymbol{\beta}} \quad \text{and} \quad \boldsymbol{\Theta} = \boldsymbol{P}\boldsymbol{\beta}, \tag{2.2}$$

we find that

$$\boldsymbol{X} \sim \mathcal{N}_{m,p}(\boldsymbol{\Theta}, \boldsymbol{D}, \boldsymbol{\Sigma}). \tag{2.3}$$

In terms of the above transformations, the above loss function (2.1) becomes

$$L(\omega, \widetilde{\boldsymbol{\Theta}}, \boldsymbol{D}^{-2}) = \mathrm{tr}\,(\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta})\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{\Theta}} - \boldsymbol{\Theta})^t\boldsymbol{D}^{-2}. \tag{2.4}$$

where $\widetilde{\boldsymbol{\Theta}} = \boldsymbol{H}\widetilde{\boldsymbol{\beta}}$ is an estimator of $\boldsymbol{\Theta}$. Writing

$$\boldsymbol{X}^t = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) \quad \text{and} \quad \boldsymbol{\Theta}^t = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m),$$

we find that $\boldsymbol{x}_i$'s are independently distributed as

$$\boldsymbol{x}_i \sim \mathcal{N}_p(\boldsymbol{\theta}_i, d_i\boldsymbol{\Sigma}), \quad i = 1, \ldots, m.$$

Here d_i's are known numbers but $\boldsymbol{\Sigma}$ is unknown which can be estimated by $n^{-1}\boldsymbol{S}$ where

$$\boldsymbol{S} = (\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\beta}})^t(\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\beta}}), \quad n = N - m,$$

and is distributed independently of $\boldsymbol{x}_i$, $i = 1, \ldots, m$, as $\mathcal{W}_p(n, \boldsymbol{\Sigma})$. Thus, the problem reduces to that of estimating $\boldsymbol{\theta}_i$ from $\boldsymbol{x}_i$ which has covariance $d_i\boldsymbol{\Sigma}$, the inequality in covariances of $\boldsymbol{x}_i$ is through the known numbers d_i.

Three types of empirical Bayes ridge regression estimators of $\boldsymbol{\Theta}$ are proposed in the following subsections.

2.1 Scalar shrinkage empirical Bayes estimator

In the model $\boldsymbol{x}_i \sim \mathcal{N}_p(\boldsymbol{\theta}_i, d_i\boldsymbol{\Sigma})$, $i = 1, \ldots, m$, where $d_1 \geq \ldots \geq d_m$, we suppose that $\boldsymbol{\theta}_i$ has a prior distribution $\mathcal{N}_p(\boldsymbol{0}, \lambda\boldsymbol{\Sigma})$. Then the posterior distribution of $\boldsymbol{\theta}_i$ given $\boldsymbol{x}_i$ has $\mathcal{N}_p(\widehat{\boldsymbol{\theta}}_i^B(\lambda), (d_i^{-1} + \lambda^{-1})^{-1}\boldsymbol{\Sigma})$ where $\widehat{\boldsymbol{\theta}}_i^B(\lambda)$ is the Bayes estimator of $\boldsymbol{\theta}_i$ given by

$$\widehat{\boldsymbol{\theta}}_i^B(\lambda) = \boldsymbol{x}_i - \frac{d_i}{d_i + \lambda}\boldsymbol{x}_i, \tag{2.5}$$

and the Bayes estimator of $\boldsymbol{\Theta}$ is $\widehat{\boldsymbol{\Theta}}^B(\lambda)$ where $\{\widehat{\boldsymbol{\Theta}}^B(\lambda)\}^t = (\widehat{\boldsymbol{\theta}}_1^B(\lambda), \ldots, \widehat{\boldsymbol{\theta}}_m^B(\lambda))$. Since $\boldsymbol{x}_i$ is marginally distributed as $\mathcal{N}_p(\boldsymbol{0}, (d_i + \lambda)\boldsymbol{\Sigma})$, we have that $E[\sum_{i=1}^m \boldsymbol{x}_i^t\boldsymbol{S}^{-1}\boldsymbol{x}_i/(d_i + \lambda)] = mp/(n - p - 1)$. Taking this moment into account, we consider the solution λ^* of the equation

$$\sum_{i=1}^m \boldsymbol{x}_i^t\boldsymbol{S}^{-1}\boldsymbol{x}_i/(d_i + \lambda^*) = (mp - 2)/(n - p + 3). \tag{2.6}$$

Also let λ_{s0} be the root of the equation

$$\sum_{i=1}^m \frac{d_i - d_m}{d_i + \lambda_{s0}} = \frac{pm - 2}{2p}, \tag{2.7}$$

and define the estimator $\hat{\lambda}^{SB}$ of λ by

$$\hat{\lambda}^{SB} = \max(\lambda^*,\ \lambda_{s0}). \tag{2.8}$$

We thus get the estimator $\widehat{\boldsymbol{\Theta}}^{SB} = (\widehat{\boldsymbol{\theta}}_1^{SB}, \ldots, \widehat{\boldsymbol{\theta}}_m^{SB})^t$ where

$$\widehat{\boldsymbol{\theta}}_i^{SB} = \widehat{\boldsymbol{\theta}}_i^B(\hat{\lambda}^{SB}) = \boldsymbol{x}_i - \frac{d_i}{d_i + \hat{\lambda}^{SB}}\boldsymbol{x}_i, \tag{2.9}$$

which we call *the scalar shrinkage empirical Bayes estimator*, denoted by SB.

Theorem 1. *Assume that $pm \geq 3$. Then the scalar shrinkage empirical Bayes estimator $\widehat{\boldsymbol{\Theta}}^{SB}$ is minimax under Strawderman's loss (2.4).*

2.2 Componentwise shrinkage empirical Bayes estimator

Suppose that $\boldsymbol{\theta}_i$ has a priori distribution $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{1/2})$ for $\boldsymbol{\Lambda} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_p)$. Then the posterior distribution of $\boldsymbol{\theta}_i$ given $\boldsymbol{x}_i$ has $\mathcal{N}_p(\widehat{\boldsymbol{\theta}}_i^B(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma}^{1/2}(d_i^{-1}\boldsymbol{I}_p + \boldsymbol{\Lambda}^{-1})^{-1}\boldsymbol{\Sigma}^{1/2})$ where $\widehat{\boldsymbol{\theta}}_i^B(\boldsymbol{\Lambda}, \boldsymbol{\Sigma})$ is the Bayes estimator of $\boldsymbol{\theta}_i$ given by

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_i^B(\boldsymbol{\Lambda}, \boldsymbol{\Sigma}) = &\left(d_i^{-1}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Sigma}^{-1/2}\right)^{-1} d_i^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i \\ =&\boldsymbol{x}_i - d_i\boldsymbol{\Sigma}^{1/2}(d_i\boldsymbol{I}_p + \boldsymbol{\Lambda})^{-1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}_i.\end{aligned} \tag{2.10}$$

Since $\boldsymbol{x}_i$ is marginally distributed as $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}^{1/2}(d_i\boldsymbol{I}_p + \boldsymbol{\Lambda})\boldsymbol{\Sigma}^{1/2})$, the estimate of the parameter $\boldsymbol{\Lambda}$ may be based on $\boldsymbol{S}$ and $\boldsymbol{X}$ by using their mariginal distributions.

Let $\boldsymbol{H}$ be an orthogonal matrix such that $\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}^t = \boldsymbol{L} = \mathrm{diag}\,(\ell_1, \ldots, \ell_p)$, $\ell_1 \geq \cdots \geq \ell_p$. For $j = 1, \ldots, p$, let λ_j^* be the solution of the equation

$$\sum_{i=1}^{m} \frac{(\boldsymbol{h}_j^t \boldsymbol{x}_i)^2/\ell_j}{d_i + \lambda_j^*} = c_0, \quad j = 1, \ldots, p, \tag{2.11}$$

where $\boldsymbol{H}^t = (\boldsymbol{h}_1, \ldots, \boldsymbol{h}_p)$ and $c_0 = (m-2)/(np+2)$. Also let λ_{c0} be the solution of the equation

$$\sum_{i=1}^{m} \frac{d_i - d_m}{d_i + \lambda_{c0}} = \frac{m-2}{2}, \tag{2.12}$$

and define the estimator $\hat{\lambda}_j^{CB}$ of λ_j by

$$\hat{\lambda}_j^{CB} = \max(\lambda_j^*, \lambda_{c0}), \quad j = 1, \ldots, p. \tag{2.13}$$

We thus consider the estimator $\widehat{\boldsymbol{\Theta}}^{CB} = (\widehat{\boldsymbol{\theta}}_1^{CB}, \ldots, \widehat{\boldsymbol{\theta}}_m^{CB})^t$ given by

$$\widehat{\boldsymbol{\theta}}_i^{CB} = \boldsymbol{x}_i - d_i \boldsymbol{H}^t \boldsymbol{\Psi}_i \boldsymbol{H} \boldsymbol{x}, \tag{2.14}$$

which we call *the componentwise shrinkage empirical Bayes estimator*, denoted by CB, where $\boldsymbol{\Psi} = \mathrm{diag}\,(\psi_1^{(i)}, \ldots, \psi_p^{(i)})$ for

$$\psi_j^{(i)} = \frac{1}{d_i + \hat{\lambda}_j^{CB}}, \quad j = 1, \ldots, p. \tag{2.15}$$

Theorem 2. *Assume that $m \geq 3$. Then the componentwise shrinkage empirical Bayes estimator $\widehat{\boldsymbol{\Theta}}^{CB}$ is minimax under Strawderman's loss (2.4).*

We can also propose a convex combination of $\hat{\theta}_i^{SB}$ and $\widehat{\boldsymbol{\theta}}_i^{CB}$ as an estimator of $\boldsymbol{\theta}_i$. For example,

$$\widehat{\boldsymbol{\theta}}_i^{CC}(c) = \frac{cd_i}{cd_i + d_1} \widehat{\boldsymbol{\theta}}_i^{SB} + \frac{d_1}{cd_i + d_1} \widehat{\boldsymbol{\theta}}_i^{CB}, \tag{2.16}$$

where c is a constant, may be considered as a viable estimator. In the simulation and empirical studies given in Section 4, we put $c = 5$. This combined estimator of $\boldsymbol{\Theta}$ is denoted by $\widehat{\boldsymbol{\Theta}}^{CC}(c)$. When d_i is large, the combined estimator $\widehat{\boldsymbol{\theta}}_i^{CC}(c)$ is close to the scalar shrinkage empirical Bayes estimator $\widehat{\boldsymbol{\theta}}_i^{SB}$. When the d_i is small, on the other hand, the componentwise shrinkage estimator $\widehat{\boldsymbol{\theta}}_i^{CB}$ will affect the risk gain effectively.

Corollary 1. *The combined estimator $\widehat{\boldsymbol{\Theta}}^{CC}(c)$ is minimax if $m \geq 3$ under Strawderman's loss.*

2.3 Matricial shrinkage empirical Bayes estimator

Suppose that $\boldsymbol{\theta}_i$ has a priori distribution $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{1/2})$ for fully unknown positive definite matrix $\boldsymbol{\Gamma}$. Then the posterior distribution of $\boldsymbol{\theta}_i$ given $\boldsymbol{x}_i$ has $\mathcal{N}_p(\widehat{\boldsymbol{\theta}}_i^{B}(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}), (d_i^{-1}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Sigma}^{-1/2})^{-1})$ where $\widehat{\boldsymbol{\theta}}_i^{B}(\boldsymbol{\Gamma}, \boldsymbol{\Sigma})$ is the Bayes estimator of $\boldsymbol{\theta}_i$ given by

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_i^{B}(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}) &= \left(d_i^{-1}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Sigma}^{-1/2}\right)^{-1} d_i^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i \\ &= \boldsymbol{x}_i - d_i \boldsymbol{\Sigma}^{1/2}(d_i \boldsymbol{I}_p + \boldsymbol{\Gamma})^{-1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}_i.\end{aligned} \tag{2.17}$$

Since $\boldsymbol{x}_i$ is marginally distributed as $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}^{1/2}(d_i \boldsymbol{I}_p + \boldsymbol{\Gamma})\boldsymbol{\Sigma}^{1/2})$, the estimate of the parameter $\boldsymbol{\Gamma}$ may be based on $\boldsymbol{S}$ and $\boldsymbol{X}$ by using their marginal distributions. However, it seems difficult to provide the estimate as a solution of an equation like (2.6) and (2.11), so that we here treat another type of estimator. Let

$$\boldsymbol{A} = \mathrm{diag}\,(d_1 + 1, \dots, d_m + 1)/(d_1 + 1), \tag{2.18}$$

and let $\boldsymbol{Q}$ be a $(p \times p)$ nonsingular matrix such that

$$\boldsymbol{Q}^t \boldsymbol{S} \boldsymbol{Q} = \boldsymbol{I}_p \quad \text{and} \quad \boldsymbol{Q}^t \boldsymbol{X}^t \boldsymbol{A}^{-1} \boldsymbol{X} \boldsymbol{Q} = \boldsymbol{F}, \tag{2.19}$$

where $\boldsymbol{F}$ is a diagonal matrix, $\boldsymbol{F} = \mathrm{diag}\,(f_1, \dots, f_p)$ and $f_1 \geq \cdots \geq f_p$. Clearly f_i's are the eigenvalues of $\boldsymbol{S}^{-1}\boldsymbol{X}^t\boldsymbol{A}^{-1}\boldsymbol{X}$. Let λ_{m0} and λ_{m1} be the solutions of the equations

$$\sum_{i=1}^{m} \frac{d_i - d_m}{d_i + \lambda_{m0}} = \frac{(p-1)(p+2)}{2p}, \tag{2.20}$$

$$\sum_{i=1}^{m} \frac{d_i - d_m}{d_i + \lambda_{m1}} = \frac{m-p-1}{2}. \tag{2.21}$$

The adaptive ridge regression estimator of $\boldsymbol{\theta}_i$ is given by

$$\widehat{\boldsymbol{\theta}}_i^{MB} = \boldsymbol{x}_i - d_i(\boldsymbol{Q}^t)^{-1}\boldsymbol{\Phi}_i(\boldsymbol{F})\boldsymbol{Q}^t\boldsymbol{x}_i, \quad i = 1, \dots, m \tag{2.22}$$

where $\boldsymbol{\Phi}_i(\boldsymbol{F}) = \mathrm{diag}\,(\phi_1^{(i)}, \dots, \phi_p^{(i)})$ and for $j = 1, \dots, p$,

$$\phi_j^{(i)} = \frac{1}{d_i + \hat{\lambda}_0^{MB}} + \frac{1}{d_i + \hat{\lambda}_j^{MB}}, \tag{2.23}$$

$$\hat{\lambda}_0^{MB} = \max(c_0 \mathrm{tr}\, \boldsymbol{F}, \lambda_{m0}), \quad c_0 = \frac{n-p+3}{(p-1)(p+2)}, \tag{2.24}$$

$$\hat{\lambda}_j^{MB} = \max(c_1 f_j, \lambda_{m1}), \quad c_1 = \frac{n+p+1}{m-p-1}. \tag{2.25}$$

It is noted that $\widehat{\boldsymbol{\theta}}_i^{MB}$ is close to the estimator proposed by Efron and Morris (1976) in the case of $d_1 = \cdots = d_m$. We can prove the minimaxity of the estimator $\widehat{\boldsymbol{\Theta}}^{MB}$ for $(\widehat{\boldsymbol{\Theta}}^{MB})^t = (\widehat{\boldsymbol{\theta}}_1^{MB}, \dots, \widehat{\boldsymbol{\theta}}_m^{MB})$.

Theorem 3. *Assume that $m \geq p+2$. Then the estimator $\widehat{\boldsymbol{\Theta}}^{MB}$ is minimax under the loss (2.4).*

We can also propose a convex combination of $\hat{\theta}_i^{SB}$ and $\widehat{\boldsymbol{\theta}}_i^{MB}$ as an estimator of $\boldsymbol{\theta}_i$. One such estimator is given by

$$\widehat{\boldsymbol{\theta}}_i^{MC}(c) = \frac{cd_i}{cd_i + d_1}\widehat{\boldsymbol{\theta}}_i^{SB} + \frac{d_1}{cd_i + d_1}\widehat{\boldsymbol{\theta}}_i^{MB}, \tag{2.26}$$

where c is a constant. In the simulation and empirical studies given in Section 4, we put $c = 5$.

Corollary 2. *The combined estimator $\widehat{\boldsymbol{\Theta}}^{MC}(c)$ is minimax if $m \geq p+2$ under Strawderman's loss.*

3 Proofs

In this sention, we prove the three theorems stated in Section 2. It may be argued that since the first two cases are special cases of the matricial estimator, only the proof of Theorem 3 is required. However, different inequalities have been used in the proofs which lead to three different conditions in equations (2.7), (2.12) and (2.20) - (2.21) respectively. Thus, we need to provide proofs for all the three theorems. In the proofs, we need the following two well known results, one due to Stein (1973, 1981) and the other due to Stein (1977) and Haff (1979), known as the Stein-Haff identity.

Lemma 1. *(Stein Identity) Let $\boldsymbol{X} = (X_1, \dots, X_p)^t$ be a p-dimensional random variable having $\mathcal{N}_p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$. Consider a vector-valued absolutely continuous function $\boldsymbol{h}(\boldsymbol{X}) = (h_1(\boldsymbol{X}), \dots, h_p(\boldsymbol{X}))^t$ with $E[|(\boldsymbol{X}-\boldsymbol{\theta})\boldsymbol{h}(\boldsymbol{X})^t|] < \infty$. Then,*

$$E\left[(\boldsymbol{X}-\boldsymbol{\theta})\{\boldsymbol{h}(\boldsymbol{X})\}^t\right] = E\left[\boldsymbol{\Sigma}\boldsymbol{\nabla}\{\boldsymbol{h}(\boldsymbol{X})\}^t\right], \tag{3.1}$$

where $\boldsymbol{\nabla} = (\partial/\partial X_1, \dots, \partial/\partial X_p)^t$.

Lemma 2. *(Stein-Haff Identity) Let $\boldsymbol{Y} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_n)$, where $\boldsymbol{y}_i$ are i.i.d. $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{V} = \boldsymbol{Y}\boldsymbol{Y}^t = \sum_{j=1}^n \boldsymbol{Y}_j\boldsymbol{Y}_j^T$. Consider a $p\times p$ matrix-valued function $\boldsymbol{G}(\boldsymbol{V}) = (g_{ij}(\boldsymbol{V}))$, where $g_{ij}(\boldsymbol{V})$ is a real-valued absolutely continuous function of the $p \times p$ matrix $\boldsymbol{V} = (v_{ij})$ and $E[|g_{ij}(\boldsymbol{V})|] < \infty$. Then,*

$$E\left[\mathrm{tr}\,\boldsymbol{G}(\boldsymbol{V})\boldsymbol{\Sigma}^{-1}\right] = E\left[(n-p-1)\mathrm{tr}\,\boldsymbol{G}(\boldsymbol{V})\boldsymbol{V}^{-1} + 2\mathrm{tr}\,\mathcal{D}_V\boldsymbol{G}(\boldsymbol{V})\right], \tag{3.2}$$

where $(\mathcal{D}_V\boldsymbol{G}(\boldsymbol{V}))_{ij} = \sum_k \mathrm{d}_{ik}g_{kj}(\boldsymbol{V})$, $\mathrm{d}_{ik} = 2^{-1}(1+\delta_{ik})\partial/\partial v_{ik}$ and $\delta_{ik} = 0$ for $i \neq k$, $\delta_{ii} = 1$.

3.1 Proof of Theorem 1

In the proof below, we may assume without any loss of generality that $\boldsymbol{\Sigma} = \boldsymbol{I}$. The risk difference between the two estimators is given by

$$\begin{aligned}
\Delta =& R(\omega, \widehat{\boldsymbol{\Theta}}^{SB}) - R(\omega, \boldsymbol{X}) \\
=& -2\sum_{i=1}^m \frac{1}{d_i} E\left[\frac{\boldsymbol{x}_i^t(\boldsymbol{x}_i - \boldsymbol{\theta}_i)}{d_i + \hat{\lambda}^{SB}}\right] + \sum_{i=1}^m E\left[\frac{\boldsymbol{x}_i^t\boldsymbol{x}_i}{(d_i+\hat{\lambda}^{SB})^2}\right] \\
=& -2\sum_{i=1}^m E\left[\frac{p}{d_i+\hat{\lambda}^{SB}} - \frac{1}{(d_i+\hat{\lambda}^{SB})^2}\boldsymbol{x}_i^t\frac{\partial\hat{\lambda}^{SB}}{\partial\boldsymbol{x}_i}\right] + \sum_{i=1}^m E\left[\frac{\boldsymbol{x}_i^t\boldsymbol{x}_i}{(d_i+\hat{\lambda}^{SB})^2}\right],
\end{aligned} \tag{3.3}$$

from the Stein identity (3.1). Using the implicit function theorem, we get from (2.6)

$$\begin{aligned}
\sum_{i=1}^m \frac{\boldsymbol{x}_i^t}{(d_i+\hat{\lambda}^{SB})^2}\frac{\partial\hat{\lambda}^{SB}}{\partial\boldsymbol{x}_i} =& 2\frac{\sum_{i=1}^m \boldsymbol{x}_i^t\boldsymbol{S}^{-1}\boldsymbol{x}_i(d_i+\hat{\lambda}^{SB})^{-3}}{\sum_{i=1}^m \boldsymbol{x}_i^t\boldsymbol{S}^{-1}\boldsymbol{x}_i(d_i+\hat{\lambda}^{SB})^{-2}} I(\lambda^* > \lambda_{s0}) \\
<& 2/(d_m + \hat{\lambda}^{SB}).
\end{aligned} \tag{3.4}$$

To evaluate the second term in (3.3), we use the Stein-Haff identity (3.2) giving

$$\begin{aligned}
&E\left[(d_i+\hat{\lambda}^{SB})^{-2}\mathrm{tr}\,\boldsymbol{x}_i\boldsymbol{x}_i^t\right] \\
&= (n-p-1)E\left[(d_i+\hat{\lambda}^{SB})^{-2}\mathrm{tr}\,\boldsymbol{x}_i\boldsymbol{x}_i^t\boldsymbol{S}^{-1}\right] + 2E\left[\mathrm{tr}\,\mathcal{D}_S[(d_i+\hat{\lambda}^{SB})^{-2}\boldsymbol{x}_i\boldsymbol{x}_i^t]\right] \\
&= (n-p-1)E\left[(d_i+\hat{\lambda}^{SB})^{-2}\mathrm{tr}\,\boldsymbol{x}_i\boldsymbol{x}_i^t\boldsymbol{S}^{-1}\right] - 4E\left[(d_i+\hat{\lambda}^{SB})^{-3}\sum_{j=1}^p\sum_{i=1}^p c_{jk}^{(i)}\mathrm{d}_{jk}(\hat{\lambda}^{SB})\right]
\end{aligned}$$

for $(c_{jk}^{(i)}) = \boldsymbol{x}_i\boldsymbol{x}_i^t$. From (2.6) and the implicit function theorem, we get

$$\mathrm{d}_{jk}(\lambda^*) = -\frac{\sum_{\ell=1}^m (d_\ell+\lambda^*)^{-1}(\boldsymbol{x}_\ell^t\boldsymbol{f}_j)(\boldsymbol{x}_\ell^t\boldsymbol{f}_k)}{\sum_{a=1}^m (d_a+\lambda^*)^{-2}\boldsymbol{x}_a^t\boldsymbol{S}^{-1}\boldsymbol{x}_a},$$

where $\boldsymbol{S}^{-1} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_p)$, see Theorem 1.11.1 of Srivastava and Khatri (1979, p.28); the definition used in this paper requires to take half of the value given there. Thus,

$$\sum_{j=1}^p\sum_{k=1}^p c_{jk}^{(i)}\mathrm{d}_{jk}(\hat{\lambda}^*) = -\frac{\sum_{\ell=1}^m (d_\ell+\lambda^*)^{-1}(\boldsymbol{x}_\ell^t\boldsymbol{S}^{-1}\boldsymbol{x}_i)^2}{\sum_{a=1}^m (d_a+\hat{\lambda}^*)^{-2}\boldsymbol{x}_a^t\boldsymbol{S}^{-1}\boldsymbol{x}_a}.$$

From the Cauchy-Schwarz inequality $(\boldsymbol{x}_\ell^t\boldsymbol{S}^{-1}\boldsymbol{x}_i)^2 \le (\boldsymbol{x}_\ell^t\boldsymbol{S}^{-1}\boldsymbol{x}_\ell)(\boldsymbol{x}_i^t\boldsymbol{S}^{-1}\boldsymbol{x}_i)$, and hence

$$\begin{aligned}
&\sum_{i=1}^m \mathrm{tr}\,\mathcal{D}_S(d_i+\hat{\lambda}^{SB})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i^t\\
=&2\sum_{i=1}^m (d_i+\hat{\lambda}^{SB})^{-3}\frac{\sum_{\ell=1}^m (d_\ell+\hat{\lambda}^{SB})^{-1}(\boldsymbol{x}_\ell^t\boldsymbol{S}^{-1}\boldsymbol{x}_i)^2}{\sum_{a=1}^m (d_a+\hat{\lambda}^{SB})^{-2}\boldsymbol{x}_a^t\boldsymbol{S}^{-1}\boldsymbol{x}_a} I(\lambda^* > \lambda_{s0})\\
\le&2(d_m+\hat{\lambda}^{SB})^{-1}\frac{\sum_{i=1}^m (d_i+\hat{\lambda}^{SB})^{-2}\boldsymbol{x}_i^t\boldsymbol{S}^{-1}\boldsymbol{x}_i\sum_{\ell=1}^m (d_\ell+\hat{\lambda}^{SB})^{-1}\boldsymbol{x}_\ell^t\boldsymbol{S}^{-1}\boldsymbol{x}_\ell}{\sum_{a=1}^m (d_a+\hat{\lambda}^{SB})^{-2}\boldsymbol{x}_a^t\boldsymbol{S}^{-1}\boldsymbol{x}_a}\\
=&2(d_m+\hat{\lambda}^{SB})^{-1}(mp-2)/(n-p+3),
\end{aligned}$$

from (2.6). Thus,

$$\begin{aligned}
&\sum_{i=1}^m E\left[(d_i+\hat{\lambda}^{SB})^{-2}\boldsymbol{x}_i^t\boldsymbol{x}_i\right]\\
\le&E\left[(n-p-1)\sum_{i=1}^m (d_i+\hat{\lambda}^{SB})^{-2}\boldsymbol{x}_i^t\boldsymbol{S}^{-1}\boldsymbol{x}_i + 4(d_m+\hat{\lambda}^{SB})^{-1}\frac{mp-2}{n-p+3}\right]\\
\le&\frac{mp-2}{n-p+3}E\left[(d_m+\hat{\lambda}^{SB})^{-1}(n-p-1)+4(d_m+\hat{\lambda}^{SB})^{-1}\right]\\
=&(mp-2)E\left[(d_m+\hat{\lambda}^{SB})^{-1}\right].
\end{aligned}\tag{3.5}$$

Hence, combining (3.3), (3.4) and (3.5), we get

$$\boldsymbol{\Delta} \le E\left[-2p\sum_{i=1}^m (d_i+\hat{\lambda}^{SB})^{-1} + (mp+2)(d_m+\hat{\lambda}^{SB})^{-1}\right].$$

Thus, the risk difference is not positive if

$$-2p\sum_{i=1}^m (d_i+\hat{\lambda}^{SB})^{-1} + (mp+2)(d_m+\hat{\lambda}^{SB})^{-1} \le 0. \tag{3.6}$$

Noting that $\sum_{i=1}^m (d_m+\hat{\lambda}^{SB})/(d_i+\hat{\lambda}^{SB}) = m - \sum_{i=1}^m (d_i-d_m)/(d_i+\hat{\lambda}^{SB})$, the inequality (3.6) is satisfied if

$$\sum_{i=1}^m (d_i-d_m)/(d_i+\hat{\lambda}^{SB}) \le (pm-2)/(2p),$$

which is guaranteed by the definition of λ_{s0}. Therefore Theorem 1 is proved. ∎

3.2 Proof of Theorem 2

Let $\boldsymbol{G} = (g_{ab}) = \boldsymbol{H}\boldsymbol{\Sigma}^{1/2}$, $\boldsymbol{G}^{-1} = (g^{ab})$, $\boldsymbol{u}_i = (u_{1i}, \ldots, u_{pi})^t = d_i^{-1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}_i$ and $\boldsymbol{\eta}_i = (\eta_{1i}, \ldots, \eta_{pi})^t = d_i^{-1/2}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}_i$. Then (2.11) can be rewritten as

$$\sum_{k=1}^{m} \frac{d_k(\sum_b g_{ab}u_{bk})^2}{d_k + \lambda_a^*} = c_0\ell_a, \quad a = 1, \ldots, p.$$

From the implicit function theorem, we get

$$\frac{\partial \lambda_a^*}{\partial u_{ji}} = 2\frac{d_i(\sum_b g_{ab}u_{bi})g_{aj}/(d_i + \lambda_a^*)}{\sum_k d_k(\sum_b g_{ab}u_{bk})^2 g_{aj}/(d_k + \lambda_a^*)^2}, \tag{3.7}$$

and from the definition of $\psi_j^{(i)} = (d_i + \hat{\lambda}_j^{CB})^{-1}$ in (2.15),

$$\partial\psi_a^{(i)}/\partial u_{ji} = -(d_i + \lambda_a^*)^{-2}(\partial\lambda_a^*/\partial u_{ji})I(\lambda_a^* > \lambda_{c0}). \tag{3.8}$$

The risk difference between the two estimators $\widehat{\boldsymbol{\Theta}}^{CB}$ and $\boldsymbol{X}$ is

$$\begin{aligned}
\Delta =& -2\sum_{i=1}^{m} E\left[(\boldsymbol{u}_i - \boldsymbol{\eta}_i)^t\boldsymbol{G}^{-1}\boldsymbol{\Psi}_i\boldsymbol{G}\boldsymbol{u}_i\right] + \sum_{i=1}^{m} E\left[\boldsymbol{x}_i^t\boldsymbol{H}^t\boldsymbol{\Psi}_i\boldsymbol{H}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}^t\boldsymbol{\Psi}_i\boldsymbol{H}\boldsymbol{x}_i\right] \\
=& -2\sum_{i=1}^{m}\sum_{j,a,b}^{p} E\left[(u_{ji} - \eta_{ji})g^{ja}\psi_a^{(i)}g_{ab}u_{bi}\right] + I_3 \qquad (3.9)\\
=& -2\sum_{i=1}^{m}\sum_{j,a,b}^{p} E\left[\frac{\partial}{\partial u_{ji}}\left\{g^{ja}\psi_a^{(i)}g_{ab}u_{bi}\right\}\right] + I_3 \\
=& -2\sum_{i=1}^{m}\sum_{j,a,b}^{p} E\left[g^{ja}\psi_a^{(i)}g_{ab}\delta_{bj}\right] - 2\sum_{i=1}^{m}\sum_{j,a,b}^{p} E\left[g^{ja}g_{ab}u_{bi}\frac{\partial\psi_a^{(i)}}{\partial u_{ji}}\right] + I_3 \\
=& -2\sum_{i=1}^{m}\sum_{a=1}^{p} E\left[\psi_a^{(i)}\right] + I_2 + I_3, \quad \text{(say)}
\end{aligned}$$

using the Stein identity (3.1) and the fact that $\sum_j g_{aj}g^{ja} = 1$, where from (3.7) and (3.8)

$$\begin{aligned}
I_2 =& 4\sum_{j,a,b}^{p} E\left[\frac{\sum_{i=1}^m g^{ja}g_{ab}u_{bi}d_i(d_i + \lambda_a^*)^{-3}g_{aj}(\sum_b g_{ab}u_{bi})}{\sum_k d_k(\sum_b g_{ab}u_{bk})^2/(d_k + \lambda_a^*)^2}I(\lambda_a^* > \lambda_{c0})\right] \\
=& 4\sum_{a=1}^{p} E\left[\frac{\sum_{i=1}^m d_i(d_i + \lambda_a^*)^{-3}(\sum_b g_{ab}u_{bi})^2}{\sum_k d_k(\sum_b g_{ab}u_{bk})^2/(d_k + \lambda_a^*)^2}I(\lambda_a^* > \lambda_{c0})\right] \\
\le& 4\sum_{a=1}^{p} E(d_m + \hat{\lambda}_a^{CB})^{-1}. \qquad (3.10)
\end{aligned}$$

Hence,

$$\boldsymbol{\Delta} \le -2\sum_{j=1}^{p} E\left[\sum_{i=1}^{m}(d_i + \hat{\lambda}_j^{CB})^{-1} - 2(d_m + \hat{\lambda}_j^{CB})^{-1}\right] + I_3. \tag{3.11}$$

From (2.11),

$$\phi \equiv \sum_{i=1}^{m}\sum_{j=1}^{p}(\boldsymbol{h}_j^t\boldsymbol{x}_i)^2/[\ell_j(d_i+\hat\lambda_j^{CB})^2]$$
$$= \sum_{i=1}^{m}\mathrm{tr}\,[\boldsymbol{\Psi}_i\boldsymbol{H}\boldsymbol{x}_i\boldsymbol{x}_i^t\boldsymbol{H}^t\boldsymbol{\Psi}_i\boldsymbol{L}^{-1}] \le c_0\sum_{j=1}^{p}(d_m+\hat\lambda_j^{CB})^{-1}. \tag{3.12}$$

Let $a_{jj} = (\boldsymbol{H}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{H})_{jj}$. Then, using the same arguments as in Sheena (1995), and the inequality $\mathrm{tr}\,(\boldsymbol{AB}) \le (\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})$ for $\boldsymbol{A}$ and $\boldsymbol{B}$ p.s.d. matrices, we get

$$I_3 \le \sum_{i=1}^{m} E\left[\left\{\mathrm{tr}\,\boldsymbol{H}^t\boldsymbol{\Psi}_i\boldsymbol{H}\boldsymbol{x}_i\boldsymbol{x}_i^t\boldsymbol{H}^t\boldsymbol{\Psi}_i\boldsymbol{H}\boldsymbol{S}^{-1}\right\}\left\{\mathrm{tr}\,\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right\}\right] = \sum_{j=1}^{p}E\left[a_{jj}\ell_j\phi\right]$$
$$= \sum_{j=1}^{p}E\left[(n-p-1)\frac{\ell_j\phi}{\ell_j} + 2\frac{\partial}{\partial\ell_j}(\ell_j\phi) + \sum_{c\neq j}\frac{(\ell_j\phi)-(\ell_c\phi)}{\ell_j-\ell_c}\right]. \tag{3.13}$$

From (2.11) and (3.12), we get

$$2\sum_{j=1}^{p}\ell_j\frac{\partial\phi}{\partial\ell_j} = 2\sum_{j=1}^{p}\sum_{i=1}^{m}\left[-\frac{(\boldsymbol{h}_j^t\boldsymbol{x}_i)^2/\ell_j}{(d_i+\hat\lambda_j^{CB})^2} - 2\frac{(\boldsymbol{h}_j^t\boldsymbol{x}_i)^2}{(d_i+\hat\lambda_j^{CB})^3}\frac{\partial\hat\lambda_j^{CB}}{\partial\ell_j}\right]$$
$$= -2\phi + 4c_0\sum_{j=1}^{p}\frac{\sum_{i=1}^{m}(\boldsymbol{h}_j^t\boldsymbol{x}_i)^2/(d_i+\lambda_j^*)^3}{\sum_{i=1}^{m}(\boldsymbol{h}_j^t\boldsymbol{x}_i)^2/(d_i+\lambda_j^*)^2}I(\lambda_j^*>\lambda_{c0})$$
$$\le -2\phi + 4c_0\sum_{j=1}^{p}(d_m+\hat\lambda_j^{CB})^{-1}. \tag{3.14}$$

Hence,

$$I_3 \le (np+2)c_0\sum_{j=1}^{p}E\left[(d_m+\hat\lambda_j^{CB})^{-1}\right],$$

and from (3.11),

$$\boldsymbol{\Delta} \le \sum_{j=1}^{p}E\left[-2\sum_{i=1}^{m}(d_i+\hat\lambda_j^{CB})^{-1} + \{4+(np+2)c_0\}(d_m+\hat\lambda_j^{CB})^{-1}\right]. \tag{3.15}$$

Since

$$\sum_{i=1}^{m}\frac{d_m+\hat\lambda_j^{CB}}{d_i+\hat\lambda_j^{CB}} \ge \sum_{i=1}^{m}\frac{d_m+\lambda_{c0}}{d_i+\lambda_{c0}} = m - \sum_{i=1}^{m}\frac{d_i-d_m}{d_i+\lambda_{c0}},$$

and $c_0 = (m-2)/(np+2)$, the right hand side of (3.15) is less than zero if

$$2\sum_{i=1}^{m}\frac{d_i-d_m}{d_i+\lambda_{c0}} - 2(m-2) + (m-2) \le 0,$$

which is guaranteed by (2.12). Therefore the proof of Theorem 2 is complete. ■

3.3 Proof of Theorem 3

Let $\boldsymbol{G} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Q}$, $a_i = (d_i+1)/(d_1+1)$, $\boldsymbol{\eta}_i = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}_i/\sqrt{a_i}$. Consider the transformations $\boldsymbol{u}_i = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}_i/\sqrt{a_i}$ and $\boldsymbol{V} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}^{-1/2}$. Then $\boldsymbol{u}_i \sim \mathcal{N}_p(\boldsymbol{\eta}_i, (d_i/a_i)\boldsymbol{I})$ and $\boldsymbol{V} \sim \mathcal{W}_p(\boldsymbol{I}, n)$. From (2.19), $\boldsymbol{V} = (\boldsymbol{G}^t)^{-1}\boldsymbol{G}^{-1}$ and $\boldsymbol{U}^t\boldsymbol{U} = (\boldsymbol{G}^t)^{-1}\boldsymbol{F}\boldsymbol{G}^{-1}$, where $\boldsymbol{U}^t = (\boldsymbol{u}_1, \dots, \boldsymbol{u}_m)$. Let $\boldsymbol{\Phi}_* = \mathrm{diag}\,(\phi_1^*, \dots, \phi_p^*)$ for $\phi_j^* = (d_m+\hat\lambda_0)^{-1} + (d_m+\hat\lambda_j)^{-1}$ where $\hat\lambda_0^{MB}$ and $\hat\lambda_j^{MB}$ are here abbreviated $\hat\lambda_0$ and $\hat\lambda_j$, and $\boldsymbol{\Psi}_* = \mathrm{diag}\,(\psi_1^*, \dots, \psi_p^*)$ for $\psi_j^* = f_j\phi_j^*$. Then it is seen that $\boldsymbol{\Phi}_i \le \boldsymbol{\Phi}_*$ for $i = 1, \dots, m$, since $d_m = \min_i\{d_i\}$.

To prove the theorem, we calculate the difference in the risks of the estimators $\widehat{\boldsymbol{\Theta}}^{MB}$ and $\boldsymbol{X}$ relative to the loss (2.4) is given by

$$\Delta = R(\omega, \widehat{\boldsymbol{\Theta}}^{MB}) - R(\omega, \boldsymbol{X}) = -2I_1 + I_2, \tag{3.16}$$

where, since $\boldsymbol{G}^t = \boldsymbol{G}^{-1}(\boldsymbol{G}\boldsymbol{G}^t) = \boldsymbol{G}^{-1}\boldsymbol{V}^{-1}$,

$$I_1 = \sum_{i=1}^m E\left[a_i d_i^{-1}(\boldsymbol{u}_i - \boldsymbol{\eta}_i)^t(\boldsymbol{G}^t)^{-1}\boldsymbol{\Phi}_i\boldsymbol{G}^{-1}\boldsymbol{V}^{-1}\boldsymbol{u}_i\right], \tag{3.17}$$

and, since $a_i \le 1$ and $\boldsymbol{\Phi}_i \le \boldsymbol{\Phi}_*$,

$$\begin{aligned}
I_2 &= \sum_{i=1}^m E\left[a_i\boldsymbol{u}_i^t\boldsymbol{G}\boldsymbol{\Phi}_i\boldsymbol{G}^{-1}(\boldsymbol{G}^t)^{-1}\boldsymbol{\Phi}_i\boldsymbol{G}^t\boldsymbol{u}_i\right] \\
&\le E\left[\mathrm{tr}\,\boldsymbol{G}\boldsymbol{\Phi}_*\boldsymbol{G}^{-1}(\boldsymbol{G}^t)^{-1}\boldsymbol{\Phi}_*\boldsymbol{G}^t\boldsymbol{U}^t\boldsymbol{U}\right] \\
&= E\left[\mathrm{tr}\,(\boldsymbol{G}^t)^{-1}\boldsymbol{\Phi}_*\boldsymbol{F}\boldsymbol{\Phi}_*\boldsymbol{G}^{-1}\right] = E\left[\mathrm{tr}\,(\boldsymbol{G}^t)^{-1}\boldsymbol{\Psi}_*^2\boldsymbol{F}^{-1}\boldsymbol{G}^{-1}\right] \\
&= E\left[(n-p-1)\mathrm{tr}\,\boldsymbol{\Psi}_*^2\boldsymbol{F}^{-1}\boldsymbol{G}^{-1}\boldsymbol{V}^{-1}(\boldsymbol{G}^t)^{-1} + 2\mathrm{tr}\,\mathcal{D}_V[(\boldsymbol{G}^t)^{-1}\boldsymbol{\Psi}_*^2\boldsymbol{F}^{-1}\boldsymbol{G}^{-1}]\right] \\
&= \sum_{j=1}^p E\left[\frac{1}{f_j}\left\{(n+p+1)(\psi_j^*)^2 - 4f_j\psi_j^*\frac{\partial\psi_j^*}{\partial f_j} - 2f_j\sum_{a>j}\frac{(\psi_j^*)^2 - (\psi_a^*)^2}{f_j - f_a}\right\}\right],
\end{aligned} \tag{3.18}$$

by using the Stein-Haff identity (3.2) and the following result due to Konno (1992):

$$\mathrm{tr}\,\mathcal{D}_V[(\boldsymbol{G}^t)^{-1}\boldsymbol{\Phi}(\boldsymbol{F})\boldsymbol{G}^{-1}] = \sum_{j=1}^p\left\{p\phi_j - f_j\frac{\partial\phi_j}{\partial f_j} - \sum_{c>j}\frac{f_j\phi_j - f_c\phi_c}{f_j - f_c}\right\}.$$

To evaluate I_1, we use some equations on the differential operator. Let $\mathcal{D}_W = (\mathrm{d}_{ij}^W)$, where $\mathrm{d}_{ij}^W = 2^{-1}(1+\delta_{ij})\partial/\partial w_{ij}$ for $\boldsymbol{W} = (w_{ij}) = \boldsymbol{U}^t\boldsymbol{U}$. Then Lo (1988) and Konno (1992) derived the following equations: For a $p \times p$ matricial function $\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{W}, \boldsymbol{V})$,

$$\boldsymbol{\nabla}_i^t\boldsymbol{T} = 2\boldsymbol{u}_i^t\mathcal{D}_W\boldsymbol{T} \tag{3.19}$$

$$\mathrm{d}_{ab}^W f_j = g_{aj}g_{bj} \tag{3.20}$$

$$\mathrm{d}_{ab}^W g^{cd} = \frac{1}{2}\sum_{s\ne c}\frac{g^{sd}(g_{ac}g_{bs} + g_{bc}g_{as})}{f_c - f_s}, \tag{3.21}$$

where $\boldsymbol{G} = (g_{ab})$, $\boldsymbol{G}^{-1} = (g^{ab})$ and $\boldsymbol{\nabla}_i^t = \partial/\partial\boldsymbol{u}_i$. Now, we evaluate I_1 with the help of the

Stein identity (3.1). Using (3.19), we get

$$\begin{aligned}
I_1 =& \sum_{i=1}^{m} E\left[\boldsymbol{\nabla}_i^t[(\boldsymbol{G}^t)^{-1}\boldsymbol{\Phi}_i\boldsymbol{G}^{-1}\boldsymbol{V}^{-1}\boldsymbol{u}_i]\right] \\
=& \sum_i E\left[\left\{\boldsymbol{\nabla}_i^t[(\boldsymbol{G}^t)^{-1}\boldsymbol{\Phi}_i\boldsymbol{G}^{-1}]\boldsymbol{V}^{-1}\right\}\boldsymbol{u}_i\right] + \sum_i E[\mathrm{tr}\,\boldsymbol{\Phi}_i] \\
=& 2\sum_i E\left[\boldsymbol{u}_i^t\mathcal{D}_W[(\boldsymbol{G}^t)^{-1}\boldsymbol{\Phi}_i\boldsymbol{G}^{-1}]\boldsymbol{V}^{-1}\boldsymbol{u}_i\right] + \sum_i E[\mathrm{tr}\,\boldsymbol{\Phi}_i] \\
=& I_{11} + I_{12}, \quad \text{(say)}.
\end{aligned} \tag{3.22}$$

We evaluate I_{11} using (3.20) and (3.21) coordinatewise. Note that $\sum_b g^{cb}g_{bj} = \delta_{cj}$, and

$$\begin{aligned}
&\left(\mathcal{D}_W[(\boldsymbol{G}^t)^{-1}\boldsymbol{\Phi}_i\boldsymbol{G}^{-1}]\right)_{a,d} \\
=& \sum_{b,c}(\mathrm{d}_{ab}^W g^{cb})\phi_c^{(i)}g^{cd} + \sum_{b,c} g^{cb}(\mathrm{d}_{ab}^W\phi_c^{(i)})g^{cd} + \sum_{b,c} g^{cb}\phi_c^{(i)}(\mathrm{d}_{ab}^W g^{cd}).
\end{aligned} \tag{3.23}$$

Since $\mathrm{d}_{ab}^W\phi_c^{(i)} = \sum_j(\mathrm{d}_{ab}^W f_j)\partial\phi_c^{(i)}/\partial f_j = \sum_j g_{aj}g_{bj}\partial\phi_c^{(i)}/\partial f_j$, we observe that

$$\begin{aligned}
\sum_{b,c} g^{cb}(\mathrm{d}_{ab}^W\phi_c^{(i)})g^{cd} =& \sum_{b,c,j} g^{cb}g_{aj}g_{bj}g^{cd}\partial\phi_c^{(i)}/\partial f_j \\
=& \sum_c g_{ac}(\partial\phi_c^{(i)}/\partial f_c)g^{cd},
\end{aligned} \tag{3.24}$$

Similarly, we obtain that

$$\begin{aligned}
\sum_{b,c}(\mathrm{d}_{ab}^W g^{cb})\phi_c^{(i)}g^{cd} =& \frac{1}{2}\sum_{b,c}\phi_c^{(i)}g^{cd}\sum_{s\neq c}\frac{g^{sb}(g_{ac}g_{bs}+g_{bc}g_{as})}{f_c-f_s} \\
=& \frac{1}{2}\sum_c g_{ac}\left(\sum_{s\neq c}\frac{\phi_c^{(i)}}{f_c-f_s}\right)g^{cd}
\end{aligned} \tag{3.25}$$

$$\sum_{b,c} g^{cb}\phi_c^{(i)}(\mathrm{d}_{ab}^W g^{cd}) = \frac{1}{2}\sum_s g_{as}\left(\sum_{c\neq s}\frac{\phi_c^{(i)}}{f_c-f_s}\right)g^{sd}. \tag{3.26}$$

Combining (3.23), (3.24), (3.25) and (3.26) gives that

$$\left(\mathcal{D}_W[(\boldsymbol{G}^t)^{-1}\boldsymbol{\Phi}_i\boldsymbol{G}^{-1}]\right)_{a,b} = \sum_c g_{ac}\left\{\frac{\partial\phi_c^{(i)}}{\partial f_c} + \frac{1}{2}\sum_{s\neq c}\frac{\phi_c^{(i)}-\phi_s^{(i)}}{f_c-f_s}\right\}g^{cb},$$

which is written in the matricial form as

$$\mathcal{D}_W[(\boldsymbol{G}^t)^{-1}\boldsymbol{\Phi}_i\boldsymbol{G}^{-1}] = \boldsymbol{G}\boldsymbol{\Phi}_i^{(1)}\boldsymbol{G}^{-1} \tag{3.27}$$

where $\boldsymbol{\Phi}_i^{(1)} = \mathrm{diag}\,(\phi_{1,i}^{(1)},\ldots,\phi_{p,i}^{(1)})$ for

$$\phi_{j,i}^{(1)} = \frac{\partial\phi_j^{(i)}}{\partial f_j} + \frac{1}{2}\sum_{a\neq j}\frac{\phi_j^{(i)}-\phi_a^{(i)}}{f_j-f_a}.$$

Note that the partial derivative of $\phi_j^{(i)}$, given by (2.23), is evaluated by

$$\begin{aligned}\frac{\partial\phi_j^{(i)}}{\partial f_j} &= -\frac{c_0}{(d_i+\hat\lambda_0)^2}I(c_0\mathrm{tr}\,\boldsymbol{F}>\lambda_{m0}) - \frac{c_1}{(d_i+\hat\lambda_j)^2}I(c_1 f_j>\lambda_{m1})\\ &\geq -\frac{c_0}{(d_m+\hat\lambda_0)^2}I(c_0\mathrm{tr}\,\boldsymbol{F}>\lambda_{m0}) - \frac{c_1}{(d_m+\hat\lambda_j)^2}I(c_1 f_j>\lambda_{m1}) = \frac{\partial\phi_j^*}{\partial f_j}. \end{aligned} \tag{3.28}$$

Since $(\hat\lambda_j-\hat\lambda_a)/(f_j-f_a)\geq 0$, we get the inequality

$$\begin{aligned}\frac{\phi_j^{(i)}-\phi_a^{(i)}}{f_j-f_a} &= -\frac{(\hat\lambda_j-\hat\lambda_a)/(f_j-f_a)}{(d_i+\hat\lambda_j)(d_i+\hat\lambda_a)}\\ &\geq -\frac{(\hat\lambda_j-\hat\lambda_a)/(f_j-f_a)}{(d_m+\hat\lambda_j)(d_m+\hat\lambda_a)} = \frac{\phi_j^*-\phi_a^*}{f_j-f_a}.\end{aligned} \tag{3.29}$$

Let $\boldsymbol{\Phi}_*^{(1)}=\mathrm{diag}\,(\phi_{1*}^{(1)},\ldots,\phi_{p*}^{(1)})$ for $\phi_{j*}^{(1)}=\partial\phi_j^*/\partial f_j+2^{-1}\sum_{a\neq j}(\phi_j^*-\phi_a^*)/(f_j-f_a)$. Then from the inequalities (3.28) and (3.29), we observe that

$$\begin{aligned}I_{11} =&2\sum_i E[\boldsymbol{u}_i^t\boldsymbol{G}\boldsymbol{\Phi}_i^{(1)}\boldsymbol{G}^{-1}\boldsymbol{V}^{-1}\boldsymbol{u}_i] = 2\sum_i E[\boldsymbol{u}_i^t\boldsymbol{G}\boldsymbol{\Phi}_i^{(1)}\boldsymbol{G}^t\boldsymbol{u}_i]\\ \geq& 2\sum_i E[\boldsymbol{u}_i^t\boldsymbol{G}\boldsymbol{\Phi}_*^{(1)}\boldsymbol{G}^t\boldsymbol{u}_i] = 2E[\mathrm{tr}\,\boldsymbol{G}\boldsymbol{\Phi}_*^{(1)}\boldsymbol{G}^t\boldsymbol{U}^t\boldsymbol{U}] = 2E[\mathrm{tr}\,\boldsymbol{\Phi}_*^{(1)}\boldsymbol{F}],\end{aligned}$$

which, from (3.22), implies that

$$I_1 \geq \sum_{j=1}^p E\left[\frac{1}{f_j}\left\{\sum_i\psi_j^{(i)}+2f_j\frac{\partial\psi_j^*}{\partial f_j}-2\psi_j^*+f_j^2\sum_{a\neq j}\frac{\psi_j^*/f_j-\psi_a^*/f_a}{f_j-f_a}\right\}\right]. \tag{3.30}$$

It is here noted that

$$\sum_j f_j\sum_{a\neq j}\frac{\psi_j^*/f_j-\psi_a^*/f_a}{f_j-f_a} = -(p-1)\sum_j\frac{\psi_j^*}{f_j}+\sum_j\sum_{a\neq j}\frac{\psi_j^*-\psi_a^*}{f_j-f_a}.$$

Then, combining (3.16), (3.17) and (3.30) gives that

$$\begin{aligned}\Delta \leq& \sum_{j=1}^p E\Big[\frac{1}{f_j}\Big\{(n+p+1)(\psi_j^*)^2-4f_j\psi_j^*\frac{\partial\psi_j^*}{\partial f_j}-2f_j\sum_{a>j}\frac{(\psi_j^*)^2-(\psi_a^*)^2}{f_j-f_a}\\ &-2\sum_{i=1}^m\psi_j^{(i)}+2(p+1)\psi_j^*-4f_j\frac{\partial\psi_j^*}{\partial f_j}-4f_j\sum_{a>j}\frac{\psi_j^*-\psi_a^*}{f_j-f_a}\Big\}\Big]\\ \leq& \sum_{j=1}^p E\Big[\frac{1}{f_j}\Big\{(n+p+1)(\psi_j^*)^2-2\sum_{i=1}^m\psi_j^{(i)}+2(p+1)\psi_j^*\\ &-4f_j\frac{\partial\psi_j^*}{\partial f_j}-2f_j\sum_{a>j}\frac{\psi_j^*-\psi_a^*}{f_j-f_a}(\psi_j^*+\psi_a^*+2)\Big\}\Big],\end{aligned} \tag{3.31}$$

since $\partial\psi_j^*/\partial f_j\geq 0$. Noting that $\psi_j^*-\psi_a^*\geq(f_j-f_a)/(d_m+\hat\lambda_0)$ for $a>j$, we observe that

$$\begin{aligned}&\sum_{j=1}^p\sum_{a>j}\frac{\psi_j^*-\psi_a^*}{f_j-f_a}(\psi_j^*+\psi_a^*+2)\geq\frac{1}{d_m+\hat\lambda_0}\sum_{j=1}^p\sum_{a>j}(\psi_j^*+\psi_a^*+2)\\ =&\frac{1}{d_m+\hat\lambda_0}\left\{(p-1)\sum_{j=1}^p\psi_j^*+(p-1)p\right\}\geq\frac{(p-1)\mathrm{tr}\,\boldsymbol{F}}{(d_m+\hat\lambda_0)^2}+\frac{(p-1)p}{d_m+\hat\lambda_0},\end{aligned} \tag{3.32}$$

where we used the equations $\sum_j \sum_{a>j} \psi_j^* = \sum_j (p-j)\psi_j^*$, $\sum_j \sum_{a>j} \psi_a^* = \sum_j (j-1)\psi_j^*$ and $\sum_j \sum_{a>j} 1 = (p-1)p/2$. Also note that the partial derivative of ψ_j^* can be evaluated as

$$\begin{aligned}\sum_{j=1}^p \frac{\partial \psi_j^*}{\partial f_j} =& \sum_{j=1}^p \left\{ \frac{1}{d_m+\hat{\lambda}_0} - \frac{c_0 f_j}{(d_m+\hat{\lambda}_0)^2} I(c_0 \mathrm{tr}\, \boldsymbol{F} > \lambda_{m0}) \right\} \\ \geq& \frac{p}{d_m+\hat{\lambda}_0} - \frac{c_0 \mathrm{tr}\, \boldsymbol{F}}{(d_m+\hat{\lambda}_0)^2} \geq \frac{p-1}{d_m+\hat{\lambda}_0},\end{aligned} \tag{3.33}$$

since $c_0 \mathrm{tr}\, \boldsymbol{F} \leq \hat{\lambda}_0 \leq d_m + \hat{\lambda}_0$. Using the inequalities (3.32) and (3.33), the r.h.s. in (3.31) can be further evaluated as

$$\begin{aligned}\Delta \leq& E\Big[\sum_j \frac{1}{f_j} \left\{ (n+p+1)(\psi_j^*)^2 - 2\sum_{i=1}^m \psi_j^{(i)} + 2(p+1)\psi_j^* \right\} \\ &- 2\frac{(p-1)\mathrm{tr}\, \boldsymbol{F}}{(d_m+\hat{\lambda}_0)^2} - 2\frac{(p-1)(p+2)}{d_m+\hat{\lambda}_0} \Big] \\ =& E[\Delta^*], \quad \text{(say)}.\end{aligned} \tag{3.34}$$

Finally, we shall show that Δ^* is not positive. Noting that

$$\sum_j \frac{(\psi_j^*)^2}{f_j} = \frac{\mathrm{tr}\, \boldsymbol{F}}{(d_m+\hat{\lambda}_0)^2} + \frac{2}{d_m+\hat{\lambda}_0} \sum_j \frac{f_j}{d_m+\hat{\lambda}_j} + \sum_j \frac{f_j}{(d_m+\hat{\lambda}_j)^2},$$

we see that Δ^* can be rewritten as $\Delta^* = \Delta_1^* + \Delta_2^*$ where

$$\begin{aligned}\Delta_1^* =& \sum_j \frac{1}{d_m+\hat{\lambda}_j} \left\{ \frac{(n+p+1)f_j}{d_m+\hat{\lambda}_j} - 2\sum_{i=1}^m \frac{d_m+\hat{\lambda}_j}{d_i+\hat{\lambda}_j} + 2(p+1) \right\}, \\ \Delta_2^* =& (n-p+3)\frac{\mathrm{tr}\, \boldsymbol{F}}{(d_m+\hat{\lambda}_0)^2} + 2\frac{n+p+1}{d_m+\hat{\lambda}_0} \sum_j \frac{f_j}{d_m+\hat{\lambda}_j} - 2\sum_{i=1}^m \frac{p}{d_i+\hat{\lambda}_0} \\ &+ 2\frac{p(p+1)}{d_m+\hat{\lambda}_0} - 2\frac{(p-1)(p+2)}{d_m+\hat{\lambda}_0}.\end{aligned}$$

For Δ_1^*, it is noted that $(n+p+1)f_j/(d_m+\hat{\lambda}_j) \leq (n+p+1)/c_1 = m-p-1$, and that $\sum_{i=1}^m (d_m+\hat{\lambda}_j)/(d_i+\hat{\lambda}_j) \geq \sum_{i=1}^m (d_m+\lambda_{m1})/(d_i+\lambda_{m1})$ since $\hat{\lambda}_j \geq \lambda_{m1}$. Hence, the inequality that $\Delta_2^* \leq 0$ is established if λ_{m1} satisfies the inequality

$$m-p-1-2\sum_{i=1}^m (d_m+\lambda_{m1})/(d_i+\lambda_{m1}) + 2(p+1) \leq 0$$

or

$$\sum_{i=1}^m (d_i-d_m)/(d_i+\lambda_{m1}) \leq (m-p-1)/2,$$

which is quaranteed by (2.21). For Δ_2^*, the same arguments are used to show that

$$\begin{aligned}&(d_m+\hat{\lambda}_0)\Delta_2^* \\ &\leq \frac{n-p+3}{c_0} + 2\frac{(n+p+1)p}{c_1} - 2p\sum_{i=1}^m \frac{d_m+\hat{\lambda}_0}{d_i+\hat{\lambda}_0} + 2p(p+1) - 2(p-1)(p+2) \\ &\leq -(p-1)(p+2) + 2mp - 2p\sum_{i=1}^m \frac{d_m+\hat{\lambda}_0}{d_i+\hat{\lambda}_0}.\end{aligned}$$

Hence, $\Delta_2^* \leq 0$ if λ_{m0} satisfies the inequality

$$\sum_{i=1}^{m}(d_i - d_m)/(d_i + \lambda_{m0}) \leq (p-1)(p+2)/(2p),$$

which is guaranteed by (2.20). Therefore the proof of Theorem 3 is complete. ■

4 Simulation and Empirical Studies

Now we investigate the risk-performances of estimators of $\mathbf{\Theta}$ numerically. The estimators we want to investigate are the least squares estimator $\boldsymbol{X}$ and the proposed estimators $\widehat{\mathbf{\Theta}}^{SB}$, $\widehat{\mathbf{\Theta}}^{CB}$, $\widehat{\mathbf{\Theta}}^{CC}$, $\widehat{\mathbf{\Theta}}^{MB}$ and $\widehat{\mathbf{\Theta}}^{MC}$, which are denoted by LS, SB, CB, CC,MB and MC, respectively, where we put $c = 5$ for the constant c in the estimators $\widehat{\mathbf{\Theta}}^{CC}$ and $\widehat{\mathbf{\Theta}}^{MC}$. The principal component regression estimators PC_1 and PC_3 are also treated where PC_1 is obtained by deleting the eigenvectors corresponding to the largest eigenvalue of $(\boldsymbol{Z}^t\boldsymbol{Z})^{-1}$ and PC_3 corresponds to the one obtained by deleting the three largest eigenvalues.

Srivastava and Solanky (2003) showed numerically that the estimator proposed by Konno (1991) is better than the LS estimator in the multicollinearity case. We thus treat the Konno's estimator, denoted by KS, for numerical comparison of estimators. Let $\widetilde{\boldsymbol{Q}}$ be a $p \times p$ nonsingular matrix such that $\widetilde{\boldsymbol{Q}}^t\boldsymbol{S}\widetilde{\boldsymbol{Q}} = \boldsymbol{I}_p$ and $\widetilde{\boldsymbol{Q}}^t\boldsymbol{X}^t\boldsymbol{D}^{-1}\boldsymbol{X}\widetilde{\boldsymbol{Q}} = \widetilde{\boldsymbol{F}} = \mathrm{diag}\,(\tilde{f}_1, \ldots, \tilde{f}_p)$. Then the Konno's estimator is given by $\widehat{\mathbf{\Theta}}^{KS} = (\widehat{\boldsymbol{\theta}}_1^{KS}, \ldots, \widehat{\boldsymbol{\theta}}_m^{KS})^t$ with

$$\widehat{\boldsymbol{\theta}}_i^{KS} = \boldsymbol{x}_i - (\widetilde{\boldsymbol{Q}}^t)^{-1}\mathbf{\Phi}^{KS}(\widetilde{\boldsymbol{F}})\widetilde{\boldsymbol{Q}}^t\boldsymbol{x}_i, \quad i = 1, \ldots, m,$$

where $\mathbf{\Phi}^{KS}(\widetilde{\boldsymbol{F}}) = \mathrm{diag}\,(\phi_1^{KS}, \ldots, \phi_p^{KS})$ for

$$\phi_j^{KS} = \min\left\{\frac{m+p-2j-1}{n-p+2j+1}\frac{1}{\tilde{f}_j}, 1\right\}.$$

Every estimator $\boldsymbol{\delta}$ is evaluated by three types of risk functions $R_j(\omega, \widetilde{\mathbf{\Theta}})$ under the loss functions $L_j(\omega, \widetilde{\mathbf{\Theta}}, \boldsymbol{D}^{-j}) = \mathrm{tr}\,(\widetilde{\mathbf{\Theta}} - \mathbf{\Theta})\mathbf{\Sigma}^{-1}(\widetilde{\mathbf{\Theta}} - \mathbf{\Theta})^t\boldsymbol{D}^{-j}$, called the L_j-loss, for $j = 0, 1, 2$. The risk functions of the above estimators and the LS estimator $\boldsymbol{X}$ are obtained from 1,000 replications through simulation experiments, and the relative efficiencies $R_j(\omega, \widetilde{\mathbf{\Theta}})/R_j(\omega, \boldsymbol{X})$, $j = 0, 1, 2$, of estimator $\widetilde{\mathbf{\Theta}}$ over $\boldsymbol{X}$ are reported. The simulation experiments are done in the following two cases:

Case 1: $p = 6$, $m = 22$, $n = 34$, $\theta_{ij} = 5(i + j/2) \times \eta$, $i = 1, \ldots, m$, $j = 1, \ldots, p$, and $\boldsymbol{D} = \mathrm{diag}\,(125.5, 94.03, 64.65, 39.79, 11.65, 6.238, 3.909, 2.325, 1.209, 0.9182,\ 0.4770,$ $0.4371,\ 0.2619,\ 0.2081,\ 0.1284,\ 0.06062,\ 0.05171,\ 0.02218,\ 0.02085,\ 0.005219,\ 0.003795,$ $0.001601)$.

Case 2: $p = 3$, $m = 10$, $n = 30$, $\theta_{ij} = (m - i + 1 + (p - j + 1)/3) \times \eta$, $i = 1, \ldots, m$, $j = 1, \ldots, p$, and $\boldsymbol{D} = \mathrm{diag}\,(700, 500, 300, 10, 5, 2, 1, 0.1, 0.01, 0.001)$.

The values of the parameters in Case 1 correspond to those in Example 1 given below. The relative efficiencies of the above estimators for the two cases are given in Tables 1 and 2, respectively. Form these tables, the following conclusions can be drawn.

(1) The empirical Bayes ridge regression estimators SB, CC and MC have very nice risk behaviors for L_0- and L_1- losses; they are highly recommended in the case of multicollinearity. Although CB has a slightly larger risk than SB, the risk performance of CB

Table 1: Relative Efficiencies of the Estimators under L_0, L_1, L_2 Losses for $\boldsymbol{D}$ = diag (125.5, 94.03, 64.65, 39.79, 11.65, 6.238, 3.909, 2.325, 1.209, 0.9182, 0.4770, 0.4371, 0.2619, 0.2081, 0.1284, 0.06062, 0.05171, 0.02218, 0.02085, 0.005219, 0.003795, 0.001601), $p = 6$, $m = 22$, $n = 34$ and $\theta_{ij} = 5(i + j/2) \times \eta$, $i = 1, \ldots, m$, $j = 1, \ldots, p$.

	η	SB	CB	CC	MB	MC	KS	PC_1	PC_3
L_0	0	0.003	0.295	0.054	0.214	0.059	0.138	0.644	0.192
	1	0.030	0.346	0.079	0.421	0.091	0.306	0.644	0.192
	2	0.083	0.429	0.130	0.503	0.139	0.319	0.652	0.218
	3	0.148	0.504	0.195	0.528	0.197	0.338	0.662	0.251
	4	0.222	0.566	0.266	0.540	0.260	0.359	0.676	0.296
L_1	0	0.409	0.785	0.697	0.824	0.740	0.140	0.955	0.864
	1	0.626	0.828	0.745	0.875	0.771	0.346	0.955	0.865
	2	0.728	0.861	0.785	0.892	0.799	0.377	0.955	0.865
	3	0.782	0.884	0.814	0.896	0.821	0.413	0.955	0.865
	4	0.817	0.901	0.838	0.900	0.840	0.441	0.959	0.883
L_2	0	0.969	0.997	0.996	0.999	0.999	0.138	0.999	0.999
	1	0.995	0.998	0.998	0.999	0.999	0.557	0.999	0.999
	2	0.998	0.999	0.999	0.999	0.999	0.613	0.999	0.999
	3	0.999	0.999	0.999	0.999	0.999	0.668	0.999	0.999
	4	0.999	0.999	0.999	0.999	0.999	0.703	0.999	0.999

Table 2: Relative Efficiencies of the Estimators under L_0, L_1, L_2 Losses for $\boldsymbol{D}$ = diag (700, 500, 300, 10, 5, 2, 1, 0.1, 0.01, 0.001), $p = 3$, $m = 10$, $n = 30$ and $\theta_{ij} = (m - i + 1 + (p - j + 1)/3) \times \eta$, $i = 1, \ldots, m$, $j = 1, \ldots, p$.

	η	SB	CB	CC	MB	MC	KS	PC_1	PC_3
L_0	0	0.003	0.242	0.023	0.225	0.027	0.183	0.552	0.011
	1	0.181	0.379	0.177	0.573	0.173	0.554	0.627	0.200
	2	0.514	0.555	0.463	0.623	0.432	0.658	0.855	0.765
	3	0.780	0.669	0.694	0.640	0.651	0.701	1.235	1.707
	4	0.928	0.739	0.831	0.654	0.784	0.720	1.766	3.025
L_1	0	0.452	0.696	0.619	0.766	0.685	0.181	0.902	0.697
	1	0.746	0.794	0.729	0.873	0.744	0.687	0.919	0.758
	2	0.854	0.857	0.827	0.888	0.826	0.768	0.968	0.941
	3	0.932	0.895	0.899	0.893	0.891	0.785	1.051	1.244
	4	0.974	0.917	0.941	0.898	0.930	0.791	1.166	1.670
L_2	0	0.998	0.999	0.999	0.999	0.999	0.176	1.000	0.999
	1	0.999	0.999	0.999	0.999	0.999	1.095	1.000	1.000
	2	0.999	0.999	0.999	0.999	0.999	1.095	1.000	1.000
	3	0.999	0.999	0.999	0.999	0.999	1.051	1.000	1.000
	4	0.999	0.999	0.999	0.999	0.999	1.028	1.000	1.000

is not bad. The matricial shrinkage estimator MB is not good in comparison with the other procedures.

(2) Konno (1991) showed the minimaxity of the estimator KS under the L_1-loss. Both tables reveal that KS is not only the best under the L_1-loss, but also behaves well relative to the L_0- and L_2- losses. This implies that the risk behaviors of KS are nice in the multicollinearity, although it is not ridge-type.

(3) Although the minimaxity of the proposed estimators are guaranteed under the L_2-loss, their risk performances are much better than the LS estimator under L_0- and L_1-loss functions.

(4) Through the tables, we see that the principal component regression estimators PC_1 and PC_3 have smaller risks for smaller values of $\mathrm{tr}\,\boldsymbol{\Theta}\boldsymbol{\Theta}^t$ and gets larger as $\mathrm{tr}\,\boldsymbol{\Theta}\boldsymbol{\Theta}^t$ increases.

We shall provide an empirical study for a set of data.

Example 1. (*Chemometrics Data*) We consider the chemometrics data analyzed by Skagerberg, MacGregor and Kiparissides (1992), Breiman and Friedman (1977) and Reinsel (1999), and Srivastava and Solanky (2003). The data were obtained from simulation of a low density tubular polyethylene reactor, and consisted of $N = 56$ observations on the $p = 6$ response variables and $m = 22$ predictor variables (temperatures); the data can be also be found in Srivastava (2002, pp 13-17). The responses are output characteristics of the polymers produced: y_1 (the number-average molecular weight), y_2 (the weight-average molecular weight), y_3 (the frequency of long chain branching), y_4 (the frequency of short chain branching), y_5 (the content of vinyl groups), y_6 (the content of vinylidene groups). Before analyzing the data, all the response variables are transformed by the logarithms and then standardized to unit variance. All the predictor variables are also standardized. As indicated by Breiman and Friedman (1997), the covariance matrix of $\boldsymbol{y}$ is

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.0000 & 0.9566 & 0.0650 & 0.2543 & 0.2551 & 0.2592 \\ 0.9566 & 1.0000 & -0.1284 & 0.2825 & 0.2655 & 0.2755 \\ 0.0650 & -0.1284 & 1.0000 & -0.4997 & -0.4839 & -0.4787 \\ 0.2543 & 0.2825 & -0.4997 & 1.0000 & 0.9744 & 0.9782 \\ 0.2551 & 0.2655 & -0.4839 & 0.9744 & 1.0000 & 0.9760 \\ 0.2592 & 0.2755 & -0.4787 & 0.9782 & 0.9760 & 1.0000 \end{pmatrix},$$

which indicates strong correlation between y_1 and y_2, and also between y_4, y_5 and y_6.

The eigenvalues of the matrix $(\boldsymbol{Z}^t\boldsymbol{Z})^{-1}$ are given by

$$\begin{aligned}\boldsymbol{D} =&(125.5, 94.03, 64.65, 39.79, 11.65, 6.238, 3.909, 2.325,\\ &1.209, 0.9182, 0.4770, 0.4371, 0.2619, 0.2081, 0.1284,\\ &0.06062, 0.05171, 0.02218, 0.02085, 0.005219, 0.003795, 0.001601),\end{aligned}$$

which means that the problem is highly ill-conditioned. We shall investigate how the proposed ridge-type regression estimators of the coefficients $\boldsymbol{\beta}$ behave for the ill-conditioned data. The estimators we treat are the least squares LS, the empirical Bayes ridge regression SB, CB, CC, MB and MC, the principal component regression estimator PC_3 which deletes the eigenvectors corresponding to the three largest eigenvalues. The solutions of the equations defined in Section 2 are given by $\lambda_{s0} = 0.536$, $\lambda_{c0} = 0.791$, $\lambda_{m0} = 35.693$, $\lambda_{m1} = 2.731$, $\lambda^* = 18.009$ and $(\lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, \lambda_5^*, \lambda_6^*) = (386.09, 287.72, 344.32, 87.02, 55.59, 229.22)$,

Table 3: Estimates of $\theta_{1,2}, \dots, \theta_{7,2}$ for the Eight Estimators LS, SB, CB, CC, MB, MC, SK and PC_3

	d_i	LS	SB	CB	CC	MB	MC	KS	PC_3
$\theta_{1,2}$	125	-1.503	-0.188	-1.135	-0.346	-1.314	-0.376	-0.913	0.000
$\theta_{2,2}$	94.0	-4.231	-0.680	-3.504	-1.275	-3.872	-1.353	-3.094	0.000
$\theta_{3,2}$	64.6	-0.386	-0.084	-0.267	-0.135	-0.334	-0.154	0.212	0.000
$\theta_{4,2}$	39.7	4.246	1.323	3.706	2.245	4.074	2.388	3.282	4.246
$\theta_{5,2}$	11.6	-1.847	-1.121	-1.790	-1.578	-1.822	-1.599	-1.164	-1.847
$\theta_{6,2}$	6.23	-2.585	-1.920	-2.515	-2.397	-2.577	-2.447	-2.127	-2.585
$\theta_{7,2}$	3.90	-2.071	-1.702	-2.027	-1.983	-2.069	-2.020	-1.959	-2.071

Table 4: Estimates of prediction errors for the Eight Estimators LS, SB, CB, CC, MB, MC, KS and PC_3

Responses	LS	SB	CB	CC	MB	MC	KS	PC_3
y_1	0.304	0.122	0.228	0.132	0.242	0.134	0.298	0.111
y_2	0.575	0.249	0.477	0.290	0.491	0.295	0.502	0.264
y_3	0.212	0.203	0.202	0.198	0.205	0.199	0.203	0.205
y_4	0.098	0.157	0.094	0.114	0.095	0.111	0.092	0.095
y_5	0.210	0.223	0.204	0.199	0.204	0.200	0.177	0.188
y_6	0.150	0.184	0.145	0.150	0.148	0.150	0.133	0.162
Average	0.258	0.190	0.225	0.180	0.231	0.181	0.234	0.171

which provide $\hat{\lambda}^{SB} = 18.009$ and $\hat{\lambda}_j^{CB} = \lambda_j^*$ for $j = 1, \dots, 6$. Also $(f_1, f_2, f_3, f_4, f_5, f_6)$ is given by $(890, 291, 106, 50, 25, 19)$, which yields $\hat{\lambda}_0^{MB} = 1624$ and $(\hat{\lambda}_1^{MB}, \hat{\lambda}_2^{MB}, \hat{\lambda}_3^{MB}, \hat{\lambda}_4^{MB}, \hat{\lambda}_5^{MB}, \hat{\lambda}_6^{MB}) = (3385, 1107, 403, 189, 94, 73)$. Table 3 gives estimates of the components $\theta_{1,2}, \dots, \theta_{7,2}$ of $\boldsymbol{\theta}_{(2)}$ in the canonical model with $\boldsymbol{\Theta} = (\boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(2)}, \dots, \boldsymbol{\theta}_{(6)}) = \boldsymbol{H\beta}$ and it explains how the proposed procedures work in the presence of the large eigenvalues of $(\boldsymbol{Z}^t\boldsymbol{Z})^{-1}$. The tabel reveals that the estimates by SB, CC and MC gets more shrunken for larger d_i, but CB, MB and KS are less shrunken.

The primary purpose of regression models may be prediction with the help of many independent variables, and the predictors constructed by the ridge-type estimators proposed in this paper are anticipated to have good performances. The prediction error of the methods considered may be estimated via the leave-one-out cross-validation as described in Srivastava (2002, p322). That is, 56 predictive errors are obtained by leaving out one observation each time. Table 4 shows the squared prediction errors estimates (PEE) for the above considered estimators, where the last row indicates the estimates of the average prediction errors. It reveals that the use of the proposed empirical Bayes estimators and the principal component estimator PC_3 provides smaller PEE than the least squares estimator (LS). Of these, SB, CC, MC and PC_3 give much smaller PEE. One weak point of SB is that it shrinks LS with the same shrinkage functions based on $\hat{\lambda}^{SB}$. This is why the scalar shrinkage estimator SB has larger PEE for y_4 and y_6 than LS although it has much smaller average (or total) PEE. From the prediction view point, the principal component regression estimator PC_3 seems the most appropriate in this

Table 5: Estimates of prediction errors for the Eight Estimators LS, SB, CB, CC, MB, MC, KS and PC_4 when the data are given without standardizing the predictor variables except for z_{21} and z_{22}

Responses	LS	SB	CB	CC	MB	MC	KS	PC_4
y_1	0.562	0.121	0.401	0.161	0.468	0.168	0.557	0.120
y_2	1.120	0.281	0.882	0.389	0.954	0.397	0.929	0.312
y_3	0.251	0.212	0.223	0.207	0.235	0.208	0.237	0.213
y_4	0.121	0.150	0.101	0.106	0.112	0.105	0.109	0.106
y_5	0.275	0.235	0.254	0.229	0.264	0.231	0.218	0.260
y_6	0.185	0.187	0.173	0.174	0.182	0.175	0.158	0.210
Average	0.419	0.198	0.339	0.211	0.369	0.214	0.368	0.204

example, although it has a larger PEE for y_6.

This story slightly changes when we treat the data without standardizing the predictor variables $z_1, \ldots, z_{20}$ except for z_{21} and z_{22}. The prediction-error estimates in this case are given in Table 5, which reveals that SB, CC, MC and PC_4 provide much smaller average PEE, and that the average PEE of SB is the smallest. The combined estimators CC and MC provide smaller PEE than LS in the sense of minimizing the PEE for all the responses as well as minimizing the average PEE. In this case, CC and MC seem appropriate. ■

5 Concluding Remarks

From the simulation results, it appears that the scalar Bayes estimator SB and the Konno estimator KS are performing much better than any other estimator, although the combination componentwise estimator CC and the combination matricial estimator MC are also very close to them. However in the combination estimators a choice of 'c' has to be made. It is very likely that a proper choice of the value of c may make them superior to SB and KS.

The numerical example confirms this fact although in this case the pricipal component estimator is also doing well, but a proper choice of the number of components may be required. For a straight forward application without resorting to heavy computation, it seems that the SB estimator may be the preferred estimator.

We conclude the paper with the note that the results on minimaxity given in Section 2 can be extended to elliptically contoured distributions using the arguments as in Kubokawa and Srivastava (2001).

Acknowledgements. We are grateful to two reviewers for their valuable comments. The research of the first author was supported in part by Natural Sciences and Engineering Research Council of Canada. The research of the second author was supported in part by grants from the Ministry of Education, Japan, Nos. 13680371, 15200021 and 15200022 and in part by a grant from COE-Economics, University of Tokyo. This work was done during the visit of the second author to the University of Toronto, 2001 and 2002, summer.

REFERENCES

Bilodeau, M. and Kariya, T. (1989). Minimax estimators in the normal MANOVA model. *J. Multivariate Anal.*, **28**, 260-270.

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

Brown, P.J. and Zidek, J.V. (1980). Adaptive Multivariate Ridge Regression. *Ann. Statist.*, **8**, 64-74.

Brown, P.J. and Zidek, J.V. (1982). Multivariate Regression Shrinkage Estimators with Unknown Covariance Matrix. *Scand. J. Statist.*, **9**, 209-215.

Efron, B. and Morris, C. (1976). Multivariate empirical Bayes estimation of covariance matrices. *Ann. Statist.*, **4**, 22-32.

Haff, L.R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, **9**, 531-542.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: biased estimation for nonorthoganal problems. *Technometrics*, **12**, 55-67.

Konno, Y. (1990). Families of minimax estimators of matrix of normal means with unknown covariance matrix. *J. Japan Statist. Soc.*, **20**, 191-201.

Konno, Y. (1991). On estimation of a matrix of normal means with unknown covariance matrix. *J. Multivariate Anal.*, **36**, 44-55.

Konno, Y. (1992). *Improved estimation of matrix of normal mean and eigenvalues in the multivariate F - distribution.* Doctoral Dissertation, Institute of Mathematics, University of Tsukuba.

Kubokawa, T. and Srivastava, M.S. (2001). Robust improvement in estimation of a mean matrix in an elliptically contoured distribution. *J. Multivariate Anal.*, **76**, 138-152.

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

Reinsel, G.C. (1999). On multivariate linear regression shrinkage and reduced-rank procedures. *J. Statist. Plann. Inf.*, **81**, 311-321.

Skagerberg, B., MacGregor, J. and Kiparissides, C. (1992). Multivariate data analysis applied to low-density polyethylene reactors. *Chemometr. Intell. Lab. Syst.*, **14**, 341-356.

Sen, A. and Srivastava, M. (1990). *Regression Analysis: Theory, Methods, and Applications.* Springer, New York.

Sheena, Y. (1995). Unbiased estimator of risk for an orthogonally invariant estimator of a covariance matrix. *J. Japan Statist. Soc.*, **25**, 35-48.

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

Srivastava, M.S., and Solanky, T.K.S. (2003). Predicting multivariate response in linear regression model. *Comm. Statist. - Simul. Compt.*, **32**, 389-409.

Srivastava, M.S., and Khatri, C.G. (1979). *An Introduction to Multivariate Statistics.* North-Holland, New York.

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

Stein, C. (1977). Lectures on multivariate estimation theory. In *Investigation on Statistical Estimation Theory* I, 4-65. Zapiski Nauchych Seminarov LOMI im. V.A. Steklova AN SSSR vol. 74, Leningrad (in Russian).

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

Strawderman, W.E. (1978). Minimax adaptive generalized ridge regression estimators. *J.Amer. Statist. Assoc.*, **73**, 623-627.