Acta Univ. Palacki. Olomuc., Fac. rer. nat.,
Mathematica **51**, 2 (2012) 73–86

Ridge Estimator Revisited*

Lubomír KUBÁČEK

Department of Mathematical Analysis and Applications of Mathematics
Faculty of Science, Palacký University
17. listopadu 12, 771 46 Olomouc, Czech Republic
e-mail: lubomir.kubacek@upol.cz

(Received March 15, 2011)

Abstract

Bad conditioned matrix of normal equations in connection with small values of model parameters is a source of problems in parameter estimation. One solution gives the ridge estimator. Some modification of it is the aim of the paper. The behaviour of it in models with constraints is investigated as well.

Key words: linear model, ridge estimator, constraints

2000 Mathematics Subject Classification: 62J05

1 Introduction

The linear model $\mathbf{Y} \sim_n (\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ is considered. Here $\mathbf{Y}$ is an n-dimensional random vector (observation vector), $\mathbf{X}\boldsymbol{\beta}$ is the mean value of it, i.e. $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, $\mathbf{X}$ is an $n \times k$ known matrix with the rank $r(\mathbf{X}) = k \leq n$, $\boldsymbol{\beta}$ is an unknown k-dimensional parameter which must be estimated and σ^2 is an unknown parameter $\sigma^2 \in (0, \infty)$.

The best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and its covariance matrix is $\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

Let the spectral decomposition of the matrix $\mathbf{X}'\mathbf{X}$ be

$$\mathbf{X}'\mathbf{X} = \sum_{i=1}^{k} \lambda_i \mathbf{f}_i \mathbf{f}_i' = \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}', \ \mathbf{f}_i'\mathbf{f}_j = \delta_{i,j} \quad \text{(the Kronecker delta)}$$
$$\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_k), \ \boldsymbol{\Lambda} = \mathrm{Diag}(\lambda_1, \dots, \lambda_k).$$

*Supported by the Council of the Czech Government MSM 6 198 959 214

The problem occurs when $\lambda_{\max} = \max\{\lambda_1, \dots, \lambda_k\}$ differs significantly from $\lambda_{\min} = \min\{\lambda_1, \dots, \lambda_k\}$, i.e. $\lambda_{\max}/\lambda_{\min}$ is large number. In this case variances of the BLUEs of different linear functions of $\boldsymbol{\beta}$ can differ significantly as well and it can be in some cases unacceptable.

It seems that the way out this problem is to use either the estimator $\tilde{\boldsymbol{\beta}}$ with the property

$$\mathrm{Var}(\tilde{\boldsymbol{\beta}}) + \left[E(\tilde{\boldsymbol{\beta}}) - \boldsymbol{\beta}\right]\left[E(\tilde{\boldsymbol{\beta}}) - \boldsymbol{\beta}\right]' \leq_L \mathrm{Var}(\widehat{\boldsymbol{\beta}}) \tag{1}$$

($\leq_L$ means the Loevner ordering positive semidefinite matrices), or the property

$$\mathrm{Tr}\left[\mathrm{Var}(\tilde{\boldsymbol{\beta}})\right] + \left[E(\tilde{\boldsymbol{\beta}}) - \boldsymbol{\beta}\right]'\left[E(\tilde{\boldsymbol{\beta}}) - \boldsymbol{\beta}\right] \leq \mathrm{Tr}\left[\mathrm{Var}(\widehat{\boldsymbol{\beta}})\right]. \tag{2}$$

The ridge estimator has the property (2) (see in more detail [3], [4]), however not the property (1).

2 Some comments to the ridge estimator

In the first step let us try to find the estimator of $\boldsymbol{\beta}$ in the form $\mathbf{AY}$, such that

$$\forall\{\mathbf{h} \in R^k\}\ \mathrm{Var}(\mathbf{h}'\mathbf{AY}) + b^2_{A,h} = \min\{\mathrm{Var}(\mathbf{h}'\mathbf{BY}) + b^2_{B,h} \colon \mathbf{B} \in \mathcal{M}^{k\times n}\} \tag{3}$$

where $b_{A,h} = E(\mathbf{h}'\mathbf{AY}) - \mathbf{h}'\boldsymbol{\beta} = \mathbf{h}'(\mathbf{AX} - \mathbf{I})\boldsymbol{\beta}$, $\mathcal{M}^{k\times n}$ is the class of $k \times n$ matrices and $\mathbf{I}$ is the identity matrix.

Lemma 2.1 *The random vector*

$$\boldsymbol{\beta}^* = \boldsymbol{\beta}\boldsymbol{\beta}'(\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' + \sigma^2\mathbf{I})^{-1}\mathbf{Y},$$

satisfies (3).

Proof Let

$$\Phi_h(\mathbf{A}) = \sigma^2\mathbf{h}'\mathbf{AA}'\mathbf{h} + \mathbf{h}'(\mathbf{AX} - \mathbf{I})\boldsymbol{\beta}\boldsymbol{\beta}'(\mathbf{X}'\mathbf{A}' - \mathbf{I})\mathbf{h}.$$

Then (see [2], p. 285)

$$\frac{\partial\Phi_h(\mathbf{A})}{\partial\mathbf{A}} = 2\sigma^2\mathbf{hh}'\mathbf{A} + 2\mathbf{hh}'\mathbf{AX}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' - 2\mathbf{hh}'\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' = \mathbf{0}$$
$$\Rightarrow \mathbf{A} = \boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')^{-1}.$$

thus $\boldsymbol{\beta}^* = \boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')^{-1}\mathbf{Y}$. □

Remark 2.2 The random vector $\boldsymbol{\beta}^*$ has the covariance matrix

$$\mathrm{Var}(\boldsymbol{\beta}^*) = \sigma^2\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')^{-2}\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'$$

and the bias

$$\mathbf{b}_\beta = E(\boldsymbol{\beta}^*) - \boldsymbol{\beta} = \big[\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')^{-1}\mathbf{X} - \mathbf{I}\big]\boldsymbol{\beta}.$$

Thus

$$\mathrm{Var}(\boldsymbol{\beta}^*) + \mathbf{b}_\beta\mathbf{b}_\beta' = \sigma^2\big[\boldsymbol{\beta}'\mathbf{X}'(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')^{-2}\mathbf{X}\boldsymbol{\beta}\big]\boldsymbol{\beta}\boldsymbol{\beta}'$$
$$+\big[\boldsymbol{\beta}'\mathbf{X}'(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')^{-1}\mathbf{X}\boldsymbol{\beta}\big]^2\boldsymbol{\beta}\boldsymbol{\beta}' - 2\big[\boldsymbol{\beta}'\mathbf{X}'(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')^{-1}\mathbf{X}\boldsymbol{\beta}\big]\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\beta}\boldsymbol{\beta}'$$

and

$$\mathrm{Tr}\,\big[\,\mathrm{Var}(\boldsymbol{\beta}^*)\big] + \mathbf{b}_\beta'\mathbf{b}_\beta = \sigma^2\big[\boldsymbol{\beta}'\mathbf{X}'(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')^{-2}\mathbf{X}\boldsymbol{\beta}\big]\boldsymbol{\beta}'\boldsymbol{\beta}$$
$$+\big[\boldsymbol{\beta}'\mathbf{X}'(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')^{-1}\mathbf{X}\boldsymbol{\beta}\big]^2\boldsymbol{\beta}'\boldsymbol{\beta} - 2\big[\boldsymbol{\beta}'\mathbf{X}'(\sigma^2\mathbf{I} + \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')^{-1}\mathbf{X}\boldsymbol{\beta}\big]\boldsymbol{\beta}'\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{\beta}.$$

Since $\boldsymbol{\beta}'\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}')^{-}\mathbf{X}\boldsymbol{\beta} = 1$, it is valid that

$$\sigma^2 = 0 \Rightarrow \mathrm{Tr}\,\big[\,\mathrm{Var}(\tilde{\boldsymbol{\beta}})\big] + \mathbf{b}_\beta'\mathbf{b}_\beta = 0.$$

The vector $\boldsymbol{\beta}^*$ is of no use for an estimation. Even an attempt to use an iteration is useless. If $\boldsymbol{\beta}_0$ is a starting vector in an iteration procedure for a determination of $\boldsymbol{\beta}^*$, then the first step leads to $\boldsymbol{\beta}^*_{(1)} = \boldsymbol{\beta}_0\boldsymbol{\beta}_0'(\sigma^2\mathbf{I}+\mathbf{X}\boldsymbol{\beta}_0\boldsymbol{\beta}_0'\mathbf{X}')^{-1}\mathbf{Y}$. It is valid that $P\{\boldsymbol{\beta}^*_{(1)} \in \mathcal{M}(\boldsymbol{\beta}_0)\} = 1$, since dimension of $\boldsymbol{\beta}_0'(\sigma^2\mathbf{I}+\mathbf{X}\boldsymbol{\beta}_0\boldsymbol{\beta}_0'\mathbf{X}')^{-1}\mathbf{Y}$ is one. An analogous situation occurs in the second and other steps, i.e. $P\{\boldsymbol{\beta}^*_{(i)} \in \mathcal{M}(\boldsymbol{\beta}_0)\} = 1$, $i = 2, 3, \ldots$ Thus the starting vector $\boldsymbol{\beta}_0$ determines the direction of the vector $\boldsymbol{\beta}^*$ and it is not admissible.

Hoerl and Kennard [3], [4] solved the problem in more efficient way. They minimized the function $\phi(\tilde{\boldsymbol{\beta}}) = \tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\beta}}$ under the condition

$$(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) = (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) + d, \quad d > 0,$$

where $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is the value of $\mathbf{t}$ for which the function

$$\phi(\mathbf{t}) = (\mathbf{y} - \mathbf{X}\mathbf{t})'(\mathbf{y} - \mathbf{X}\mathbf{t})$$

attains its minimum. They obtained the estimator $\tilde{\boldsymbol{\beta}}$ of the form $\tilde{\boldsymbol{\beta}} = (c\mathbf{I} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ (ridge estimator), where $c > 0$ can be chosen in such a way that $\mathrm{Tr}\,\big[\,\mathrm{Var}(\tilde{\boldsymbol{\beta}})\big] + \big[E(\tilde{\boldsymbol{\beta}}) - \boldsymbol{\beta}\big]'\big[E(\tilde{\boldsymbol{\beta}}) - \boldsymbol{\beta}\big]$ is significantly smaller than $\mathrm{Tr}\,\big[\,\mathrm{Var}(\widehat{\boldsymbol{\beta}})\big]$ mainly in the case that $\|\boldsymbol{\beta}\|$ is relatively small with respect to σ and the matrix $\mathbf{X}'\mathbf{X}$ is bad conditioned.

In [7] new reasons for utilization of the ridge estimator are given and in [5] a general view on the philosophy of the ridge estimator is analyzed.

3 Modification

Let us try to find explicit value for c in the expression for the ridge estimator. The spectral decomposition of the matrix $\mathbf{X}'\mathbf{X}$ from Introduction is used. The quantity $\mathrm{Tr}\left[\mathrm{Var}(\tilde{\boldsymbol{\beta}})\right] + \mathbf{b}'_\beta \mathbf{b}_\beta$ is a function of c and it can be expressed as

$$\Phi(c) = \mathrm{Tr}\left[\mathrm{Var}(\tilde{\boldsymbol{\beta}})\right] + \mathbf{b}'_\beta \mathbf{b}_\beta = \sigma^2 \,\mathrm{Tr}\left[(c\mathbf{I} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(c\mathbf{I} + \mathbf{X}'\mathbf{X})^{-1}\right]$$
$$+\boldsymbol{\beta}'\left[(c\mathbf{I}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}-\mathbf{I}\right]^2\boldsymbol{\beta} = \sigma^2\,\mathrm{Tr}\left[(c\mathbf{F}\mathbf{F}'+\mathbf{F}\boldsymbol{\Lambda}\mathbf{F}')^{-1}\mathbf{F}\boldsymbol{\Lambda}\mathbf{F}'(c\mathbf{F}\mathbf{F}'+\mathbf{F}\boldsymbol{\Lambda}\mathbf{F}')^{-1}\right]$$
$$+\boldsymbol{\beta}'\left[(c\mathbf{F}\mathbf{F}' + \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}')^{-1}\mathbf{F}\boldsymbol{\Lambda}\mathbf{F}' - \mathbf{F}\mathbf{F}'\right]^2\boldsymbol{\beta} = \sigma^2\sum_{i=1}^k \frac{\lambda_i}{(c+\lambda_i)^2} + \sum_{i=1}^k \frac{c^2(\mathbf{f}'_i\boldsymbol{\beta})^2}{(c+\lambda_i)^2}.$$

Thus

$$\frac{d\Phi(c)}{dc} = \sum_{i=1}^k \frac{2\lambda_i(c(\mathbf{f}'_i\boldsymbol{\beta})^2 - \sigma^2)}{(c+\lambda_i)^3}$$

and in the case that $\boldsymbol{\beta}$ is approximately known, it is possible to solve the equation

$$\sum_{i=1}^k \frac{\lambda_i(c(\mathbf{f}'_i\boldsymbol{\beta})^2 - \sigma^2)}{(c+\lambda_i)^3} = 0$$

for c.

Let the function

$$\Phi(\mathbf{t}) = \mathbf{t}'\overline{\mathbf{D}}\mathbf{t} + \lambda\left[(\mathbf{y}-\mathbf{X}\mathbf{t})'(\mathbf{y}-\mathbf{X}\mathbf{t}) - (\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}) - d\right] \tag{4}$$

instead the function

$$\Phi(\mathbf{t}) = \mathbf{t}'\mathbf{t} + \lambda\left[(\mathbf{y}-\mathbf{X}\mathbf{t})'(\mathbf{y}-\mathbf{X}\mathbf{t}) - (\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}) - d\right]$$

be considered. Her $\overline{\mathbf{D}}$ is a positive definite matrix which will be determined later.

Theorem 3.1 *The solution of the optimization problem (4) is*

$$\mathbf{t} = (\mathbf{D} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

where $\mathbf{D} = \overline{\mathbf{D}}/\lambda$. *The MSE-optimum choice of the matrix* $\mathbf{D}$ *is*

$$\mathbf{D} = \sigma^2\mathbf{F}\,\mathrm{Diag}\left(\frac{1}{(\mathbf{f}'_1\boldsymbol{\beta})^2}, \dots, \frac{1}{(\mathbf{f}'_k\boldsymbol{\beta})^2}\right)\mathbf{F}',$$

where $\mathbf{X}'\mathbf{X} = \mathbf{F}\boldsymbol{\Lambda}\mathbf{F}'$, $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_k)$, $\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1, & \dots, & 0 \\ \dots & \dots & \dots \\ 0, & \dots, & \lambda_k \end{pmatrix}$.

Proof Let $\mathbf{U} = \mathrm{Diag}(u_1, \dots, u_k)$ and $\mathbf{D} = \mathbf{F}\mathbf{U}\mathbf{F}'$. Then the MSE of the estimator $\mathbf{t}$ can be written as

$$\mathrm{Tr}\left[\mathrm{Var}(\mathbf{t})\right] + \mathbf{b}'_\beta \mathbf{b}_\beta = \sigma^2\, \mathrm{Tr}\left\{\left[\mathbf{F}(\mathbf{U}+\boldsymbol{\Lambda})\mathbf{F}'\right]^{-1}\mathbf{F}\boldsymbol{\Lambda}\mathbf{F}'\left[\mathbf{F}(\mathbf{U}+\boldsymbol{\Lambda})\mathbf{F}'\right]^{-1}\right\}$$
$$+\boldsymbol{\beta}'\left\{\mathbf{F}\boldsymbol{\Lambda}\mathbf{F}'\left[\mathbf{F}(\mathbf{U}+\boldsymbol{\Lambda})\mathbf{F}'\right]^{-1} - \mathbf{F}\mathbf{F}'\right\}\left\{\left[\mathbf{F}(\mathbf{U}+\boldsymbol{\Lambda})\mathbf{F}'\right]^{-1}\mathbf{F}\boldsymbol{\Lambda}\mathbf{F}' - \mathbf{F}\mathbf{F}'\right\}\boldsymbol{\beta}$$
$$= \sigma^2\, \mathrm{Tr}\left[\mathbf{F}\,\mathrm{Diag}\left(\frac{1}{u_1+\lambda_1},\dots,\frac{1}{u_k+\lambda_k}\right)\mathrm{Diag}\left(\lambda_1,\dots,\lambda_k\right)\right.$$
$$\left.\times\,\mathrm{Diag}\left(\frac{1}{u_1+\lambda_1},\dots,\frac{1}{u_k+\lambda_k}\right)\mathbf{F}'\right]$$
$$+\boldsymbol{\beta}'\left[\mathbf{F}\,\mathrm{Diag}\left(\frac{\lambda_1}{u_1+\lambda_1},\dots,\frac{\lambda_k}{u_k+\lambda_k}\right)\mathbf{F}' - \mathbf{F}\mathbf{F}'\right]$$
$$\times\left[\mathbf{F}\,\mathrm{Diag}\left(\frac{\lambda_1}{u_1+\lambda_1},\dots,\frac{\lambda_k}{u_k+\lambda_k}\right)\mathbf{F}' - \mathbf{F}\mathbf{F}'\right]\boldsymbol{\beta} = \sum_{i=1}^{k}\frac{\sigma^2\lambda_i + (\mathbf{f}'_i\boldsymbol{\beta})^2u_i^2}{(u_i+\lambda_i)^2}.$$

Here the relationship

$$\left[\mathbf{F}(\mathbf{U}+\boldsymbol{\Lambda})\mathbf{F}'\right]^{-1} = \mathbf{F}(\mathbf{U}+\boldsymbol{\Lambda})^{-1}\mathbf{F}'$$

was taken into account.

The optimum entries $u_1,\dots,u_k$ of the matrix $\mathbf{U}$ can be now easily find.

$$\Phi(\mathbf{U}) = \sum_{i=1}^{k}\frac{\sigma^2\lambda_i + (\mathbf{f}'_i\boldsymbol{\beta})^2u_i^2}{(u_i+\lambda_i)^2},$$
$$\frac{\partial\Phi(\mathbf{U})}{\partial u_i} = 2\frac{(\mathbf{f}'_i\boldsymbol{\beta})^2u_i(u_i+\lambda_i) - \left[\sigma^2\lambda_i + (\mathbf{f}'_i\boldsymbol{\beta})^2u_i^2\right]}{(u_i+\lambda_i)^3}.$$

Thus $u_i = \sigma^2/(\mathbf{f}'_i\boldsymbol{\beta})^2$, $i=1,\dots,k$ and

$$\mathbf{D} = \sigma^2\mathbf{F}\,\mathrm{Diag}\left(\frac{1}{(\mathbf{f}'_1\boldsymbol{\beta})^2},\dots,\frac{1}{(\mathbf{f}'_k\boldsymbol{\beta})^2}\right)\mathbf{F}'.$$

□

Remark 3.2 The matrix $\mathbf{D}$ must be determined by the iteration. We start with some $\boldsymbol{\beta}^{(0)}$, then obtain the matrix $\mathbf{D}_1^{(0)}$, by the help of it we obtain the estimator $\tilde{\boldsymbol{\beta}}^{(1)}$, etc. The choice of the matrix $\mathbf{D}$ from Theorem 3.1 gives

$$\mathrm{Tr}\left[\mathrm{Var}(\mathbf{t})\right] + \mathbf{b}'_\beta\mathbf{b}_\beta = \sum_{i=1}^{k}\frac{\lambda_i\sigma^2 + (\mathbf{f}'_i\boldsymbol{\beta})^2\frac{\sigma^4}{(\mathbf{f}'_i\boldsymbol{\beta})^4}}{\left(\frac{\sigma^2}{(\mathbf{f}'_i\boldsymbol{\beta})^2}+\lambda_i\right)^2} = \sigma^2\sum_{i=1}^{k}\frac{(\mathbf{f}'_i\boldsymbol{\beta})^2}{\lambda_i(\mathbf{f}_i\boldsymbol{\beta})^2+\sigma^2},$$

what can be significantly smaller than $\mathrm{Tr}\left[\mathrm{Var}(\widehat{\boldsymbol{\beta}})\right] = \sigma^2\sum_{i=1}^{k}\frac{1}{\lambda_i}$.

4 Model with the type I constraints

The model is

$$\mathbf{Y} \sim_n (\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}), \quad \mathbf{g} + \mathbf{G}\boldsymbol{\beta} = \mathbf{0},$$

where $r(\mathbf{X}_{n,k}) = k < n,\ r(\mathbf{G}_{q,k}) = q < k$. The matrix $\mathbf{X}$, the vector $\mathbf{g}$ and the matrix $\mathbf{G}$ are given.

Obviously $\mathcal{M}(\mathbf{G}') \subset \mathcal{M}(\mathbf{X}')$. In this case the BLUE of the parameter $\boldsymbol{\beta}$ is

$$\widehat{\widehat{\boldsymbol{\beta}}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\big[\mathbf{G}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\big]^{-1}\big[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} + \mathbf{g}\big]$$

and its covariance matrix is

$$\begin{aligned}\operatorname{Var}(\widehat{\widehat{\boldsymbol{\beta}}}) &= \sigma^2\Big\{(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\big[\mathbf{G}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\big]^{-1}\mathbf{G}(\mathbf{X}'\mathbf{X})^{-1}\Big\}\\ &= \sigma^2\left(\mathbf{M}_{G'}\mathbf{X}'\mathbf{X}\mathbf{M}_{G'}\right)^+,\end{aligned}$$

where $\mathbf{M}_{G'} = \mathbf{I} - \mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1}\mathbf{G}$ and $^+$ is notation for the Moore–Penrose generalized inverse (see [6]).

If the idea of Hoerl and Kennard is a little bit generalized, we seek for an estimator $\tilde{\tilde{\boldsymbol{\beta}}}$ which satisfy the constraints $\mathbf{g} + \mathbf{G}\tilde{\tilde{\boldsymbol{\beta}}} = \mathbf{0}$ and also the constraints $(\mathbf{y} - \mathbf{X}\tilde{\tilde{\boldsymbol{\beta}}})'(\mathbf{y} - \mathbf{X}\tilde{\tilde{\boldsymbol{\beta}}}) = (\mathbf{y} - \mathbf{X}\widehat{\widehat{\boldsymbol{\beta}}})'(\mathbf{y} - \mathbf{X}\widehat{\widehat{\boldsymbol{\beta}}}) + d$ and at the same time it will minimize the quantity $\tilde{\tilde{\boldsymbol{\beta}}}'\overline{\mathbf{D}}\tilde{\tilde{\boldsymbol{\beta}}}$.

Lemma 4.1 *The estimator* $\tilde{\tilde{\boldsymbol{\beta}}}$ *is*

$$\begin{aligned}\tilde{\tilde{\boldsymbol{\beta}}} &= (\mathbf{D} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{D} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\big[\mathbf{G}(\mathbf{D} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\big]^{-1}\\ &\times\big[\mathbf{G}(\mathbf{D} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} + \mathbf{g}\big] = \mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)}(\mathbf{D} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \mathbf{G}^-_{m(D+X'X)}\mathbf{g}\end{aligned}$$

where $\mathbf{D} = \sigma^2\mathbf{F}\,\mathrm{Diag}\left(\frac{1}{(\mathbf{f}_1'\boldsymbol{\beta})^2}, \dots, \frac{1}{(\mathbf{f}_k'\boldsymbol{\beta})^2}\right)\mathbf{F}'$.

Proof The auxiliary Lagrange function is

$$\Phi(\mathbf{t}) = \mathbf{t}'\overline{\mathbf{D}}\mathbf{t} - \lambda\big[(\mathbf{y} - \mathbf{X}\mathbf{t})'(\mathbf{y} - \mathbf{X}\mathbf{t}) - (\mathbf{y} - \mathbf{X}\widehat{\widehat{\boldsymbol{\beta}}})'(\mathbf{y} - \mathbf{X}\widehat{\widehat{\boldsymbol{\beta}}}) - d\big] + 2\boldsymbol{\kappa}'(\mathbf{G}\mathbf{t} + \mathbf{g}),$$

where λ is the Lagrange multiplier and $\boldsymbol{\kappa}$ is a vector of the Lagrange multipliers. Thus

$$\begin{aligned}\frac{\partial\Phi(\mathbf{t})}{\partial\mathbf{t}} &= 2\overline{\mathbf{D}}\mathbf{t} + \lambda(-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{t}) + 2\mathbf{G}'\boldsymbol{\kappa} = \mathbf{0}\\ &\Rightarrow (\mathbf{D} + \mathbf{X}'\mathbf{X})\mathbf{t} + \mathbf{G}'\frac{\boldsymbol{\kappa}}{\lambda} = \mathbf{X}'\mathbf{y},\end{aligned}$$

where $\mathbf{D} = \overline{\mathbf{D}}/\lambda$ and

$$\tilde{\tilde{\boldsymbol{\beta}}} = (\mathbf{D} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{D} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\frac{\boldsymbol{\kappa}}{\lambda}.$$

Further

$$\begin{aligned}
\mathbf{0} = \mathbf{g} + \mathbf{G}\tilde{\tilde{\boldsymbol\beta}} &= \mathbf{g} + \mathbf{G}(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \mathbf{G}(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\frac{\boldsymbol\kappa}{\lambda}\\
\Rightarrow \frac{\boldsymbol\kappa}{\lambda} &= \big[\mathbf{G}(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\big]^{-1}\big[\mathbf{G}(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} + \mathbf{g}\big]\\
\Rightarrow \tilde{\tilde{\boldsymbol\beta}} &= (\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\Big[\mathbf{G}(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\Big]^{-1}\\
&\qquad \times\big[\mathbf{G}(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} + \mathbf{g}\big]\\
&= \Big\{\mathbf{I} - (\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\big[\mathbf{G}(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\big]^{-1}\mathbf{G}\Big\}(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\\
&\qquad -(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\big[\mathbf{G}(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\big]^{-1}\mathbf{g}\\
&= \mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)}(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \mathbf{G}^{-}_{m(D+X'X)}\mathbf{g},
\end{aligned}$$

where $\mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)}$ is the projection matrix on $\mathcal{K}er(\mathbf{G}) = \{\mathbf{u}\colon \mathbf{u}\in R^k, \mathbf{G}\mathbf{u} = \mathbf{0}\}$ in the norm given by the positive definite matrix $\mathbf{D}+\mathbf{X}'\mathbf{X}$.

The bias of the estimator $\tilde{\tilde{\boldsymbol\beta}}$ is

$$\mathbf{b}_\beta = E(\tilde{\tilde{\boldsymbol\beta}}) - \boldsymbol\beta = \mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)}\big[(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} - \mathbf{I}\big]\boldsymbol\beta.$$

and

$$\mathrm{Var}(\tilde{\tilde{\boldsymbol\beta}}) = \sigma^2\mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)}(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\left(\mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)}\right)'$$

Since the bias of the estimator $\tilde{\boldsymbol\beta} = (\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is

$$E(\tilde{\boldsymbol\beta}) - \boldsymbol\beta = \big[(\mathbf{D}+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} - \mathbf{I}\big]\boldsymbol\beta,$$

it is obvious that $\mathbf{b}_\beta = \mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)}\big[E(\tilde{\boldsymbol\beta}) - \boldsymbol\beta\big]$ and analogously for the covariance matrix

$$\mathrm{Var}(\tilde{\tilde{\boldsymbol\beta}}) = \mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)}\,\mathrm{Var}(\tilde{\boldsymbol\beta})\left(\mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)}\right)'.$$

Thus the estimator $\tilde{\boldsymbol\beta}$ from the model without constraints can be used in the formula for the estimator $\tilde{\tilde{\boldsymbol\beta}}$ in the model with the type I constraints. The bias and the covariance matrix are reasonably diminished by the projection matrix $\mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)}$ which fully respects the constraints. □

5 Model with the type II constraints

The model considered is

$$\mathbf{Y} \sim_n (\mathbf{X}\boldsymbol\beta_1, \sigma^2\mathbf{I}), \quad \mathbf{g} + \mathbf{G}_1\boldsymbol\beta_1 + \mathbf{G}_2\boldsymbol\beta_2 = \mathbf{0},$$

where $r(\mathbf{X}_{(n,k_1)}) = k_1 < n,\ r(\mathbf{G}_{1,(q,k_1)}, \mathbf{G}_{2,(q,k_2)}) = q < k_1 + k_2,\ r(\mathbf{G}_{2,(q,k_2)}) = k_2 < q$. The matrix $\mathbf{X}$, the vector $\mathbf{g}$ and the matrices $\mathbf{G}_1,\ \mathbf{G}_2$ are given.

The BLUEs of the vector parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are

$$\hat{\hat{\boldsymbol{\beta}}}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\big[\mathbf{M}_{G_2}\mathbf{G}_1(\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\mathbf{M}_{G_2}\big]^+ \times \big[\mathbf{g} + \mathbf{G}_1(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\big],$$

$$\hat{\hat{\boldsymbol{\beta}}}_2 = -\big[(\mathbf{G}_2')^-_{m[G_1'(X'X)^{-1}G_1]}\big]'\big[\mathbf{g} + \mathbf{G}_1(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\big]$$

(in more detail see [1]).

In both estimators the effect of the bad conditioned matrix $\mathbf{X}'\mathbf{X}$ is fully manifested.

The modification of the Kennard and Hoerl approach can be made in two ways.

The first one starts with a minimization of the quantity $\tilde{\tilde{\boldsymbol{\beta}}}_1'\overline{\mathbf{D}}_1\tilde{\tilde{\boldsymbol{\beta}}}_1$ and the other starts with a minimization of the quantity $\tilde{\tilde{\boldsymbol{\beta}}}_1'\overline{\mathbf{D}}_1\tilde{\tilde{\boldsymbol{\beta}}}_1 + \tilde{\tilde{\boldsymbol{\beta}}}_2'\overline{\mathbf{D}}_2\tilde{\tilde{\boldsymbol{\beta}}}_2$.

Let us consider the minimization of $\tilde{\tilde{\boldsymbol{\beta}}}_1\overline{\mathbf{D}}_1\tilde{\tilde{\boldsymbol{\beta}}}_1$, i.e.

$$\Phi(\mathbf{t}_1, \mathbf{t}_2) = \mathbf{t}_1'\overline{\mathbf{D}}_1\mathbf{t}_1 + \lambda\big[(\mathbf{y} - \mathbf{X}\mathbf{t}_1)'(\mathbf{y} - \mathbf{X}\mathbf{t}_1) - (\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}_1)'(\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}_1) - d\big] - 2\boldsymbol{\kappa}'(\mathbf{g} + \mathbf{G}_1\mathbf{t}_1 + \mathbf{G}_2\mathbf{t}_2),$$

$$\frac{\partial\Phi(\mathbf{t}_1, \mathbf{t}_2)}{\partial\mathbf{t}_1} = 2\overline{\mathbf{D}}_1\mathbf{t}_1 + \lambda(-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{t}_1) - 2\mathbf{G}_1'\boldsymbol{\kappa},$$

$$\frac{\partial\Phi(\mathbf{t}_1, \mathbf{t}_2)}{\partial\mathbf{t}_2} = -2\mathbf{G}_2'\boldsymbol{\kappa}.$$

Let $\mathbf{D}_1 = \frac{1}{\lambda}\overline{\mathbf{D}}_1$. Thus the following relationships can be written.

$$(\overline{\mathbf{D}}_1 + \lambda\mathbf{X}'\mathbf{X})\mathbf{t}_1 - \lambda\mathbf{X}'\mathbf{y} - \mathbf{G}_1'\boldsymbol{\kappa} = \mathbf{0},$$

$$\mathbf{t}_1 = (\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + (\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\frac{\boldsymbol{\kappa}}{\lambda},$$

$$\mathbf{g} + \mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} + \mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\frac{\boldsymbol{\kappa}}{\lambda} + \mathbf{G}_2\mathbf{t}_2 = \mathbf{0}$$

$$\begin{pmatrix} \mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1', & \mathbf{G}_2 \\ \mathbf{G}_2', & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\kappa}/\lambda \\ \mathbf{t}_2 \end{pmatrix} = -\begin{pmatrix} \mathbf{g} + \mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ \mathbf{0} \end{pmatrix}.$$

Regarding the Pandora–Box theorem [6], we obtain

$$\begin{pmatrix} \boldsymbol{\kappa}/\lambda \\ \mathbf{t}_2 \end{pmatrix} = \begin{pmatrix} \boxed{\text{aa}}, & \boxed{\text{ab}} \\ \boxed{\text{ba}}, & \boxed{\text{bb}} \end{pmatrix} \begin{pmatrix} -\big[\mathbf{g} + \mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\big] \\ \mathbf{0} \end{pmatrix},$$

$$\boxed{\text{aa}} = \big[\mathbf{M}_{G_2}\mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\mathbf{M}_{G_2}\big]^+,$$

$$\boxed{\text{ab}} = (\mathbf{G}_2')^-_{m[G_1(D_1+X'X)^{-1}G_1']},$$

$$\boxed{\text{ba}} = \big[(\mathbf{G}_2')^-_{m[G_1(D_1+X'X)^{-1}G_1']}\big]',$$

$$\boxed{\text{bb}} = -\big[(\mathbf{G}_2')^-_{m[G_1(D_1+X'X)^{-1}G_1']}\big]'\mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'(\mathbf{G}_2')^-_{m[G_1(D_1+X'X)^{-1}G_1']}.$$

Thus the following theorem can be stated.

Theorem 5.1 *In the regular model with the type constraints II the estimators* $\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)}$ *and* $\tilde{\tilde{\boldsymbol{\beta}}}_2^{(1)}$ *minimizing the quantity* $\tilde{\tilde{\boldsymbol{\beta}}}_1'\mathbf{D}_1\tilde{\tilde{\boldsymbol{\beta}}}_1$ *and satisfying the constraints*

$$(\mathbf{y}-\mathbf{X}\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)})'(\mathbf{y}-\mathbf{X}\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)}) - \big[(\mathbf{y}-\mathbf{X}\widehat{\widehat{\boldsymbol{\beta}}}_1)'(\mathbf{y}-\mathbf{X}\widehat{\widehat{\boldsymbol{\beta}}}_1) + d\big] = 0$$

and

$$\mathbf{g} + \mathbf{G}_1\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)} + \mathbf{G}_2\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)} = \mathbf{0}.$$

are

$$\begin{aligned}\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)} &= (\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\big[\mathbf{M}_{G_2}\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1} \\ &\quad\times \mathbf{G}_1'\mathbf{M}_{G_2}\big]^+\big[\mathbf{g}+\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\big], \\ \tilde{\tilde{\boldsymbol{\beta}}}_2^{(1)} &= -\big[(\mathbf{G}_2')^-_{m[G_1(D_1+X'X)^{-1}G_1']}\big]'\big[\mathbf{g}+\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\big].\end{aligned}$$

Since $\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)}$ can be written as

$$\begin{aligned}\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)} &= \Big(\mathbf{I} - (\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\mathbf{M}_{G_2}\big[\mathbf{M}_{G_2}\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\mathbf{M}_{G_2}\big]^+\mathbf{M}_{G_2}\mathbf{G}_1\Big) \\ &\times (\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\big[\mathbf{M}_{G_2}\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\mathbf{M}_{G_2}\big]^+\mathbf{g} \\ &= \mathbf{P}^{(D_1+X'X)}_{\mathcal{K}er(M_{G_2}G_1)}(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1' \\ &\quad\times \big[\mathbf{M}_{G_2}\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\mathbf{M}_{G_2}\big]^+\mathbf{g},\end{aligned}$$

the bias of the estimator $\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)}$ is

$$E(\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)}) - \boldsymbol{\beta}_1 = \mathbf{P}^{(D_1+X'X)}_{\mathcal{K}er(M_{G_2}G_1)}\big[(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}-\mathbf{I}\big]\boldsymbol{\beta}_1.$$

The same reasoning for the utilization of the matrix

$$\mathbf{D}_1 = \sigma^2\mathbf{F}\,\mathrm{Diag}\left(\frac{1}{(\mathbf{f}_1'\boldsymbol{\beta}_1)^2},\dots,\frac{1}{(\mathbf{f}_k'\boldsymbol{\beta}_1)^2}\right)\mathbf{F}'$$

can be made as at the end of the section 4.

As far as the estimator $\tilde{\tilde{\boldsymbol{\beta}}}_2^{(1)}$ is concerned, there exists just one solution of the equation

$$\mathbf{g} + \mathbf{G}_1\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)} + \mathbf{G}_2\tilde{\tilde{\boldsymbol{\beta}}}_2^{(1)} = \mathbf{0}$$

for $\tilde{\tilde{\boldsymbol{\beta}}}_2^{(1)}$, since $r(\mathbf{G}_2) = k_2 < q$. Thus

$$\begin{aligned}\tilde{\tilde{\boldsymbol{\beta}}}_2^{(1)} &= -(\mathbf{G}_2'\mathbf{G}_2)^{-1}\mathbf{G}_2'(\mathbf{G}_1\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)} + \mathbf{g}) \\ &= -\big[(\mathbf{G}_2')^-_{m[G_1(D_1+X'X)^{-1}G_1']}\big]'\big[\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}+\mathbf{g}\big].\end{aligned}$$

Thus the matrix

$$\mathbf{D}_1 = \sigma^2 \mathbf{F}\,\mathrm{Diag}\left(\frac{1}{(\mathbf{f}_1\boldsymbol{\beta}_1)^2}, \dots, \frac{1}{(\mathbf{f}_k\boldsymbol{\beta}_1)^2}\right)\mathbf{F}'$$

can be used for the estimator $\tilde{\tilde{\boldsymbol{\beta}}}_2^{(1)}$ as well.

Let us consider a little more general problem, i.e. a minimization of the function $\phi(\mathbf{t}_1, \mathbf{t}_2) = \mathbf{t}_1'\overline{\mathbf{D}}_1\mathbf{t}_1 + \mathbf{t}_2'\overline{\mathbf{D}}_2\mathbf{t}_2$ under the conditions

$$\begin{gathered}\big[(\mathbf{y} - \mathbf{X}\mathbf{t}_1)'(\mathbf{y} - \mathbf{X}\mathbf{t}_1) - (\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}_1)'(\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}_1) - d\big],\\ \mathbf{G}_1\mathbf{t}_1 + \mathbf{G}_2\mathbf{t}_2 + \mathbf{g} = \mathbf{0}.\end{gathered}$$

The Lagrangian auxiliary function is

$$\begin{aligned}\Phi(\mathbf{t}_1, \mathbf{t}_2) &= \mathbf{t}_1'\overline{\mathbf{D}}_1\mathbf{t}_1 + \mathbf{t}_2'\overline{\mathbf{D}}_2\mathbf{t}_2 + \lambda\big[(\mathbf{y} - \mathbf{X}\mathbf{t}_1)'(\mathbf{y} - \mathbf{X}\mathbf{t}_1)\\ &- (\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}_1)'(\mathbf{y} - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}_1) - d\big] + 2\boldsymbol{\kappa}'(\mathbf{G}_1\mathbf{t}_1 + \mathbf{G}_2\mathbf{t}_2 + \mathbf{g}).\end{aligned}$$

Then

$$\begin{aligned}\frac{\partial\Phi(\mathbf{t}_1, \mathbf{t}_2)}{\partial\mathbf{t}_1} &= 2\overline{\mathbf{D}}_1\mathbf{t}_1 + \lambda(-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{t}_1) + 2\mathbf{G}_1\boldsymbol{\kappa},\\ \frac{\partial\Phi(\mathbf{t}_1, \mathbf{t}_2)}{\partial\mathbf{t}_2} &= 2\overline{\mathbf{D}}_2\mathbf{t}_2 + 2\mathbf{G}_2'\boldsymbol{\kappa}\end{aligned}$$

and

$$\begin{aligned}\mathbf{t}_1 &= (\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - (\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\frac{\boldsymbol{\kappa}}{\lambda},\\ \mathbf{t}_2 &= -\mathbf{D}_2^{-1}\mathbf{G}_2'\frac{\boldsymbol{\kappa}}{\lambda},\end{aligned}$$

where $\mathbf{D}_1 = \overline{\mathbf{D}}_1/\lambda$ and $\mathbf{D}_2 = \overline{\mathbf{D}}_2/\lambda$.

Since

$$\begin{aligned}\mathbf{0} = \mathbf{g} + \mathbf{G}_1\tilde{\tilde{\boldsymbol{\beta}}}_1 + \mathbf{G}_2\tilde{\tilde{\boldsymbol{\beta}}}_2 &= \mathbf{g} + (\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\\ &-\big[\mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1' + \mathbf{G}_2\mathbf{D}_2^{-1}\mathbf{G}_2'\big]^{-1}\frac{\boldsymbol{\kappa}}{\lambda},\end{aligned}$$

we have

$$\frac{\boldsymbol{\kappa}}{\lambda} = \big[\mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1' + \mathbf{G}_2\mathbf{D}_2^{-1}\mathbf{G}_2'\big]^{-1}\big[\mathbf{g} + \mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\big].$$

The following theorem can be stated.

Theorem 5.2 *In the regular model with the constraints II*

$$\mathbf{Y} \sim_n (\mathbf{X}\boldsymbol{\beta}_1, \sigma^2\mathbf{I}), \quad \mathbf{g} + \mathbf{G}_1\boldsymbol{\beta}_1 + \mathbf{G}_2\boldsymbol{\beta}_2 = \mathbf{0},$$

where

$$r(\mathbf{X}_{n,k_1}) = k_1 < n,\ r(\mathbf{G}_{1,(q,k_1)}, \mathbf{G}_{2,(q,k_2)}) = q < k_1 + k_2,\ r(\mathbf{G}_{2,(q,k_2)}) = k_2 < q,$$

the estimators $\tilde{\tilde{\boldsymbol{\beta}}}_1^{(2)}$ *and* $\tilde{\tilde{\boldsymbol{\beta}}}_2^{(2)}$ *minimizing the quantity*

$$(\tilde{\tilde{\boldsymbol{\beta}}}_1^{(2)})'\overline{\mathbf{D}}_1\tilde{\tilde{\boldsymbol{\beta}}}_1^{(2)} + (\tilde{\tilde{\boldsymbol{\beta}}}_2^{(2)})'\overline{\mathbf{D}}_2\tilde{\tilde{\boldsymbol{\beta}}}_2^{(2)}$$

are

$$\begin{aligned}
\tilde{\tilde{\boldsymbol{\beta}}}_1^{(2)} &= (\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\big[\mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1' \\
&\quad + \mathbf{G}_2\mathbf{D}_2^{-1}\mathbf{G}_2'\big]^{-1}\big[\mathbf{g} + \mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\big], \\
\tilde{\tilde{\boldsymbol{\beta}}}_2^{(2)} &= -\mathbf{D}_2^{-1}\mathbf{G}_2'\big[\mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1' + \mathbf{G}_2\mathbf{D}_2^{-1}\mathbf{G}_2'\big]^{-1} \\
&\quad \times \big[\mathbf{g} + \mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\big],
\end{aligned}$$

where $\mathbf{D}_1 = \overline{\mathbf{D}}_1/\lambda$ *and* $\mathbf{D}_2 = \overline{\mathbf{D}}_2/\lambda$.

The problem of the MSE-optimization of the matrix $\mathbf{D}_2$ is out of the scope of the paper. For the sake of simplicity the choice $\mathbf{D}_2 = \mathbf{I}$ can be used.

6 Numerical examples

Let singular value decomposition of the matrix $\mathbf{X}$ be

$$\mathbf{X}_{n,k} = \mathbf{J}_{8,3}\boldsymbol{\Lambda}_{3,3}^{1/2}\mathbf{F}_{3,3}', \quad \mathbf{J}'\mathbf{J} = \mathbf{I}_3, \boldsymbol{\Lambda}^{1/2} = \mathrm{Diag}\left(\sqrt{4}, \sqrt{4}, \sqrt{0.1}\right), \quad \mathbf{F} = \mathbf{I}_3$$

$\sigma^2 = 1$ and

$$\boldsymbol{\beta} = \begin{pmatrix} 0.5 \\ 0.4 \\ 0.2 \end{pmatrix}, \quad \mathbf{X}'\mathbf{X} = \mathrm{Diag}\,(4, 4, 0.1), \quad \mathbf{D} = \mathbf{F}\,\mathrm{Diag}\left(\frac{1}{0.5^2}, \frac{1}{0.4^2}, \frac{1}{0.2^2}\right)\mathbf{F}'.$$

Thus

$$\begin{aligned}
\mathrm{Var}(\widehat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-1} = \mathrm{Diag}(0.25, 0.25, 10), \\
\tilde{\boldsymbol{\beta}} = (\mathbf{D} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} &= \left[\mathrm{Diag}\left(\frac{1}{0.5^2}, \frac{1}{0.4^2}, \frac{1}{0.2^2}\right) + \mathrm{Diag}\,(4, 4, 0.1)\right]^{-1} \\
&\times \mathrm{Diag}\left(\sqrt{4}, \sqrt{4}, \sqrt{0.1}\right)\mathbf{J}'\mathbf{y} = \mathrm{Diag}(0.250000, 0.195122, 0.012599)\mathbf{J}'\mathbf{y}, \\
\mathrm{Var}(\tilde{\boldsymbol{\beta}}) &= \mathrm{Diag}(0.062500, 0.038073, 0.000159) \\
E(\tilde{\boldsymbol{\beta}}) &= \mathrm{Diag}(0.250000, 0.195122, 0.012599)\mathbf{J}'\mathbf{J}\boldsymbol{\Lambda}^{1/2}\mathbf{F}'\boldsymbol{\beta} \\
&= (0.250000, 0.156098, 0.000797)'.
\end{aligned}$$

The bias of $\tilde{\boldsymbol{\beta}}$ is

$$\begin{aligned}
\mathbf{b}_\beta &= (0.250000, 0.156098, 0.000797)' - (0.5, 0.4, 0.2)' \\
&= (-0.250000, -0.243902, -0.199203)'.
\end{aligned}$$

Thus

$\operatorname{Tr}\left[\operatorname{Var}(\widehat{\boldsymbol{\beta}})\right] = 10.5$ and $\operatorname{Tr}\left[\operatorname{Var}(\tilde{\boldsymbol{\beta}})\right] + \mathbf{b}'_\beta \mathbf{b}_\beta = 0.100732 + 0.161670 = 0.262402.$

The effect of the optimization is expressive.

Let this model with constraints $(2,-2,-1)\boldsymbol{\beta} = 0$ be considered, i.e. $\mathbf{g} = \mathbf{0}$ and $\mathbf{G} = (2,-2,-1)$. It is valid that

$$\mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)} = \begin{pmatrix} 0.462415, & 0.537585, & 0.268793 \\ 0.419579, & 0.580421, & -0.209789 \\ 0.085671, & -0.085671, & 0.957164 \end{pmatrix}$$

and

$$E(\tilde{\tilde{\boldsymbol{\beta}}}) - \boldsymbol{\beta} = \mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)}\left[E(\tilde{\boldsymbol{\beta}}) - \boldsymbol{\beta}\right] = \mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)} \begin{pmatrix} -0.250000 \\ -0.243902 \\ -0.199203 \end{pmatrix} = \begin{pmatrix} -0.300266 \\ -0.204670 \\ -0.191192 \end{pmatrix}.$$

Since

$$\operatorname{Var}(\tilde{\tilde{\boldsymbol{\beta}}}) = \mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)} \operatorname{Var}(\tilde{\boldsymbol{\beta}}) \left(\mathbf{P}^{(D+X'X)}_{\mathcal{K}er(G)}\right)',$$

we have

$$\operatorname{Tr}\left[\operatorname{Var}(\tilde{\tilde{\boldsymbol{\beta}}})\right] + \left[E(\tilde{\tilde{\boldsymbol{\beta}}}) - \boldsymbol{\beta}\right]'\left[E(\tilde{\tilde{\boldsymbol{\beta}}}) - \boldsymbol{\beta}\right] = 0.049099 + 0.168604 = 0.217703.$$

The covariance matrix of the BLUE is

$$\operatorname{Var}(\widehat{\widehat{\boldsymbol{\beta}}}) = (\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\left[\mathbf{G}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}'\right]^{-1}\mathbf{G}(\mathbf{X}'\mathbf{X}^{-1})^{-1}$$

and thus

$$\operatorname{Tr}\left[\operatorname{Var}(\widehat{\widehat{\boldsymbol{\beta}}})\right] = 2.110007.$$

what is essentially larger than 0.217703.

Let now the type II constraints be considered, e.g.

$$\boldsymbol{\beta} = \boldsymbol{\beta}_1 = (0.5,\ 0.4,\ 0.2)',\ \mathbf{G}_1 = (2,-2,-1),\ \mathbf{G}_2 = 2,\ \mathbf{g} = -0.6,\ \beta_2 = 0.3.$$

Thus

$$\begin{aligned} \tilde{\tilde{\boldsymbol{\beta}}}^{(1)}_1 &= \tilde{\boldsymbol{\beta}}_1 - \mathbf{G}'_1\mathbf{M}_{G_2}\left[\mathbf{M}_{G_2}\mathbf{G}_1(\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1\mathbf{M}_{G_2}\right]^{+}\mathbf{M}_{G_2}(\mathbf{G}_1\tilde{\boldsymbol{\beta}}_1 + \mathbf{g}) \\ &= \tilde{\boldsymbol{\beta}}_1 = (\mathbf{D}_1 + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \end{aligned}$$

since $\mathbf{M}_{G_2} = \mathbf{0}$ and

$$\tilde{\tilde{\beta}}^{(1)}_2 = -\left[(\mathbf{G}'_2)^{-}_{m[G_1(X'X)^{-1}G'_1]}\right]'(\mathbf{g} + \mathbf{G}_1\tilde{\boldsymbol{\beta}}_1) = \frac{1}{2}\left[0.6 - (2,-2,-1)\tilde{\boldsymbol{\beta}}_1\right].$$

$$\mathbf{b}^{(1)}_{\beta_1} = E(\tilde{\tilde{\boldsymbol{\beta}}}^{(1)}_1) - \boldsymbol{\beta}_1 = E(\tilde{\boldsymbol{\beta}}_1) - \boldsymbol{\beta}_1 = (-0.250000, -0.243902, -0.199203)',$$

$$\operatorname{Var}(\tilde{\tilde{\boldsymbol{\beta}}}^{(1)}_1) = \operatorname{Var}(\tilde{\boldsymbol{\beta}}_1) = \operatorname{Diag}(0.062500, 0.038073, 0.000159),$$

$$\operatorname{Tr}\left[\operatorname{Var}(\tilde{\tilde{\boldsymbol{\beta}}}^{(1)}_1)\right] + (\mathbf{b}^{(1)}_{\beta_1})'\mathbf{b}^{(1)}_{\beta_1} = 0.262402,$$

$$\operatorname{Var}(\widehat{\widehat{\boldsymbol{\beta}}}_1) = \operatorname{Var}(\widehat{\boldsymbol{\beta}_1}) = \operatorname{Diag}(0.25, 0.25, 10) \Rightarrow \operatorname{Tr}\left[\operatorname{Var}(\widehat{\widehat{\boldsymbol{\beta}}}_1)\right] = 10.5$$

The difference $\mathrm{Tr}\left[\mathrm{Var}(\widehat{\widehat{\boldsymbol{\beta}}}_1)\right]-\left\{\mathrm{Tr}\left[\mathrm{Var}(\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)})\right]+(\mathbf{b}_{\beta_1}^{(1)})'\mathbf{b}_{\beta_1}^{(1)}\right\}$ is the same as in the first example.

Since

$$\left[(\mathbf{G}_2')^-_{m[G_1(X'X)^{-1}G_1']}\right]'=\frac{1}{2}\quad\text{and}\quad(\mathbf{g}+\mathbf{G}_1\tilde{\boldsymbol{\beta}}_1)=\left[-0.6+(2,-2,-1)\tilde{\boldsymbol{\beta}}_1\right],$$

we have

$$\tilde{\tilde{\beta}}_2^{(1)}=-\frac{1}{2}\left[-0.6+(2,-2,-1)\tilde{\beta}_1\right]$$

and

$$E\left(\tilde{\tilde{\beta}}_2^{(1)}\right)-\beta_2=0.206496-0.3=-0.093504=b_{\beta_2}^{(1)}.$$

Further

$$\mathrm{Var}(\tilde{\tilde{\beta}}_2^{(1)})=$$

$$=\frac{1}{4}(2,-2,-1)\,\mathrm{Diag}(0.062500,0.038073,0.000159)\begin{pmatrix}2\\-2\\-1\end{pmatrix}=0.100613.$$

Thus

$$\mathrm{Tr}\left[\mathrm{Var}\left(\tilde{\tilde{\beta}}_2^{(1)}\right)\right]+\left(b_{\beta_2}^{(1)}\right)^2=0.109356$$

$$\mathrm{Var}(\widehat{\widehat{\beta}}_2)=\left[(\mathbf{G}_2')^-_{m[G_1(X'X)^{-1}G_1']}\right]'\mathbf{G}_1(\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'(\mathbf{G}_2')^-_{m[G_1(X'X)^{-1}G_1']}=3.$$

As far as the estimators $\tilde{\tilde{\boldsymbol{\beta}}}_1^{(2)}$ and $\tilde{\tilde{\beta}}_2^{(2)}$ ($\mathbf{D}_2=\mathbf{I}$) are concerned, we obtain

$$\tilde{\tilde{\boldsymbol{\beta}}}_1^{(2)}=(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}-(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\left[\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'+\mathbf{G}_2\mathbf{G}_2'\right]^{-1}\times\left[\mathbf{g}+\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right],$$

$$\mathrm{Var}\left(\tilde{\tilde{\boldsymbol{\beta}}}_1^{(2)}\right)=\left\{\mathbf{I}-(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\left[\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'+\mathbf{G}_2\mathbf{G}_2'\right]^{-1}\mathbf{G}_1\right\}\times\mathrm{Var}(\tilde{\boldsymbol{\beta}}_1)\left\{\mathbf{I}-(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\left[\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'+\mathbf{G}_2\mathbf{G}_2'\right]^{-1}\mathbf{G}_1\right\}'$$

$$=\begin{pmatrix}0.050857, & 0.008001, & 0.000853\\ 0.008001, & 0.032676, & -0.000493\\ 0.000853, & -0.000493, & 0.000183\end{pmatrix},$$

$$E\left(\tilde{\tilde{\boldsymbol{\beta}}}_1^{(2)}\right)-\boldsymbol{\beta}_1=E(\tilde{\boldsymbol{\beta}}_1)-(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'\left[\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1'+\mathbf{G}_2\mathbf{G}_2'\right]^{-1}\times\left[\mathbf{g}+\mathbf{G}_1E(\tilde{\boldsymbol{\beta}}_1)\right]-\boldsymbol{\beta}_1=\begin{pmatrix}-0.148785\\-0.322899\\-0.215333\end{pmatrix},$$

$$\mathrm{Tr}\left[\mathrm{Var}\left(\tilde{\tilde{\boldsymbol{\beta}}}_1^{(2)}\right)\right]+\left(\mathbf{b}_{\beta_1}^{(2)}\right)'\mathbf{b}_{\beta_1}^{(2)}=0.083716+0.172769=0.256485.$$

The behaviour of $\tilde{\tilde{\boldsymbol{\beta}}}_1^{(1)}$ and $\tilde{\tilde{\boldsymbol{\beta}}}_1^{(2)}$ is similar; the MSEs of both estimators are almost the same.

The estimator $\tilde{\tilde{\beta}}_2^{(2)}$ is

$$\tilde{\tilde{\beta}}_2^{(2)} = -\mathbf{G}_2'\big[\mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}_1' + \mathbf{G}_2\mathbf{G}_2'\big]^{-1}\big[\mathbf{g} + \mathbf{G}_1(\mathbf{D}_1+\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\big]$$
$$= -2\frac{1}{4.930085}\big[-0.6 + (2,-2,-1)\tilde{\boldsymbol{\beta}}_1\big],$$

we have

$$E\left(\tilde{\tilde{\beta}}_2^{(2)}\right) - \beta_2 = b_{\beta_2}^{(2)} = -0.132460$$

and

$$\mathrm{Var}\left(\tilde{\tilde{\beta}}_2^{(2)}\right) =$$
$$= 0.405673^2\,(2,-2,-1)\begin{pmatrix} 0.062500, & 0, & 0\\ 0, & 0.038073, & 0\\ 0, & 0, & 0.000159\end{pmatrix}\begin{pmatrix}2\\-2\\-1\end{pmatrix} = 0.066232.$$

Thus

$$\mathrm{Var}\left(\tilde{\tilde{\beta}}_2^{(2)}\right) + \left(b_{\beta_2}^{(2)}\right)^2 = 0.066232 + 0.017546 = 0.083778$$

and $\mathrm{Var}(\widehat{\widehat{\beta}}_2) = 3$, what is essentialy larger than 0.083778.

The MSE of the estimator $\tilde{\tilde{\beta}}_2^{(2)}$ equal to 0.083778 is smaller than the MSE of the estimator $\tilde{\tilde{\beta}}_2^{(1)}$ equal to 0.109356.

References

[1] Fišerová, E., Kubáček, L., Kunderová, P.: Linear Statistical Models: Regularity and Singularities. *Academia*, Praha, 2007.

[2] Harville, D. A.: Matrix Algebra for a Statistician's Perspective. *Springer*, New York–Berlin–Heidelberg–Barcelona–Hong Kong–London–Milan–Paris–Singapore–Tokyo, 1997.

[3] Hoerl, A. E., Kennard, R. W.: *Ridge regression: Biased eastimation for nonorthogonal problems.* Technometrics **12** (1970), 55–67.

[4] Hoerl, A. E., Kennard, R. W.: *Ridge regression: Application to nonorthogonal problems.* Technometrics **12** (1970), 69–82.

[5] Goldstein, M., Smith, A. F. M.: *Ridge–type estimators for regression analysis.* Journal of the Royal Statistical Society, Series B **36** (1974), 284–291

[6] Rao, C. R., Mitra, S. K.: Generalized Inverse of Matrices and its Applications. *Wiley*, New York–London–Sydney–Toronto, 1971.

[7] Theobald, C. M.: *Generalizations of mean square error applied to ridge regression.* Journal of the Royal Statistical Society, Series B **36** (1974), 103–106.