Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 79–86
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Minimax estimation in linear regression with ellipsoidal constraints

Maciej Wilczyński*

Institute of Mathematics and Computer Science, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

Received 1 December 2004; accepted 28 September 2005
Available online 15 December 2005

Abstract

In the paper we consider estimation of the regression coefficients in a linear regression model $\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\epsilon}$ under ellipsoid constraints on the parameter space and the weighted squared error loss function. We prove that the minimax linear decision rule $\boldsymbol{d}^*$ is also minimax when the class $\mathscr{D}$ of possible estimators of the regression coefficients $\boldsymbol{\beta}$ is unrestricted. We derive that result assuming that the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, which define the loss and the constraints, have the same eigenvectors as the matrix $\boldsymbol{X}'\boldsymbol{X}$.

Keywords: Linear regression; Minimax estimator

1. Introduction

Suppose that

$$\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\epsilon}, \tag{1}$$

is the linear regression model in which $\boldsymbol{Y} \in R^n$ is the column vector of observations, $\boldsymbol{X} \in R^{n\times p}$ is the known deterministic design matrix of rank p, $\boldsymbol{\beta} \in R^p$ is the column vector of the unknown regression coefficients and $\boldsymbol{\epsilon} \in R^n$ is the column vector of the un-observable experimental errors, which has mean vector zero and known covariance matrix $\sigma^2\boldsymbol{I}_n$.

Let $\boldsymbol{A} \in R^{p\times p}$ and $\boldsymbol{B} \in R^{p\times p}$ be given positive definite and non-negative definite matrices and let ρ be a positive real number. In the paper we consider minimax estimation of the regression coefficients $\boldsymbol{\beta}$ in model (1) with a restricted parameter space

$$\Theta=\{\boldsymbol{\beta} : \boldsymbol{\beta}'\boldsymbol{B\beta}\leqslant\rho\}. \tag{2}$$

We assume that the loss in taking $\widehat{\boldsymbol{\beta}} \in R^p$ as an estimate of $\boldsymbol{\beta}$ has the form

$$L(\widehat{\boldsymbol{\beta}},\boldsymbol{\beta})=(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\boldsymbol{A}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}). \tag{3}$$

We estimate $\boldsymbol{\beta}$ on the basis of an observation $\boldsymbol{Y}$ by $\widehat{\boldsymbol{\beta}}(\boldsymbol{Y})$, where $\widehat{\boldsymbol{\beta}}$ is an arbitrary element of the set of all estimators

$$\mathscr{D}=\{\text{all Borel measurable transformation } \widehat{\boldsymbol{\delta}} : R^n \to R^p\}.$$

* Tel.: +48 71 3202103; fax: +48 71 3280751.
E-mail addresses: maciej.wilczynski@pwr.wroc.pl, wilczyn@im.pwr.wroc.pl (M. Wilczyński).

0378-3758/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2005.09.009

When the vector $\boldsymbol{\beta}$ is a priori known to be constrained it is then advisable to incorporate this information in the estimation procedure, because it may enhance the efficiency of the resulting estimators. One way to utilize this information is to use the minimax principle. Such an approach to estimating the regression coefficient in the linear model (1) under the weighted squared error loss function (3) was suggested by Kuks and Olman (1971, 1972). In their investigations the disturbance $\boldsymbol{\epsilon}$ has the mean vector zero, a known covariance matrix V and there is some prior information about $\boldsymbol{\beta}$. Then, this approach was considered by many authors, e.g. Läuter (1975), Hoffmann (1979), Pilz (1986), Stahlecker (1987), Trenkler and Stahlecker (1987), Gaffke and Heiligers (1989), Pilz (1991), Drygas (1993), Stahlecker and Trenkler (1993), Rao and Toutenburg (1995) and Blaker (2000). In each of these references the parameter space is defined as a known subset Θ of R^p and attention is restricted to the class of linear estimators

$$\mathscr{D}_L = \{\widehat{\boldsymbol{\delta}} \in \mathscr{D} : \widehat{\boldsymbol{\delta}}(\boldsymbol{y}) = \boldsymbol{C}\boldsymbol{y} \text{ for some matrix } \boldsymbol{C} \in R^{p\times n}\}.$$

Application of the minimax principle leads to a minimax linear estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}(\boldsymbol{Y})$ which satisfies

$$\inf_{\widehat{\beta}\in\mathscr{D}_L} \sup_{\substack{P\in\mathscr{P}\\ \beta\in R^p}} R(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}, P) = \sup_{\substack{P\in\mathscr{P}\\ \beta\in R^p}} R(\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}, \boldsymbol{\beta}, P),$$

where

$$R(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}, P) = E_P[L(\widehat{\boldsymbol{\beta}}(\boldsymbol{Y}), \boldsymbol{\beta})], \quad P \in \mathscr{P}, \quad \boldsymbol{\beta} \in \Theta$$

is the risk function of an estimator $\widehat{\boldsymbol{\beta}}$. Here $E_P(\cdot)$ denotes the expectation with respect to an unknown distribution $P \in \mathscr{P}$ of the random vector $\boldsymbol{\epsilon}$ and the class $\mathscr{P}$ is defined as

$$\mathscr{P} = \{P : P \text{ is a distribution of } \boldsymbol{\epsilon} \text{ for which } E_P(\boldsymbol{\epsilon}) = \boldsymbol{0},\ Cov_P(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n\}. \tag{4}$$

Unfortunately, only relatively few explicit solutions to such a minimax problem are available and frequently these results are derived under additional distributional or structural assumptions. Nevertheless, the simplicity of estimators that are linear in the data makes it natural to ask how much is lost by restricting attention to this class. It is well known, for example, that for $\Theta = R^p$ there is no loss of efficiency. In this special case the least-squares estimator of $\boldsymbol{\beta}$, given by

$$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}(\boldsymbol{Y}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

is minimax linear and remains minimax when the class of possible estimators of the regression coefficients is unrestricted, i.e. when it is defined as $\mathscr{D}$.

In the paper we prove that when $\Theta = \{\boldsymbol{\beta} : \boldsymbol{\beta}'\boldsymbol{B}\boldsymbol{\beta} \leqslant \rho\}$ then the resulting minimax linear estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$ is also minimax in $\mathscr{D}$ and satisfies

$$\inf_{\widehat{\beta}\in\mathscr{D}} \sup_{\substack{P\in\mathscr{P}\\ \beta\in\Theta}} R(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}, P) = \sup_{\substack{P\in\mathscr{P}\\ \beta\in\Theta}} R(\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}, \boldsymbol{\beta}, P),$$

provided that the rather restrictive assumption is fulfilled:

Condition 1. The matrices $\boldsymbol{X}'\boldsymbol{X}, \boldsymbol{A}$ and $\boldsymbol{B}$ have the same eigenvectors.

2. Main result

Let $\boldsymbol{M}$ be a symmetric matrix with a spectral decomposition $\boldsymbol{M} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}'$, in which $\boldsymbol{Q}$ is orthogonal and $\boldsymbol{\Lambda}$ is diagonal. We define $\boldsymbol{M}_+ = \boldsymbol{Q}\boldsymbol{\Lambda}_+\boldsymbol{Q}'$, where $\boldsymbol{\Lambda}_+$ is obtained by replacing each negative element of $\boldsymbol{\Lambda}$ by zero. The following theorem, proved by Blaker (2000), is not new but unifies previous separate approaches to the problem of minimax linear estimation in the special case in which Condition 1 holds (cf. Gaffke and Heiligers, 1989 or Hoffmann, 1979).

Theorem 1. *Consider the linear regression model* (1) *with the restricted parameter space* $\boldsymbol{\Theta}$ *defined by* (2) *and the loss function given by* (3). *If the matrices* $\boldsymbol{X}'\boldsymbol{X}, \boldsymbol{A}$ *and* $\boldsymbol{B}$ *have the same eigenvectors then the minimax linear estimator of* $\boldsymbol{\beta}$ *has the form*

$$\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}(\boldsymbol{Y}) = (\boldsymbol{I}_p - h\boldsymbol{B}^{1/2}\boldsymbol{A}^{-1/2})_+ \widehat{\boldsymbol{\beta}}_{\mathrm{LS}}(\boldsymbol{Y}), \tag{5}$$

where h satisfies

$$\sigma^2 \operatorname{tr}\{(X'X)^{-1}(h^{-1}A^{1/2}B^{1/2} - B)_+\} = \rho. \tag{6}$$

The minimax linear risk is

$$\inf_{\widehat{\beta}\in\mathscr{D}_L} \sup_{\substack{P\in\mathscr{P}\\ \beta\in\Theta}} R(\widehat{\beta}, \beta, P) = \sup_{\substack{P\in\mathscr{P}\\ \beta\in\Theta}} R(\widehat{\beta}_{\mathrm{ML}}, \beta, P) = \sigma^2 \operatorname{tr}\{(X'X)^{-1}(A - hA^{1/2}B^{1/2})_+\}.$$

To generalize the above result and prove that $\widehat{\beta}_{\mathrm{ML}}(Y)$ remains minimax in $\mathscr{D}$ we first transform the regression model to its canonical form where all matrices in question are diagonal (cf. Blaker, 2000). So, let the singular value decomposition of X be $X = U\tilde{D}V'$, where $U \in R^{n\times n}$, $V \in R^{p\times p}$ are orthogonal matrices and $\tilde{D} \in R^{n\times p}$ is a matrix with elements $d_1 \geqslant d_2 \geqslant \cdots \geqslant d_p > 0$ in positions $(1, 1), (2, 2), \ldots, (p, p)$ and zero everywhere else. Then

$$X'X = V\tilde{D}'\tilde{D}V' = VD^2V'$$

and $D = \operatorname{diag}(d_1, \ldots, d_p)$, with $d_i = \lambda_i^{1/2}$ for $1 \leqslant i \leqslant p$, where $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p > 0$ are eigenvalues of $X'X$. Setting $Z = U'Y$, $\gamma = V'\beta$, $\eta = EZ = U'X\beta$ and $\varepsilon = U'\epsilon$, we transform (1) to the canonical model

$$Z = \tilde{D}\gamma + \varepsilon, \quad E_{\tilde{P}}\varepsilon = 0, \ \operatorname{Cov}_{\tilde{P}}(\varepsilon) = \sigma^2 I_n, \tag{7}$$

where $\tilde{P} \in \mathscr{P}$ denotes a distribution of $\varepsilon = U'\epsilon$ when ϵ is distributed according to $P \in \mathscr{P}$ ($\tilde{P} \in \mathscr{P}$, because U is orthogonal). This means that the components Z_i of $Z = (Z_1, \ldots, Z_n)'$ satisfy the equation

$$Z_i = \begin{cases} d_i\gamma_i + \varepsilon_i & \text{for } i = 1, \ldots, p, \\ \varepsilon_i & \text{for } i = p+1, \ldots, n. \end{cases} \tag{8}$$

Moreover, Condition 1 implies that there exist diagonal matrices $\tilde{A} = \operatorname{diag}(\tilde{a}_1, \ldots, \tilde{a}_p)$ and $\tilde{B} = \operatorname{diag}(\tilde{b}_1, \ldots, \tilde{b}_p)$ for which $A = V\tilde{A}V'$ and $B = V\tilde{B}V'$. Define

$$\tilde{L}(\widehat{\gamma}, \gamma) = (\widehat{\gamma} - \gamma)'\tilde{A}(\widehat{\gamma} - \gamma) = \sum_{i=1}^{p} \tilde{a}_i(\widehat{\gamma}_i - \gamma_i)^2, \tag{9}$$

$$\Gamma = \left\{\gamma : \gamma'\tilde{B}\gamma = \sum_{i=1}^{p} \tilde{b}_i\gamma_i^2 \leqslant \rho\right\} \tag{10}$$

Then we have the following lemma:

Lemma 1. *Let $\widehat{\beta}$ be any rule in $\mathscr{D}$ and let $\widehat{\gamma}(z) = V'\widehat{\beta}(Uz)$, $z \in R^p$. Then $\widehat{\beta}(Y)$ is a minimax estimator of $\beta \in \Theta$ in the linear regression model* (1) *with the loss function $L(\widehat{\beta}, \beta)$ if and only if $\widehat{\gamma}(Z)$ is minimax for estimating $\gamma \in \Gamma$ in the canonical regression model* (7) *with the loss $\tilde{L}(\widehat{\gamma}, \gamma)$. Moreover, $\widehat{\beta}$ and $\widehat{\gamma}$ have the same minimax risk.*

Proof. First we note that for $\gamma = V'\beta$ and $Y = UZ$ we have

$$\beta'B\beta = \gamma'\tilde{B}\gamma,$$

$$L(\widehat{\beta}(Y), \beta) = (\widehat{\beta}(Y) - \beta)'A(\widehat{\beta}(Y) - \beta) = (\widehat{\gamma}(Z) - \gamma)'\tilde{A}(\widehat{\gamma}(Z) - \gamma) = \tilde{L}(\widehat{\gamma}(Z), \gamma).$$

The rest of the proof follows immediatetely from the fact that the risk function $\tilde{R}$ of $\widehat{\gamma}$ satisfies

$$\tilde{R}(\widehat{\gamma}, \gamma, \tilde{P}) = E_{\tilde{P}}\tilde{L}(\widehat{\gamma}(Z), \gamma) = E_P L(\widehat{\beta}(Y), \beta) = R(\widehat{\beta}, \beta, P),$$

where, as before, $\tilde{P}$ denotes a distribution of $\varepsilon = U'\epsilon$ when ϵ is distributed according to P. $\square$

Now we are ready to state the following theorem, which is the main result of the paper.

Theorem 2. *Consider the linear regression model* (1) *with the restricted parameter space* $\boldsymbol{\Theta}=\{\boldsymbol{\beta}:\boldsymbol{\beta}'\boldsymbol{B}\boldsymbol{\beta}\leqslant\rho\}$ *and the loss function given by* (3). *If the matrices* $\boldsymbol{X}'\boldsymbol{X}$, $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *have the same eigenvectors then the minimax linear estimator* $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}(\boldsymbol{Y})$, *defined by* (5) *and* (6), *is minimax in* $\mathscr{D}$. *The minimax risk is*

$$\inf_{\widehat{\beta}\in\mathscr{D}}\sup_{\substack{P\in\mathscr{P}\\ \beta\in\Theta}} R(\widehat{\boldsymbol{\beta}},\boldsymbol{\beta},P)=\sup_{\substack{P\in\mathscr{P}\\ \beta\in\Theta}} R(\widehat{\boldsymbol{\beta}}_{\mathrm{ML}},\boldsymbol{\beta},P)=\sigma^2\operatorname{tr}\{(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{A}-h\boldsymbol{A}^{1/2}\boldsymbol{B}^{1/2})_+\}.$$

Proof. Lemma 1 implies that to prove the theorem it suffices to show that $\widehat{\boldsymbol{\gamma}}_{\mathrm{ML}}(\boldsymbol{Z})\overset{\mathrm{def}}{=}\boldsymbol{V}'\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}(\boldsymbol{U}\boldsymbol{Z})$ is the minimax estimator of $\boldsymbol{\gamma}\in\Gamma$ in the canonical regresion model (7) with the loss function $\tilde{L}(\widehat{\boldsymbol{\gamma}},\boldsymbol{\gamma})$. For this we observe first that, by (5) and (6),

$$\widehat{\boldsymbol{\gamma}}_{\mathrm{ML}}(\boldsymbol{Z})=\boldsymbol{V}'\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}(\boldsymbol{U}\boldsymbol{Z})=(\boldsymbol{I}_p-h\tilde{\boldsymbol{B}}^{1/2}\tilde{\boldsymbol{A}}^{-1/2})_+\boldsymbol{D}^{-2}\tilde{\boldsymbol{D}}\boldsymbol{Z}=(c_1^*Z_1,\ldots,c_p^*Z_p)',$$

where

$$c_i^*=d_i^{-1}(1-h(\tilde{b}_i/\tilde{a}_i)^{1/2})_+\quad\text{for } 1\leqslant i\leqslant p.$$

Here h is determined from $\sum_{i=1}^p\tilde{b}_i\gamma_i^{*2}=\rho$ and $\boldsymbol{\gamma}^*=(\gamma_1^*,\ldots,\gamma_p^*)'$ is any point from Γ for which

$$\gamma_i^{*2}=d_i^{-2}\sigma^2((\tilde{a}_i/\tilde{b}_i)^{1/2}/h-1)_+\quad\text{if } \tilde{b}_i>0. \tag{11}$$

Note that

$$c_i^*=\begin{cases}\dfrac{d_i\gamma_i^{*2}}{\sigma^2+(d_i\gamma_i^*)^2} & \text{if } \tilde{b}_i>0,\\ d_i^{-1} & \text{if } \tilde{b}_i=0.\end{cases} \tag{12}$$

Moreover, as shown by Blaker (2000), the minimax linear risk of $\widehat{\boldsymbol{\gamma}}_{\mathrm{ML}}(\boldsymbol{Z})$ is

$$\begin{aligned}\sup_{\substack{P\in\mathscr{P}\\ \gamma\in\Gamma}}\tilde{R}(\widehat{\boldsymbol{\gamma}}_{\mathrm{ML}},\boldsymbol{\gamma},P)&=\sup_{\gamma\in\Gamma}\sum_{i=1}^p\tilde{a}_i[c_i^{*2}\sigma^2+(c_i^*d_i-1)^2\gamma_i^2]=\sum_{\{i:\tilde{b}_i=0\}}\tilde{a}_ic_i^{*2}\sigma^2\\&\quad+\sum_{\{i:\tilde{b}_i>0\}}\tilde{a}_i[c_i^{*2}\sigma^2+(c_i^*d_i-1)^2\gamma_i^{*2}]=\sigma^2\sum_{i=1}^p\tilde{a}_id_i^{-2}(1-h(\tilde{b}_i/\tilde{a}_i)^{1/2})_+.\end{aligned} \tag{13}$$

To prove minimaxity of $\widehat{\boldsymbol{\gamma}}_{\mathrm{ML}}$ in $\mathscr{D}$ we construct a sequence (π_k) of prior distributions on $\mathscr{P}$, for which the corresponding sequence of Bayes risks converges to the supremum of the risk of $\widehat{\boldsymbol{\gamma}}_{\mathrm{ML}}$. From this we deduce minimaxity. For this we fix $\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_p)'\in\Gamma$ and define

$$J(\boldsymbol{\gamma})=\{i\in\{1,\ldots,p\}:\gamma_i\neq 0\}.$$

Next we denote by P_γ a distribution of $\boldsymbol{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_n)'$ according to which $\varepsilon_1,\ldots,\varepsilon_n$ are independent random variables such that

$$P_\gamma\left(\varepsilon_i=\frac{\sigma^2}{d_i\gamma_i}\right)=1-P_\gamma(\varepsilon_i=-d_i\gamma_i)=\frac{(d_i\gamma_i)^2}{\sigma^2+(d_i\gamma_i)^2}\quad\text{if } i\in J(\gamma),$$

and

$$P_\gamma(\varepsilon_i=\sigma)=P_\gamma(\varepsilon_i=-\sigma)=1/2\quad\text{if } i\in\{1,\ldots,n\}\backslash J(\boldsymbol{\gamma}).$$

Then $P_\gamma\in\mathscr{P}$, because $E_{P_\gamma}(\boldsymbol{\varepsilon})=\boldsymbol{0}$ and $Cov_{P_\gamma}(\boldsymbol{\varepsilon})=\sigma^2\boldsymbol{I}_n$. Moreover, (8) implies that under P_γ components of the observed random vector $\boldsymbol{Z}=\tilde{\boldsymbol{D}}\boldsymbol{\gamma}+\boldsymbol{\varepsilon}$ are independent random variables with the two-point distributions given by

$$P_\gamma\left(Z_i=\frac{\sigma^2+(d_i\gamma_i)^2}{d_i\gamma_i}\right)=1-P_\gamma(Z_i=0)=\frac{(d_i\gamma_i)^2}{\sigma^2+(d_i\gamma_i)^2}\quad\text{if } i\in J(\boldsymbol{\gamma})$$

and

$$P_\gamma(Z_i=\sigma)=P_\gamma(Z_i=-\sigma)=1/2 \quad \text{if } i\in\{1,\dots,n\}\backslash J(\gamma).$$

Furthermore, we assume that the prior π_k chooses $\boldsymbol{\gamma}=(\gamma_1,\dots,\gamma_p)'\in\Gamma$, and hence $P_\gamma\in\mathscr{P}$, in such a way that $\gamma_1,\dots,\gamma_p$ are independent random variables with the distribution

$$\pi_k(\gamma_i=\gamma^*_{ki})=\pi_k(\gamma_i=-\gamma^*_{ki})=\begin{cases}1/2 & \text{if } \gamma^*_{ki}>0,\\ 1 & \text{if } \gamma^*_{ki}=0,\end{cases} \tag{14}$$

where $\boldsymbol{\gamma}^*_k=(\gamma^*_{k1},\dots,\gamma^*_{kp})'\in\Gamma$ is defined by

$$\gamma^*_{ki}=\begin{cases}\gamma^*_i & \text{if } \tilde{b}_i>0,\\ k & \text{if } \tilde{b}_i=0.\end{cases} \tag{15}$$

Note that for such a prior π_k we have $\gamma_i=0$ when $i\notin J(\boldsymbol{\gamma}^*_k)$. Moreover, when $i\in J(\boldsymbol{\gamma}^*_k)$, the random variable Z_i takes on two values: 0, z^*_{ki} if $\gamma_i=\gamma^*_{ki}$ and 0, $-z^*_{ki}$ if $\gamma_i=-\gamma^*_{ki}$, where $z^*_{ki}=(\sigma^2+(d_i\gamma^*_{ki})^2)/(d_i\gamma^*_{ki})$. Finally, the posterior distribution of γ_i, given $Z_i=0$, is symmetric about 0. Thus the π_k-Bayes estimator $\widehat{\boldsymbol{\gamma}}_k=(\widehat{\gamma}_{k1},\dots,\widehat{\gamma}_{kp})'$ of $\boldsymbol{\gamma}=(\gamma_1,\dots,\gamma_p)'$ has the form

$$\widehat{\gamma}_{ki}(\mathbf{Z})=\begin{cases}\gamma^*_{ki} & \text{if } Z_i=(\sigma^2+(d_i\gamma^*_{ki})^2)/(d_i\gamma^*_{ki}),\\ 0 & \text{if } Z_i=0,\\ -\gamma^*_{ki} & \text{if } Z_i=-(\sigma^2+(d_i\gamma^*_{ki})^2)/(d_i\gamma^*_{ki})\end{cases} \qquad \text{for } i\in J(\boldsymbol{\gamma}^*_k),$$

and

$$\widehat{\gamma}_{ki}(\mathbf{Z})=0 \quad \text{for } i\notin J(\boldsymbol{\gamma}^*_k).$$

Hence, using (15) and (12), the π_k-Bayes rule is simply $\widehat{\boldsymbol{\gamma}}_k(\mathbf{Z})=(c^*_{k1}Z_1,\dots,c^*_{kp}Z_p)'$, where

$$c^*_{ki}=\frac{d_i\gamma^{*2}_{ki}}{\sigma^2+(d_i\gamma^*_{ki})^2}=\begin{cases}\dfrac{d_ik^2}{\sigma^2+(d_ik)^2} & \text{if } \tilde{b}_i=0,\\ c^*_i & \text{if } \tilde{b}_i>0.\end{cases} \tag{16}$$

Moreover, the risk function of this decision rule is given by

$$\tilde{R}(\widehat{\boldsymbol{\gamma}}_k,\boldsymbol{\gamma},P)=E_P\tilde{L}(\widehat{\boldsymbol{\gamma}}_k,\boldsymbol{\gamma})=E_P\sum_{i=1}^p\tilde{a}_i(c^*_{ki}Z_i-\gamma_i)^2=\sum_{i=1}^p\tilde{a}_i[c^{*2}_{ki}\sigma^2+(c^*_{ki}d_i-1)^2\gamma_i^2]$$

and, by (14)–(16), the corresponding Bayes risk is

$$\begin{aligned}r(\widehat{\boldsymbol{\gamma}}_k,\pi_k)&=E_{\pi_k}\tilde{R}(\widehat{\boldsymbol{\gamma}}_k,\boldsymbol{\gamma},P)=\sum_{i=1}^p\tilde{a}_i[c^{*2}_{ki}\sigma^2+(c^*_{ki}d_i-1)^2\gamma^{*2}_{ki}]\\&=\sum_{\{i:\tilde{b}_i>0\}}\tilde{a}_i[c^{*2}_i\sigma^2+(c^*_id_i-1)^2\gamma^{*2}_i]+\sum_{\{i:\tilde{b}_i=0\}}\tilde{a}_i\frac{(\sigma k)^2}{\sigma^2+(d_ik)^2}.\end{aligned}$$

It is easily checked that, by (12) and (13),

$$\begin{aligned}\lim_{k\to\infty}r(\widehat{\boldsymbol{\gamma}}_k,\pi_k)&=\sum_{\{i:\tilde{b}_i>0\}}\tilde{a}_i[c^{*2}_i\sigma^2+(c^*_id_i-1)^2\gamma^{*2}_i]+\sum_{\{i:\tilde{b}_i=0\}}\tilde{a}_ic^{*2}_i\sigma^2\\&=\sup_{\substack{P\in\mathscr{P}\\ \gamma\in\Gamma}}\tilde{R}(\widehat{\boldsymbol{\gamma}}_{\mathrm{ML}},\boldsymbol{\gamma},P).\end{aligned}$$

Hence

$$\begin{aligned}\inf_{\widehat{\gamma}\in\mathscr{D}} \sup_{\substack{P\in\mathscr{P}\\ \gamma\in\Gamma}} \tilde{R}(\widehat{\gamma},\gamma,P) &\geqslant \lim_{k\to\infty} \inf_{\widehat{\gamma}\in\mathscr{D}} r(\widehat{\gamma},\pi_k) = \lim_{k\to\infty} r(\widehat{\gamma}_k,\pi_k) = \sup_{\substack{P\in\mathscr{P}\\ \gamma\in\Gamma}} \tilde{R}(\widehat{\gamma}_{\mathrm{ML}},\gamma,P)\\ &\geqslant \inf_{\widehat{\gamma}\in\mathscr{D}} \sup_{\substack{P\in\mathscr{P}\\ \gamma\in\Gamma}} \tilde{R}(\widehat{\gamma},\gamma,P),\end{aligned}$$

because $\sup_{\substack{P\in\mathscr{P}\\ \gamma\in\Gamma}} \tilde{R}(\widehat{\gamma},\gamma,P)\geqslant r(\widehat{\gamma},\pi_k)$ for all $k\geqslant 1$ and $\widehat{\gamma}\in\mathscr{D}$. This implies minimaxity of $\widehat{\gamma}_{\mathrm{ML}}$ in $\mathscr{D}$. □

3. The case of non-commuting matrices

We have proved the main result assuming that Condition 1 is satisfied. It is an open question whether there exist examples where the minimax linear estimator fails to be minimax in $\mathscr{D}$ if $\boldsymbol{X}'\boldsymbol{X}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ are non-commuting matrices. Below, we answer another question showing that $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$ may be minimax in $\mathscr{D}$ when Condition 1 does not hold.

Consider the linear regression model (1) in which $p=n$, $\rho=\sigma=1$, $\boldsymbol{X}=\boldsymbol{I}_n$, $\boldsymbol{B}=[b_{ij}]$ is a positive definite matrix and $\boldsymbol{A}=\boldsymbol{\Lambda}^{-1}\boldsymbol{B}\boldsymbol{\Lambda}^{-1}$, where $\boldsymbol{\Lambda}=\mathrm{diag}(\sqrt{b_{11}},\ldots,\sqrt{b_{nn}})$. Moreover, assume that

$$\min_{1\leqslant i\leqslant n} b_{ii}\geqslant 1 \quad\text{and}\quad \kappa>\max_{1\leqslant i\leqslant n}\sqrt{b_{ii}},\qquad \text{where } \kappa=\frac{\mathrm{tr}(\boldsymbol{B})+1}{\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{B})}. \tag{17}$$

Then, from Gaffke and Heiligers (1989), we can easily deduce that the minimax linear estimator of $\boldsymbol{\beta}$ has the form

$$\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}(\boldsymbol{Y})=(\boldsymbol{I}_n-\kappa^{-1}\boldsymbol{\Lambda})\boldsymbol{Y}=\mathrm{diag}\left(\frac{\alpha_1}{b_{11}+\alpha_1},\ldots,\frac{\alpha_n}{b_{nn}+\alpha_n}\right)\boldsymbol{Y}, \tag{18}$$

where

$$\alpha_i=\kappa\sqrt{b_{ii}}-b_{ii},\quad i=1,\ldots,n, \tag{19}$$

are positive numbers that sum to one (cf. (17)). Furthermore, the risk function of $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$, which is given by

$$\begin{aligned}R(\widehat{\boldsymbol{\beta}}_{\mathrm{ML}},\boldsymbol{\beta},P)&=\mathrm{tr}(\boldsymbol{I}_n-\kappa^{-1}\boldsymbol{\Lambda})'\boldsymbol{A}(\boldsymbol{I}_n-\kappa^{-1}\boldsymbol{\Lambda})+\kappa^{-2}\boldsymbol{\beta}'\boldsymbol{\Lambda}'\boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{\beta}\\ &=\mathrm{tr}(\boldsymbol{I}_n-\kappa^{-1}\boldsymbol{\Lambda})'\boldsymbol{A}(\boldsymbol{I}_n-\kappa^{-1}\boldsymbol{\Lambda})+\kappa^{-2}\boldsymbol{\beta}'\boldsymbol{B}\boldsymbol{\beta},\end{aligned}$$

attains its global maximum over $\mathscr{P}$ and Θ at any $\boldsymbol{\beta}\in\Theta$ which satisfies $\boldsymbol{\beta}'\boldsymbol{B}\boldsymbol{\beta}=1$. To prove minimaxity of $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$ in $\mathscr{D}$ we first show that this decision rule is Bayes with respect to some prior distribution π on $\mathscr{P}$. For this purpose we put

$$p_0=\frac{1}{2^{n-1}}\frac{1}{\kappa^2},\quad p_i=\frac{\alpha_i^2}{b_{ii}}\frac{1}{\kappa^2},\quad \boldsymbol{\beta}_i=(1/\sqrt{b_{ii}})\boldsymbol{e}_i,\quad i=1,\ldots,n,$$

where $\boldsymbol{e}_1=(1,0,\ldots,0)',\ldots,\boldsymbol{e}_n=(0,0,\ldots,1)'$ is the standard orthonormal base in R^n. Then $p_0,\ldots,p_n$ are positive numbers and $\boldsymbol{\beta}_i\in\Theta$, $i=1,\ldots,n$, because $\boldsymbol{\beta}_i'\boldsymbol{B}\boldsymbol{\beta}_i=1$. Finally we denote by Δ_i, $i=1,\ldots,n$, the subset of R^n containing 2^{n-1} points $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_n)$ such that $\xi_i=1$ and $\xi_j=\pm 1$ for all $j=1,\ldots,n$, $j\neq i$.

Now, let P_i, $i=1,\ldots,n$, be a distribution of the observed random vector $\boldsymbol{Y}=\boldsymbol{\beta}_i+\boldsymbol{\epsilon}$ according to which $\boldsymbol{Y}$ is discrete and its probability mass function is given by

$$P_i(\boldsymbol{Y}=\boldsymbol{y})=\begin{cases}1-(2^{n-1}p_0+p_i) & \text{if } \boldsymbol{y}=\boldsymbol{0},\\ p_i & \text{if } \boldsymbol{y}=(\kappa/\alpha_i)\boldsymbol{e}_i,\\ p_0 & \text{if } \boldsymbol{y}=\kappa(\sum_{j=1}^n \xi_j\boldsymbol{e}_j)\end{cases} \tag{20}$$

for all $\boldsymbol{\xi}=(\xi_1,\ldots,\xi_n)\in\Delta_i$. Note that this definition is consistent, because (17) implies that $\kappa\sqrt{b_{ii}}>b_{ii}\geqslant 1$ and thus

$$p_i+\sum_{\{\xi\in\Delta_i\}} p_0=p_i+2^{n-1}p_0=\frac{b_{ii}+\alpha_i^2}{b_{ii}\kappa^2}\leqslant\frac{b_{ii}+\alpha_i}{b_{ii}\kappa^2}=\frac{\kappa\sqrt{b_{ii}}}{b_{ii}\kappa^2}=\frac{1}{\kappa\sqrt{b_{ii}}}\leqslant 1.$$

Moreover, under P_i, $i=1,\dots,n$, the distribution of the error vector $\boldsymbol{\epsilon}=\boldsymbol{Y}-\boldsymbol{\beta}_i$ belongs to $\mathscr{P}$, because

$$E_{P_i}\boldsymbol{Y}=p_i(\kappa/\alpha_i)\boldsymbol{e}_i+p_0\sum_{\{\xi\in\Delta_i\}}\kappa\left(\sum_{j=1}^n\xi_j\boldsymbol{e}_j\right)=((\kappa/\alpha_i)p_i+\kappa 2^{n-1}p_0)\boldsymbol{e}_i=\boldsymbol{\beta}_i,$$

$$\begin{aligned}E_{P_i}\boldsymbol{Y}\boldsymbol{Y}'&=p_i(\kappa/\alpha_i)^2\boldsymbol{e}_i\boldsymbol{e}_i'+p_0\sum_{\{\xi\in\Delta_i\}}\kappa^2\left(\sum_{j=1}^n\xi_j\boldsymbol{e}_j\right)\left(\sum_{j=1}^n\xi_j\boldsymbol{e}_j\right)'\\&=p_i(\kappa/\alpha_i)^2\boldsymbol{e}_i\boldsymbol{e}_i'+2^{n-1}p_0\kappa^2\left(\sum_{j=1}^n\boldsymbol{e}_j\boldsymbol{e}_j'\right)=\boldsymbol{\beta}_i\boldsymbol{\beta}_i'+\boldsymbol{I}_n.\end{aligned}$$

Now, let the prior π chooses a distribution P of $\boldsymbol{Y}$, and hence the corresponding distribution from $\mathscr{P}$ of the error $\boldsymbol{\epsilon}=\boldsymbol{Y}-E_P\boldsymbol{Y}$, in such a way that

$$\pi(P=P_i)=\pi(P=\overline{P}_i)=\alpha_i/2,\quad i=1,\dots,n,$$

where $\overline{P}_i$ is a distribution of $-\boldsymbol{Y}$ when $\boldsymbol{Y}$ is distributed according to P_i, $i=1,\dots,n$. Then $E_{\overline{P}_i}\boldsymbol{Y}=-\boldsymbol{\beta}_i\in\Theta$ and $\overline{P}_i(\boldsymbol{Y}=\boldsymbol{y})=P_i(\boldsymbol{Y}=-\boldsymbol{y})$, $\boldsymbol{y}\in R^n$, $i=1,\dots,n$. Therefore, by (20), the π-Bayes estimator $\widehat{\boldsymbol{\beta}}_\pi(\boldsymbol{y})=E(\boldsymbol{\beta}\,|\,\boldsymbol{Y}=\boldsymbol{y})$ of $\boldsymbol{\beta}$ has the form

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_\pi(\boldsymbol{y})&=\frac{\sum_{i=1}^n\frac{\alpha_i}{2}[P_i(\boldsymbol{Y}=\boldsymbol{y})-P_i(\boldsymbol{Y}=-\boldsymbol{y})]\boldsymbol{\beta}_i}{\sum_{i=1}^n\frac{\alpha_i}{2}[P_i(\boldsymbol{Y}=\boldsymbol{y})+P_i(\boldsymbol{Y}=-\boldsymbol{y})}\\&=\begin{cases}\boldsymbol{0}&\text{if }\boldsymbol{y}=\boldsymbol{0},\\\xi_i\boldsymbol{\beta}_i&\text{if }\boldsymbol{y}=(\kappa/\alpha_i)\xi_i\boldsymbol{e}_i,\\\sum_{j=1}^n\xi_j\alpha_j\boldsymbol{\beta}_j&\text{if }\boldsymbol{y}=\kappa\left(\sum_{j=1}^n\xi_j\boldsymbol{e}_j\right)\end{cases}\end{aligned}$$

for all $\xi_1=\pm1,\dots,\xi_n=\pm1$. Hence, using (18) and (19), the π-Bayes rule is simply $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$. This implies that $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$ is minimax in $\mathscr{D}$, because $\pi(\omega_\pi)=1$, where ω_π is the set of points at which the risk function of $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$ takes on its maximum.

To complete the example, we note that it is always possible to choose a matrix $\boldsymbol{B}$ so that (17) holds and $\boldsymbol{A}=\boldsymbol{\Lambda}^{-1}\boldsymbol{B}\boldsymbol{\Lambda}^{-1}$, $\boldsymbol{B}$ are non-commuting matrices (and therefore have different sets of eigenvectors).

Acknowledgements

I would like to thank the referees whose comments led to many improvements.

References

Blaker, H., 2000. Minimax estimation in linear regression under restrictions. J. Statist. Plann. Inference 90, 35–55.

Drygas, H., 1993. Reparametrization methods in linear minimax estimation. In: Matusita, K., Puri, M.L., Hayakawa, T. (Eds.), Proceedings of the Third Pacific Area Statistical Conference. Zeist, Netherlands (VSP), pp. 87–95.

Gaffke, N., Heiligers, B., 1989. Bayes, admissible, and minimax linear estimator in linear models with restricted parameter space. Statistics 20, 487–508.

Hoffmann, K., 1979. Characterization of minimax linear estimators in linear regression. Math. Oper. Statist. Ser. Statist. 10, 19–26.

Läuter, H., 1975. A minimax linear estimator for linear parameters under restrictions in form of inequalities. Math. Oper. Statist. Ser. Statist. 6, 689–695.

Kuks, J., Olman, W., 1971. Minimax linear estimation of regression coefficient I. Izwestija Akademija Nauk Estonskoj SSR 20, 480–482 (in Russian).

Kuks, J., Olman, W., 1972. Minimax linear estimation of regression coefficient II. Izwestija Akademija Nauk Estonskoj SSR 21, 66–72 (in Russian).

Pilz, J., 1986. Minimax linear regression estimation with symmetric parameter restriction. J. Statist. Plann. Inference 13, 297–318.

Pilz, J., 1991. Bayesian Estimation and Experimental Design in Linear Regression Models. Wiley, New York.

Rao, C.R., Toutenburg, H., 1995. Linear Models, Least Squares and Alternatives. Springer, Heidelberg.
Stahlecker, P., 1987. A priori information und minimax-schätzung im linearen regressions modell, Mathematical System in Economics, vol. 108. Athenäum, Frankfurt (in German).
Stahlecker, P., Trenkler, G., 1993. Minimax estimation in linear regression with singular covariance structure and convex polyhedral constraints. J. Statist. Plann. Inference 13, 297–318.
Trenkler, G., Stahlecker, P., 1987. Quasi minimax estimation in the linear regression model. Statistics 36, 185–196.