PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 45, Fasc. 2 (2025), pp. 211–225
Published online 1.2.2026
doi:10.37190/0208-4147.00256

MINIMAX ESTIMATORS IN LINEAR MODELS WITH RESTRICTED PARAMETER SPACE

BY

MACIEJ KOWAL (WROCŁAW) AND MACIEJ WILCZYŃSKI (WROCŁAW)

Abstract. In the paper we consider estimation of the regression coefficients in a linear regression model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\beta}$ is assumed to lie in a given ellipsoid Θ. The estimation is performed using a weighted squared error loss function. Under specific conditions, we find an explicit formula for an estimator $\widehat{\boldsymbol{\beta}}_M$ of $\boldsymbol{\beta}$, which is minimax in the class $\mathcal{D}$ of all decision rules, not limited to linear estimators. We generalize the results of Wilczyński (2007), who found a minimax decision rule $\widehat{\boldsymbol{\beta}}_M$ under considerably more restrictive conditions.

2020 Mathematics Subject Classification: Primary 62H12; Secondary 62C20, 62J05, 62F30.

Key words and phrases: minimax estimator.

1. INTRODUCTION

Consider the linear regression model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.1}$$

where $\boldsymbol{Y} \in \mathbb{R}^n$ is the vector of observations, $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ is a known design matrix of full rank p, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of unknown regression coefficients, and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ is a random error vector with mean zero and known covariance matrix $\boldsymbol{I}_n$.

Let $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ and $\boldsymbol{B} \in \mathbb{R}^{p\times p}$ be fixed positive and nonnegative definite matrices, respectively.

Suppose we know in advance that the unknown vector $\boldsymbol{\beta}$ belongs to a given ellipsoidal set Θ, defined by

$$\Theta = \{\boldsymbol{\beta} : \boldsymbol{\beta}'\boldsymbol{B}\boldsymbol{\beta} \leqslant 1\}. \tag{1.2}$$

In this paper, we address the problem of finding a minimax estimator $\widehat{\boldsymbol{\beta}}_M$ for $\boldsymbol{\beta} \in \Theta$ under the weighted squared error loss function

$$L(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}) = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{A}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \tag{1.3}$$

© Probability and Mathematical Statistics, 2025

We estimate the unknown vector $\boldsymbol{\beta}$ based on an observation $\boldsymbol{Y}$ using an estimator $\widehat{\boldsymbol{\beta}}(\boldsymbol{Y})$, where $\widehat{\boldsymbol{\beta}}$ is an arbitrary element of the set of all estimators

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

The risk function of an estimator $\widehat{\boldsymbol{\beta}} \in \mathcal{D}$ is defined to be

$$R(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}, P) = \mathbb{E}_P[L(\widehat{\boldsymbol{\beta}}(\boldsymbol{Y}), \boldsymbol{\beta})], \quad P \in \mathcal{P},\ \boldsymbol{\beta} \in \Theta.$$

Here, $\mathbb{E}_P(\cdot)$ denotes the expectation with respect to an unknown distribution $P \in \mathcal{P}$ of the random vector $\boldsymbol{\varepsilon}$, and the class $\mathcal{P}$ is defined by

$$\mathcal{P} = \{P : P \text{ is a distribution of } \boldsymbol{\varepsilon} \text{ such that } \mathbb{E}_P(\boldsymbol{\varepsilon}) = \mathbf{0},\ \mathrm{Cov}_P(\boldsymbol{\varepsilon}) = \boldsymbol{I}_n\}.$$

A minimax estimator $\widehat{\boldsymbol{\beta}}_M$ minimizes the maximum risk $\sup_{P\in\mathcal{P},\,\beta\in\Theta} R(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}, P)$ over all estimators, i.e.

$$\inf_{\widehat{\beta}\in\mathcal{D}} \sup_{\substack{P\in\mathcal{P}\\ \beta\in\Theta}} R(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}, P) = \sup_{\substack{P\in\mathcal{P}\\ \beta\in\Theta}} R(\widehat{\boldsymbol{\beta}}_M, \boldsymbol{\beta}, P),$$

When it is known a priori that the vector $\boldsymbol{\beta}$ lies within a restricted parameter space Θ (not necessarily an ellipsoid), finding the minimax estimator $\widehat{\boldsymbol{\beta}}_M$ is generally difficult. As a result, this problem is often replaced by the simpler task of finding a minimax *linear* estimator $\widehat{\boldsymbol{\beta}}_{ML} \in \mathcal{D}_L$. Such an estimator minimizes the maximum risk among all linear estimators, that is,

$$\inf_{\widehat{\boldsymbol{\beta}}\in\mathcal{D}_L} \sup_{P\in\mathcal{P},\boldsymbol{\beta}\in\Theta} R(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}, P) = \sup_{P\in\mathcal{P},\boldsymbol{\beta}\in\Theta} R(\widehat{\boldsymbol{\beta}}_{ML}, \boldsymbol{\beta}, P).$$

The class of linear decision rules $\mathcal{D}_L$ is defined by

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

The use of the minimax linear principle to estimate the regression coefficient in the linear model (1.1) under the weighted squared error loss function (1.3) was proposed by Kuks and Olman (1971, 1972). In their work, the disturbance $\boldsymbol{\varepsilon}$ has mean zero and a known covariance matrix, and some prior information about $\boldsymbol{\beta}$ is available. This approach was subsequently studied by several authors, including 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 all of these studies, the parameter space is assumed to be a known compact subset $\Theta \subset \mathbb{R}^p$, and the focus is restricted to the class $\mathcal{D}_L$ of linear estimators.

Although the risk function of a linear estimator is relatively simple under the loss (1.3), determining the admissible linear or minimax linear estimators of $\boldsymbol{\beta}$

remains challenging when Θ is a known compact subset of $\mathbb{R}^p$. This issue was comprehensively analyzed by Gaffke and Heiligers (1989).

Using Theorem 3 and Lemma 3 of Gaffke and Heiligers (1989), we obtain the following result, whose proof is postponed to the Appendix.

PROPOSITION 1.1. *Consider the linear regression model* (1.1) *with the restricted parameter space* (1.2) *and the loss function* (1.3). *Let the number* κ *and the matrices* $\boldsymbol{E}$, $\boldsymbol{H}$ *be defined by*

$$\kappa = \frac{1+\operatorname{tr}(\boldsymbol{H}\boldsymbol{E})}{\operatorname{tr}((\boldsymbol{H}^{1/2}\boldsymbol{E}^2\boldsymbol{H}^{1/2})^{1/2})},$$
$$\boldsymbol{E} = \boldsymbol{A}^{1/2}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}^{1/2}, \quad \boldsymbol{H} = \boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2}.$$

If

$$\boldsymbol{N}^* := \kappa \cdot \boldsymbol{H}^{-1/2}(\boldsymbol{H}^{1/2}\boldsymbol{E}^2\boldsymbol{H}^{1/2})^{1/2}\boldsymbol{H}^{-1/2} - \boldsymbol{E} \succeq \boldsymbol{0}, \tag{1.4}$$

that is, if $\boldsymbol{N}^*$ *is nonnegative definite, then the minimax linear estimator of* $\boldsymbol{\beta}$ *is*

$$\widehat{\boldsymbol{\beta}}_{ML}(\boldsymbol{Y}) = \boldsymbol{A}^{-1/2}\boldsymbol{N}^*(\boldsymbol{N}^*+\boldsymbol{E})^{-1}\boldsymbol{E}\boldsymbol{A}^{-1/2}\boldsymbol{X}'\boldsymbol{Y}. \tag{1.5}$$

REMARK 1.1. To the best of our knowledge, the explicit form of the minimax linear estimator $\widehat{\boldsymbol{\beta}}_{ML}(\boldsymbol{Y})$ is not known when the matrix $\boldsymbol{N}^*$ is not nonnegative definite and the matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{X}'\boldsymbol{X}$ do not commute.

Minimax linear estimation of $\boldsymbol{\beta} \in \Theta$ under the loss (1.3) becomes considerably simpler when the matrices $\boldsymbol{X}'\boldsymbol{X}$, $\boldsymbol{A}$, and $\boldsymbol{B}$ have the same eigenvectors (and are thus commuting matrices) and Θ is an ellipsoid. In such cases, Blaker (2000) derived an explicit form of the minimax linear decision rule $\widehat{\boldsymbol{\beta}}_{ML}$ (also in cases not resulting from Lemma 3 of Gaffke and Heilligers (1989)). Wilczyński (2007) generalized this result, proving that the minimax linear decision rule $\widehat{\boldsymbol{\beta}}_{ML}$ found by Blaker is minimax among all estimators, i.e. $\widehat{\boldsymbol{\beta}}_{M} = \widehat{\boldsymbol{\beta}}_{ML}$.

The assumption that $\boldsymbol{X}'\boldsymbol{X}$, $\boldsymbol{A}$, and $\boldsymbol{B}$ are commuting matrices is quite restrictive. However, the example provided by Wilczyński (2007) demonstrates that when the loss is (1.3) and Θ is an ellipsoid, the minimax linear estimator $\widehat{\boldsymbol{\beta}}_{ML}$ can still be minimax in $\mathcal{D}$ even when this assumption does not hold.

The above example raises the following question regarding the estimation of the vector $\boldsymbol{\beta} \in \{\boldsymbol{\beta} : \boldsymbol{\beta}'\boldsymbol{B}\boldsymbol{\beta} \leqslant 1\}$ in the linear regression model (1.1) with the loss function (1.3): Is the minimax linear estimator $\widehat{\boldsymbol{\beta}}_{ML}$ also minimax within the class of all estimators, not only the linear ones?

In this paper, we provide a partial answer to this question. Specifically, under the assumption that (1.4) holds, we present a sufficient (but not necessary) condition to ensure that $\widehat{\boldsymbol{\beta}}_{M} = \widehat{\boldsymbol{\beta}}_{ML}$, even when $\boldsymbol{X}'\boldsymbol{X}$, $\boldsymbol{A}$, and $\boldsymbol{B}$ are noncommuting matrices.

2. MAIN RESULTS

Consider the linear regression model

$$\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad \mathbb{E}_P\boldsymbol{\epsilon} = \boldsymbol{0}, \operatorname{Cov}_P(\boldsymbol{\epsilon}) = \boldsymbol{I}_n, \tag{2.1}$$

where $\boldsymbol{D} = [d_{ij}]$ is the $n \times p$ matrix with all elements zero except $d_{11} = \cdots = d_{pp} = 1$.

The following result is crucial to this paper. Its proof is deferred to the Appendix.

PROPOSITION 2.1. *There exist a matrix $\tilde{\boldsymbol{B}} \in \mathbb{R}^{p\times p}$ and a diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{p\times p}$, both positive definite, such that the minimax estimation problem in the model* (1.1) *is equivalent to finding the minimax estimator $\widehat{\boldsymbol{\gamma}}(\boldsymbol{Z})$ of $\boldsymbol{\gamma} \in \Gamma = \{\boldsymbol{\gamma} : \boldsymbol{\gamma}'\tilde{\boldsymbol{B}}\boldsymbol{\gamma} \leqslant 1\}$ under the loss function $\tilde{L}(\widehat{\boldsymbol{\gamma}}, \boldsymbol{\gamma}) = (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ in the linear model* (2.1).

The main result of the paper is as follows.

THEOREM 2.1. *Consider the linear regression model* (2.1) *with the restricted parameter space $\Gamma = \{\boldsymbol{\gamma} : \boldsymbol{\gamma}'\tilde{\boldsymbol{B}}\boldsymbol{\gamma} \leqslant 1\}$ and the loss function*

$$\tilde{L}(\widehat{\boldsymbol{\gamma}}, \boldsymbol{\gamma}) = (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}),$$

where $\tilde{\boldsymbol{B}} = [b_{ij}] \in \mathbb{R}^{p\times p}$ and $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$ are positive definite matrices. Define

$$\kappa = \frac{\operatorname{tr}(\tilde{\boldsymbol{B}}) + 1}{\operatorname{tr}(\boldsymbol{\Lambda}\tilde{\boldsymbol{B}})}. \tag{2.2}$$

If the following two conditions hold:

(C1) *the matrix $\kappa\boldsymbol{\Lambda} - \boldsymbol{I}_p$ is nonnegative definite,*

(C2) $$\max_{1\leqslant i\leqslant p} \frac{(\kappa\lambda_i - 1)^2}{(\kappa\lambda_i)^2} + \max_{1\leqslant i\leqslant p} \frac{1}{(\kappa\lambda_i)^2 b_{ii}} \leqslant 1,$$

then the minimax linear estimator $\widehat{\boldsymbol{\gamma}}_{ML}$ of $\boldsymbol{\gamma}$ is also minimax among all decision rules and

$$\widehat{\boldsymbol{\gamma}}_M(\boldsymbol{Z}) = \widehat{\boldsymbol{\gamma}}_{ML}(\boldsymbol{Z}) = (\boldsymbol{I}_p - \kappa^{-1}\boldsymbol{\Lambda}^{-1})\boldsymbol{D}'\boldsymbol{Z}. \tag{2.3}$$

Proof. Using Gaffke's result (cf. Gaffke and Heiligers (1989), Lemma 3), we first observe that $\widehat{\boldsymbol{\gamma}}_{ML}$ is a minimax linear estimator under condition (C1). Then, under additional condition (C2), we construct a prior distribution π on $\mathcal{P}$ with respect to which $\widehat{\boldsymbol{\gamma}}_{ML}$ is a Bayes decision rule. Finally, we deduce the minimaxity of $\widehat{\boldsymbol{\gamma}}_{ML}$ in $\mathcal{D}$ from the fact that $\pi(\omega_\pi) = 1$, where ω_π denotes the set of all points at which the risk function of $\widehat{\boldsymbol{\gamma}}_{ML}$ attains its maximum. The details of the argument are provided in the Appendix. ∎

3. EXAMPLES

In this subsection we will illustrate Theorem 2.1 with two examples in which the nondiagonal matrices $\tilde{\boldsymbol{B}} = [b_{ij}]$ and the diagonal matrices $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$ satisfy both conditions (C1) and (C2). Consequently, the minimax estimator $\widehat{\boldsymbol{\gamma}}_M$ of $\boldsymbol{\gamma}$ is given by (2.3). In each case, the formula for $\widehat{\boldsymbol{\gamma}}_M$ cannot be deduced from the result of Wilczyński (2007), because $\tilde{\boldsymbol{B}}$, $\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}$ and $\boldsymbol{D}'\boldsymbol{D}$ are not commuting matrices. Without loss of generality, we assume throughout that $\lambda_1 \geqslant \cdots \geqslant \lambda_p = 1$. Note that, under this assumption, condition (C1) can be rewritten in a more readable form as (see (2.2))

$$\text{(C1)} \quad \sum_{i=1}^{p} b_{ii}(\lambda_i - 1) \leqslant 1.$$

Furthermore, since $\lambda_1 = \max_{1 \leqslant i \leqslant p} \lambda_i$, condition (C2) takes the form

$$\text{(C2)} \quad \frac{(\kappa\lambda_1 - 1)^2}{(\kappa\lambda_1)^2} + \max_{1 \leqslant i \leqslant p} \frac{1}{(\kappa\lambda_i)^2 b_{ii}} \leqslant 1.$$

REMARK 3.1. For fixed positive values of $b_{11}, \ldots, b_{pp}$, condition (C1) is satisfied when the numbers $\lambda_1 \geqslant \cdots \geqslant \lambda_p = 1$ are sufficiently close to 1. For example, when $\lambda_1 - 1 \leqslant 1/\sum_{i=1}^{p-1} b_{ii}$, we have $\sum_{i=1}^{p} b_{ii}(\lambda_i - 1) \leqslant 1$, because

$$\sum_{i=1}^{p} b_{ii}(\lambda_i - 1) = \sum_{i=1}^{p-1} b_{ii}(\lambda_i - 1) \leqslant (\lambda_1 - 1) \sum_{i=1}^{p-1} b_{ii}.$$

EXAMPLE 3.1. Let $\tilde{\boldsymbol{B}} = [b_{ij}] \in \mathbb{R}^{p \times p}$ be a positive definite matrix such that $\min_{1 \leqslant i \leqslant p} b_{ii} \geqslant 1$. Furthermore, let $\lambda_1 \geqslant \cdots \geqslant \lambda_p = 1$ be real numbers such that condition (C1) is satisfied (see Remark 3.1). If we assume additionally that $\lambda_1^2 \leqslant \min_{1 \leqslant i \leqslant p} b_{ii}$ then condition (C2) also holds, since

$$\begin{aligned}
&\frac{(\kappa\lambda_1 - 1)^2}{(\kappa\lambda_1)^2} + \max_{1 \leqslant i \leqslant p} \frac{1}{(\kappa\lambda_i)^2 b_{ii}} \leqslant \frac{(\kappa\lambda_1 - 1)^2}{(\kappa\lambda_1)^2} + \max_{1 \leqslant i \leqslant p} \frac{1}{(\kappa\lambda_i)^2} \max_{1 \leqslant i \leqslant p} \frac{1}{b_{ii}} \\
&\quad = \frac{(\kappa\lambda_1 - 1)^2}{(\kappa\lambda_1)^2} + \frac{1}{(\kappa\lambda_p)^2} \frac{1}{\min_{1 \leqslant i \leqslant p} b_{ii}} = \frac{(\kappa\lambda_1 - 1)^2}{(\kappa\lambda_1)^2} + \frac{1}{(\kappa\lambda_1)^2} \frac{\lambda_1^2}{\min_{1 \leqslant i \leqslant p} b_{ii}} \\
&\quad \leqslant \frac{(\kappa\lambda_1 - 1)^2}{(\kappa\lambda_1)^2} + \frac{1}{(\kappa\lambda_1)^2} = \frac{(\kappa\lambda_1)^2 - 2\kappa\lambda_1 + 2}{(\kappa\lambda_1)^2} \leqslant \frac{(\kappa\lambda_1)^2}{(\kappa\lambda_1)^2} = 1.
\end{aligned}$$

Here, the next to last inequality follows from the fact that $\kappa\lambda_1 \geqslant 1$, because the matrix $\kappa\boldsymbol{\Lambda} - \boldsymbol{I}_p$ is nonnegative definite.

In the previous example, each of the values $b_{11}, \ldots, b_{pp}$ was assumed to be greater than or equal to 1. In the next example, some of these values may be slightly less than 1.

EXAMPLE 3.2. Let $b_{11}, \dots, b_{pp}$ be positive real numbers such that

$$-2\kappa_0 + 1 + \frac{1}{b_{kk}} \leqslant 0, \tag{3.1}$$

where $b_{kk} = \min_{1 \leqslant i \leqslant p} b_{ii}$, and the number $\kappa_0 > 1$ is defined by

$$\kappa_0 := \frac{b_{11} + \cdots + b_{pp} + 1}{b_{11} + \cdots + b_{pp}}.$$

Of course, inequality (3.1) holds when $b_{kk} \geqslant 1$, because $\kappa_0 > 1$. Nevertheless, (3.1) may also be satisfied when $b_{kk} < 1$, provided that

$$b_{kk} \in [1 - 2/p, 1) \quad \text{and} \quad p b_{kk} \leqslant b_{11} + \cdots + b_{pp} \leqslant \frac{2 b_{kk}}{1 - b_{kk}}. \tag{3.2}$$

Now, let $\tilde{\boldsymbol{B}} = [b_{ij}] \in \mathbb{R}^{p \times p}$ be a positive definite matrix whose diagonal elements either satisfy condition (3.2) or are all greater than or equal to 1 (see Remark 3.2 below). Furthermore, let $\lambda_1 \geqslant \cdots \geqslant \lambda_p = 1$ be real numbers such that condition (C1) is satisfied (see Remark 3.1). Assume additionally that

$$-2\kappa_0 + 1 + \frac{\lambda_1^2}{b_{kk}} \leqslant 0. \tag{3.3}$$

(Clearly, $\lambda_1 = \cdots = \lambda_p = 1$ when equality holds in inequality (3.1).) Then condition (C2) is also satisfied, since

$$\begin{aligned}
&\frac{(\kappa\lambda_1 - 1)^2}{(\kappa\lambda_1)^2} + \max_{1 \leqslant i \leqslant p} \frac{1}{(\kappa\lambda_i)^2 b_{ii}} \leqslant \frac{(\kappa\lambda_1 - 1)^2}{(\kappa\lambda_1)^2} + \max_{1 \leqslant i \leqslant p} \frac{1}{(\kappa\lambda_i)^2} \max_{1 \leqslant i \leqslant p} \frac{1}{b_{ii}} \\
&= \frac{(\kappa\lambda_1 - 1)^2}{(\kappa\lambda_1)^2} + \frac{1}{(\kappa\lambda_p)^2} \frac{1}{b_{kk}} = \frac{(\kappa\lambda_1 - 1)^2}{(\kappa\lambda_1)^2} + \frac{1}{(\kappa\lambda_1)^2} \frac{\lambda_1^2}{b_{kk}} \\
&= \frac{(\kappa\lambda_1)^2 - 2\kappa\lambda_1 + 1 + \lambda_1^2/b_{kk}}{(\kappa\lambda_1)^2} \leqslant \frac{(\kappa\lambda_1)^2 - 2\kappa_0 + 1 + \lambda_1^2/b_{kk}}{(\kappa\lambda_1)^2} \leqslant \frac{(\kappa\lambda_1)^2}{(\kappa\lambda_1)^2} = 1.
\end{aligned}$$

Here, the next to last inequality follows from the assumption (3.3) and the fact that

$$\kappa = \frac{b_{11} + \cdots + b_{pp} + 1}{b_{11}\lambda_1 + \cdots + b_{pp}\lambda_p} \geqslant \frac{b_{11} + \cdots + b_{pp} + 1}{\lambda_1 (b_{11} + \cdots + b_{pp})} = \frac{\kappa_0}{\lambda_1}.$$

REMARK 3.2. The class of non-diagonal positive definite matrices satisfying conditions described in the previous example is nonempty. It includes, for instance, every strictly (row) diagonally dominant matrix $\tilde{\boldsymbol{B}} = [b_{ij}] \in \mathbb{R}^{p \times p}$ whose diagonal elements either satisfy condition (3.2) or are all greater than or equal to 1.

4. CONCLUDING REMARKS

In Theorem 2.1, we showed that under condition (C1) the minimax linear estimator $\widehat{\boldsymbol{\gamma}}_{ML}$ remains minimax over the class of all estimators, provided the rather restrictive condition (C2) also holds. This raises the question: can the conclusion of this theorem be extended to the case where only (C1) holds? We believe the answer is affirmative.

However, we expect that proving such a generalization will require constructing a prior distribution whose support is larger than that of the least-favorable prior π used in the original proof. Recall that this prior π, with respect to which $\widehat{\boldsymbol{\gamma}}_{ML}$ is a Bayes estimator, has the following properties. It chooses specific discrete distributions $P_1, \overline{P}_1, \ldots, P_p, \overline{P}_p$ of the random vector $\boldsymbol{Z}$ under which

- $\mathrm{Cov}_{P_i}\boldsymbol{Z} = \boldsymbol{I}_p$ and $\mathbb{E}_{P_i}\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{\gamma}_i := 1/\sqrt{b_{ii}})\boldsymbol{e}_i$,
- $\overline{P}_i$ is the distribution of $-\boldsymbol{Z}$ when $\boldsymbol{Z}$ is distributed according to P_i.

In the minimax estimation problem described in Theorem 2.1, the unknown vector $\boldsymbol{\gamma}$ lies in the set $\Gamma = \{\boldsymbol{\gamma} : \boldsymbol{\gamma}'\tilde{\boldsymbol{B}}\boldsymbol{\gamma} \leqslant 1\}$. Thus, the prior distribution π is quite limited, as it takes into account only the $2p$ distributions P of $\boldsymbol{Z}$ for which $\mathbb{E}_P\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{\gamma}$ with

$$\boldsymbol{\gamma} \in \Gamma_0 := \left\{ \pm\frac{\boldsymbol{e}_1}{\sqrt{b_{11}}}, \pm\frac{\boldsymbol{e}_2}{\sqrt{b_{22}}}, \ldots, \pm\frac{\boldsymbol{e}_p}{\sqrt{b_{pp}}} \right\} \subset \Gamma.$$

However, the following fact is worth noting. Since $\widehat{\boldsymbol{\gamma}}_{ML}$ is Bayes with respect to this sparse, least-favorable prior π, we immediately obtain:

COROLLARY 4.1. *Theorem* 2.1 *remains valid even when the uncountable parameter space* $\Gamma = \{\boldsymbol{\gamma} : \boldsymbol{\gamma}'\tilde{\boldsymbol{B}}\boldsymbol{\gamma} \leqslant 1\}$ *is replaced by its finite subset* Γ_0 *defined above.*

REMARK 4.1. The results of the paper can be applied to a model slightly more general than (1.1), in which the covariance matrix of the observed vector $\boldsymbol{Y}$ is a known positive definite matrix $\boldsymbol{V}$, that is,

$$\mathbb{E}\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}, \quad \mathrm{Cov}(\boldsymbol{Y}) = \boldsymbol{V}.$$

This model can be reduced to the simpler case (1.1) by defining $\tilde{\boldsymbol{Y}} := \boldsymbol{V}^{-1/2}\boldsymbol{Y}$, $\tilde{\boldsymbol{X}} := \boldsymbol{V}^{-1/2}\boldsymbol{X}$, which gives

$$\mathbb{E}\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta}, \quad \mathrm{Cov}(\tilde{\boldsymbol{Y}}) = \boldsymbol{I}.$$

5. APPENDIX

5.1. Result of Gaffke and Heiligers (1989). For completeness, we recall the following result that is used to prove Proposition 1.1. We state it using the authors' original notation.

Let $\boldsymbol{Y}$ denote the n-dimensional vector of observations and let

$$\mathbb{E}\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{a}, \quad \operatorname{Cov}(\boldsymbol{Y}) = \boldsymbol{V}, \tag{5.1}$$

where $\boldsymbol{X}$ is a given $n \times p$ matrix of full rank p and $\boldsymbol{V}$ is a given positive definite $n \times n$ matrix. Suppose that the unknown vector $\boldsymbol{a}$ is constrained to the set

$$\boldsymbol{a} \in \Omega := \{\boldsymbol{\omega} \in \mathbb{R}^p : \boldsymbol{\omega}'\boldsymbol{H}\boldsymbol{\omega} \leqslant 1\},$$

where $\boldsymbol{H}$ is a positive definite matrix, and that the loss function has the form

$$L(\widehat{\boldsymbol{a}}, \boldsymbol{a}) = (\widehat{\boldsymbol{a}} - \boldsymbol{a})'(\widehat{\boldsymbol{a}} - \boldsymbol{a}).$$

Let $\boldsymbol{E} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$. In this setting, Gaffke and Heiligers (1989) proved (see Theorem 3 and Lemma 3 therein) that if the matrix

$$\boldsymbol{N}^* := \frac{1 + \operatorname{tr}(\boldsymbol{H}\boldsymbol{E})}{\operatorname{tr}\{(\boldsymbol{H}^{1/2}\boldsymbol{E}^2\boldsymbol{H}^{1/2})^{1/2}\}}\, \boldsymbol{H}^{-1/2}(\boldsymbol{H}^{1/2}\boldsymbol{E}^2\boldsymbol{H}^{1/2})^{1/2}\boldsymbol{H}^{-1/2} - \boldsymbol{E} \tag{5.2}$$

is nonnegative definite, then the minimax linear estimator of $\boldsymbol{a} \in \Omega$ is given by

$$\widehat{\boldsymbol{a}}_{ML}(\boldsymbol{Y}) = \boldsymbol{N}^*(\boldsymbol{E} + \boldsymbol{N}^*)^{-1}\boldsymbol{E}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Y}. \tag{5.3}$$

5.2. Proofs.

Proof of Proposition 1.1. For any $\boldsymbol{\beta}, \widehat{\boldsymbol{\beta}} \in \mathbb{R}^p$, define $\boldsymbol{a} := \boldsymbol{A}^{1/2}\boldsymbol{\beta}$ and $\widehat{\boldsymbol{a}} := \boldsymbol{A}^{1/2}\widehat{\boldsymbol{\beta}}$. Since $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{A}^{-1/2}\boldsymbol{a}$, the linear model (1.1) can be rewritten under this reparametrization as

$$\mathbb{E}\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A}^{-1/2}\boldsymbol{a}, \quad \operatorname{Cov}(\boldsymbol{Y}) = \boldsymbol{I}_n.$$

In the notation of Gaffke and Heiligers, the loss function (1.3) takes the form

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{A}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (\boldsymbol{A}^{1/2}\widehat{\boldsymbol{\beta}} - \boldsymbol{A}^{1/2}\boldsymbol{\beta})'(\boldsymbol{A}^{1/2}\widehat{\boldsymbol{\beta}} - \boldsymbol{A}^{1/2}\boldsymbol{\beta}) = (\widehat{\boldsymbol{a}} - \boldsymbol{a})'(\widehat{\boldsymbol{a}} - \boldsymbol{a}).$$

Moreover, with $\boldsymbol{H} := \boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2}$, the ellipsoidal constraint (1.2) is given by

$$\{\boldsymbol{\beta} : \boldsymbol{\beta}'\boldsymbol{B}\boldsymbol{\beta} \leqslant 1\} = \{\boldsymbol{a} : (\boldsymbol{A}^{-1/2}\boldsymbol{a})'\boldsymbol{B}\boldsymbol{A}^{-1/2}\boldsymbol{a} \leqslant 1\} = \{\boldsymbol{a} : \boldsymbol{a}'\boldsymbol{H}\boldsymbol{a} \leqslant 1\}.$$

To complete the proof of Proposition 1.1, it suffices to apply the result of Gaffke and Heiligers (cf. (5.1)–(5.3)) with $\boldsymbol{X} := \boldsymbol{X}\boldsymbol{A}^{-1/2}$, $\boldsymbol{H} := \boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2}$, $\boldsymbol{V} := \boldsymbol{I}_n$, noting that $\widehat{\boldsymbol{\beta}}_{ML} = \boldsymbol{A}^{-1/2}\widehat{\boldsymbol{a}}_{ML}$. ∎

Proof of Proposition 2.1. Let the singular value decomposition of $\boldsymbol{X}$ be $\boldsymbol{X} = \boldsymbol{U}\tilde{\boldsymbol{D}}\boldsymbol{V}'$, where $\boldsymbol{U} \in \mathbb{R}^{n\times n}$, $\boldsymbol{V} \in \mathbb{R}^{p\times p}$ are unitary matrices and $\tilde{\boldsymbol{D}} = [\tilde{d}_{ij}] \in \mathbb{R}^{n\times p}$

is a matrix with $\tilde{d}_{11} \geqslant \cdots \geqslant \tilde{d}_{pp} > 0$ and $d_{ij} = 0$ for all distinct i, j. Then

$$\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{V}\tilde{\boldsymbol{D}}'\tilde{\boldsymbol{D}}\boldsymbol{V}' = \boldsymbol{V}\boldsymbol{L}^2\boldsymbol{V}',$$

where $\boldsymbol{L} = \operatorname{diag}(\tilde{d}_{11}, \ldots, \tilde{d}_{pp})$. Setting $\tilde{\boldsymbol{Z}} = \boldsymbol{U}'\boldsymbol{Y}$, $\tilde{\boldsymbol{\gamma}} = \boldsymbol{V}'\boldsymbol{\beta}$, and $\tilde{\boldsymbol{\epsilon}} = \boldsymbol{U}'\boldsymbol{\epsilon}$, we transform (1.1) into the linear model

$$\tilde{\boldsymbol{Z}} = \tilde{\boldsymbol{D}}\tilde{\boldsymbol{\gamma}} + \tilde{\boldsymbol{\epsilon}}, \quad \mathbb{E}_{\tilde{P}}\tilde{\boldsymbol{\epsilon}} = \boldsymbol{0},\ \operatorname{Cov}_{\tilde{P}}(\tilde{\boldsymbol{\epsilon}}) = \boldsymbol{I}_n, \tag{5.4}$$

where $\tilde{P} \in \mathcal{P}$ denotes the distribution of $\tilde{\boldsymbol{\epsilon}} = \boldsymbol{U}'\boldsymbol{\epsilon}$ when $\boldsymbol{\epsilon}$ is distributed according to $P \in \mathcal{P}$ ($\tilde{P} \in \mathcal{P}$, because $\boldsymbol{U}$ is orthogonal). Under these assumptions, Wilczyński proved the following lemma (see Wilczyński (2007), Lemma 1):

LEMMA 5.1. *Let $\widehat{\boldsymbol{\beta}}$ be any rule in $\mathcal{D}$ and define $\widehat{\boldsymbol{\gamma}}(\tilde{\boldsymbol{z}}) := \boldsymbol{V}'\widehat{\boldsymbol{\beta}}(\boldsymbol{U}\tilde{\boldsymbol{z}}),\ \tilde{\boldsymbol{z}} \in \mathbb{R}^n$. Then $\widehat{\boldsymbol{\beta}}_M(\boldsymbol{Y})$ is a minimax estimator of $\boldsymbol{\beta} \in \{\boldsymbol{\beta} : \boldsymbol{\beta}'\boldsymbol{B}\boldsymbol{\beta} \leqslant 1\}$ in the linear regression model* (1.1) *with the loss function* (1.3) *if and only if $\widehat{\boldsymbol{\gamma}}_M(\tilde{\boldsymbol{Z}}) := \boldsymbol{V}'\widehat{\boldsymbol{\beta}}_M(\boldsymbol{U}\tilde{\boldsymbol{Z}})$ is a minimax estimator of $\tilde{\boldsymbol{\gamma}} \in \{\boldsymbol{\gamma} : \boldsymbol{\gamma}'\boldsymbol{V}'\boldsymbol{B}\boldsymbol{V}\boldsymbol{\gamma} \leqslant 1\}$ in the linear regression model* (5.4) *with the loss function $\tilde{L}(\widehat{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\gamma}}) := (\widehat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}})'\boldsymbol{V}'\boldsymbol{A}\boldsymbol{V}(\widehat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}})$. Moreover, $\widehat{\boldsymbol{\beta}}_M$ and $\widehat{\boldsymbol{\gamma}}_M$ have the same minimax risk.*

To simplify the estimation problem, we set $\tilde{\boldsymbol{\eta}} = (\tilde{\eta}_1, \ldots, \tilde{\eta}_p)' := \boldsymbol{L}\tilde{\boldsymbol{\gamma}}$. In this parametrization, $\tilde{Z}_i = \tilde{\eta}_i + \tilde{\epsilon}_i$ for $i = 1, \ldots, p$ and $\tilde{Z}_i = \tilde{\epsilon}_i$ for $i = p+1, \ldots, n$, i.e.

$$\tilde{\boldsymbol{Z}} = \boldsymbol{D}\tilde{\boldsymbol{\eta}} + \tilde{\boldsymbol{\epsilon}}, \quad \mathbb{E}_{\tilde{P}}\tilde{\boldsymbol{\epsilon}} = \boldsymbol{0},\ \operatorname{Cov}_{\tilde{P}}(\tilde{\boldsymbol{\epsilon}}) = \boldsymbol{I}_n. \tag{5.5}$$

Using arguments similar to those employed in the proof of Lemma 5.1, we conclude that $\widehat{\boldsymbol{\gamma}}_M(\tilde{\boldsymbol{Z}})$ is a minimax decision rule in the previous problem if and only if $\widehat{\boldsymbol{\eta}}_M(\tilde{\boldsymbol{Z}}) := \boldsymbol{L}\widehat{\boldsymbol{\gamma}}_M(\tilde{\boldsymbol{Z}})$ is a minimax estimator of $\tilde{\boldsymbol{\eta}} \in \{\boldsymbol{\eta} : \boldsymbol{\eta}'\boldsymbol{B}_1\boldsymbol{\eta} \leqslant 1\}$ with the loss $(\widehat{\boldsymbol{\eta}} - \tilde{\boldsymbol{\eta}})'\boldsymbol{A}_1(\widehat{\boldsymbol{\eta}} - \tilde{\boldsymbol{\eta}})$, where $\boldsymbol{A}_1 := \boldsymbol{L}^{-1}\boldsymbol{V}'\boldsymbol{A}\boldsymbol{V}\boldsymbol{L}^{-1}$ and $\boldsymbol{B}_1 := \boldsymbol{L}^{-1}\boldsymbol{V}'\boldsymbol{B}\boldsymbol{V}\boldsymbol{L}^{-1}$.

To simplify the estimation problem even further, note that in view of the spectral decomposition of a symmetric matrix (see Rao (1973)), there exist a unitary matrix $\tilde{\boldsymbol{P}} \in \mathbb{R}^{p\times p}$ and a diagonal matrix $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$ such that

$$\boldsymbol{B}_1^{-1/2}(\boldsymbol{B}_1^{1/2}\boldsymbol{A}_1\boldsymbol{B}_1^{1/2})^{1/2}\boldsymbol{B}_1^{-1/2} = \tilde{\boldsymbol{P}}\boldsymbol{\Lambda}\tilde{\boldsymbol{P}}' = \sum_{i=1}^{p}\lambda_i\tilde{\boldsymbol{P}}_i\tilde{\boldsymbol{P}}_i'. \tag{5.6}$$

This implies that the matrix $\boldsymbol{A}_1$ can be rewritten as

$$\boldsymbol{A}_1 = \tilde{\boldsymbol{P}}\boldsymbol{\Lambda}\tilde{\boldsymbol{P}}'\boldsymbol{B}_1\tilde{\boldsymbol{P}}\boldsymbol{\Lambda}\tilde{\boldsymbol{P}}'.$$

Let $\boldsymbol{P}$ be the unitary matrix of the form

$$\boldsymbol{P} = \begin{bmatrix} \tilde{\boldsymbol{P}} & \boldsymbol{0} \\ \boldsymbol{0}' & \boldsymbol{I}_{n-p} \end{bmatrix}$$

where $\mathbf{0} \in \mathbb{R}^{p\times(n-p)}$ is a rectangular submatrix of $\boldsymbol{P}$. Setting $\boldsymbol{Z} = \boldsymbol{P}'\tilde{\boldsymbol{Z}}, \boldsymbol{\gamma} = \tilde{\boldsymbol{P}}'\tilde{\boldsymbol{\eta}}$, $\boldsymbol{\epsilon} = \boldsymbol{P}'\tilde{\boldsymbol{\epsilon}}$, we transform the linear model (5.5) into the linear model (2.1), because $\boldsymbol{P}$ is a unitary matrix and $\boldsymbol{P}'\boldsymbol{z}$ keeps $(z_{p+1}, \ldots, z_n)$ unchanged for any $\boldsymbol{z} = (z_1, \ldots, z_n)' \in \mathbb{R}^p$.

Let $\tilde{\boldsymbol{A}} \in \mathbb{R}^{p\times p}$ and $\tilde{\boldsymbol{B}} \in \mathbb{R}^{p\times p}$ be defined as

$$\tilde{\boldsymbol{B}} = \tilde{\boldsymbol{P}}'\boldsymbol{B}_1\tilde{\boldsymbol{P}}, \quad \tilde{\boldsymbol{A}} = \boldsymbol{\Lambda}\tilde{\boldsymbol{P}}'\boldsymbol{B}_1\tilde{\boldsymbol{P}}\boldsymbol{\Lambda} = \boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}.$$

Again, by arguments similar to those used in the proof of Lemma 5.1, we conclude the following: The minimax estimation of $\tilde{\boldsymbol{\eta}} \in \{\boldsymbol{\eta} : \boldsymbol{\eta}'\boldsymbol{B}_1\boldsymbol{\eta} \leqslant 1\}$ in the model (5.5), under the loss function $(\widehat{\boldsymbol{\eta}} - \tilde{\boldsymbol{\eta}})'\boldsymbol{A}_1(\widehat{\boldsymbol{\eta}} - \tilde{\boldsymbol{\eta}})$, is equivalent to finding the minimax estimator $\widehat{\boldsymbol{\gamma}}(\boldsymbol{Z})$ of $\boldsymbol{\gamma} \in \Gamma = \{\boldsymbol{\gamma} : \boldsymbol{\gamma}'\tilde{\boldsymbol{B}}\boldsymbol{\gamma} \leqslant 1\}$ in the model (2.1), under the loss fuction $\tilde{L}(\widehat{\boldsymbol{\gamma}}, \boldsymbol{\gamma}) = (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\tilde{\boldsymbol{A}}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$, where $\tilde{\boldsymbol{A}} = \boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}$. This completes the proof of the proposition. ∎

Proof of Theorem 2.1. First, we apply Proposition 1.1 to identify the counterpart of condition (1.4) and to obtain the explicit formula for $\widehat{\boldsymbol{\gamma}}_{ML}$.

Since the matrices $\boldsymbol{X}$, $\boldsymbol{B}$ and $\boldsymbol{A}$ appearing in this proposition must be replaced by $\boldsymbol{D}$, $\tilde{\boldsymbol{B}}$ and $\tilde{\boldsymbol{A}}$, respectively, we obtain $\boldsymbol{E} = \tilde{\boldsymbol{A}}$, because $\boldsymbol{D}'\boldsymbol{D} = \boldsymbol{I}_p$. Define

$$\boldsymbol{W} := \boldsymbol{H}^{-1/2}(\boldsymbol{H}^{1/2}\tilde{\boldsymbol{A}}^2\boldsymbol{H}^{1/2})^{1/2}\boldsymbol{H}^{-1/2}.$$

Then $(\boldsymbol{H}^{1/2}\tilde{\boldsymbol{A}}^2\boldsymbol{H}^{1/2})^{1/2} = \boldsymbol{H}^{1/2}\boldsymbol{W}\boldsymbol{H}^{1/2}$ and $\tilde{\boldsymbol{A}}^2 = \boldsymbol{W}\boldsymbol{H}\boldsymbol{W}$. Since $\tilde{\boldsymbol{A}} = \boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}$ and $\boldsymbol{H} = \tilde{\boldsymbol{A}}^{-1/2}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{A}}^{-1/2}$, it follows that $\boldsymbol{W} = \tilde{\boldsymbol{A}}^{1/2}\boldsymbol{\Lambda}\tilde{\boldsymbol{A}}^{1/2}$. Thus the matrix $\boldsymbol{N}^*$ defined in Proposition 1.1 takes the form

$$\boldsymbol{N}^* := \kappa\cdot\boldsymbol{H}^{-1/2}(\boldsymbol{H}^{1/2}\tilde{\boldsymbol{A}}^2\boldsymbol{H}^{1/2})^{1/2}\boldsymbol{H}^{-1/2} - \tilde{\boldsymbol{A}} = \tilde{\boldsymbol{A}}^{1/2}(\kappa\boldsymbol{\Lambda} - \boldsymbol{I}_p)\tilde{\boldsymbol{A}}^{1/2}, \tag{5.7}$$

and condition (1.4) is equivalent to condition (C1) introduced in Theorem 2.1.

The number κ, defined in Proposition 1.1, satisfies (2.2), because

$$\begin{aligned}\operatorname{tr}((\boldsymbol{H}^{1/2}\tilde{\boldsymbol{A}}^2\boldsymbol{H}^{1/2})^{1/2}) &= \operatorname{tr}(\boldsymbol{H}^{1/2}\boldsymbol{W}\boldsymbol{H}^{1/2}) = \operatorname{tr}(\boldsymbol{W}\boldsymbol{H})\\ &= \operatorname{tr}(\tilde{\boldsymbol{A}}^{1/2}\boldsymbol{\Lambda}\tilde{\boldsymbol{A}}^{1/2}\tilde{\boldsymbol{A}}^{-1/2}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{A}}^{-1/2}) = \operatorname{tr}(\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}),\end{aligned}$$

and

$$\operatorname{tr}(\boldsymbol{H}\tilde{\boldsymbol{A}}) = \operatorname{tr}(\tilde{\boldsymbol{A}}^{-1/2}\tilde{\boldsymbol{B}}\tilde{\boldsymbol{A}}^{-1/2}\tilde{\boldsymbol{A}}) = \operatorname{tr}(\tilde{\boldsymbol{B}}).$$

Finally, when $\kappa\boldsymbol{\Lambda} - \boldsymbol{I}_p \succeq \mathbf{0}$, (1.5) and (5.7), together with the identity $\boldsymbol{E} = \tilde{\boldsymbol{A}}$, yield

$$\widehat{\boldsymbol{\gamma}}_{ML}(\boldsymbol{Z}) = (\boldsymbol{I}_p - \kappa^{-1}\boldsymbol{\Lambda}^{-1})\boldsymbol{D}'\boldsymbol{Z}.$$

To prove minimaxity of $\widehat{\boldsymbol{\gamma}}_{ML}$ in $\mathcal{D}$ we first show that this decision rule is Bayes with respect to some prior distribution π on $\mathcal{P}$. Let

$$\alpha_i := (\kappa\lambda_i - 1)b_{ii}, \quad i = 1, \ldots, p. \tag{5.8}$$

Since $\kappa\boldsymbol{\Lambda}-\boldsymbol{I}_p$ and $\tilde{\boldsymbol{B}}=[b_{ij}]$ are positive definite matrices, the numbers $\alpha_1,\dots,\alpha_p$ are positive. Moreover, from (2.2) we obtain

$$\sum_{i=1}^{p}\alpha_i=\sum_{i=1}^{p}(\kappa\lambda_i-1)b_{ii}=\kappa\,\mathrm{tr}(\boldsymbol{\Lambda}\tilde{\boldsymbol{B}})-\mathrm{tr}(\tilde{\boldsymbol{B}})=\mathrm{tr}(\tilde{\boldsymbol{B}})+1-\mathrm{tr}(\tilde{\boldsymbol{B}})=1.$$

The construction of π begins by defining an auxiliary random vector $\boldsymbol{\Psi}$. Let

$$z_i=\begin{cases}\dfrac{\sqrt{b_{ii}}}{b_{ii}+\alpha_i} & \text{for } i=1,\dots,p,\\ 1 & \text{for } i=p+1,\dots,n,\end{cases}$$

and let

$$c=\max_{1\leqslant j\leqslant p} z_j.$$

Furthermore, let the components $\Psi_1,\dots,\Psi_n$ of $\boldsymbol{\Psi}=(\Psi_1,\dots,\Psi_n)'$ be independent random variables defined as follows:

- For $j=1,\dots,p$, the variable Ψ_j takes the values $-c/z_j$, 0, and c/z_j with probabilities $(z_j/c)^2/2$, $1-(z_j/c)^2$ and $(z_j/c)^2/2$, respectively.
- For $j=p+1,\dots,n$, the variable Ψ_j takes the values -1 and 1 with equal probabilities of $1/2$.

Clearly, each of these variables has mean zero and variance 1. Consequently, since the variables are independent, the covariance matrix of $\boldsymbol{\Psi}$ is the identity matrix $\boldsymbol{I}_n$. Specifically,

$$\boldsymbol{I}_n=\mathbb{E}(\boldsymbol{\Psi}\boldsymbol{\Psi}')=\sum_{\boldsymbol{\psi}\in\Delta}p(\boldsymbol{\psi})\times\boldsymbol{\psi}\boldsymbol{\psi}'.$$

Here Δ is a subset of $\mathbb{R}^n$ with cardinality $N=3^p\cdot 2^{n-p}$, defined by

$$\Delta=\{\boldsymbol{\psi}=(\psi_1,\dots\psi_n):\psi_j\in\{-c/z_j,0,c/z_j\},\ 1\leqslant j\leqslant p,\\ \psi_j=\pm1,\ j=p+1,\dots,n\},$$

and $p(\boldsymbol{\psi})$, where $\boldsymbol{\psi}\in\Delta$, is the probability that the random vector $\boldsymbol{\Psi}$ takes the value $\boldsymbol{\psi}$. Clearly, due to the independence of $\Psi_1,\dots,\Psi_n$, we have

$$p(\boldsymbol{\psi})=p(\psi_1,\dots,\psi_n)=\prod_{j=1}^{p}((z_j/c)^2/2)^{|\mathrm{sgn}(\psi_j)|}(1-(z_j/c)^2)^{1-|\mathrm{sgn}(\psi_j)|}\\ \cdot\prod_{j=p+1}^{n}1/2,$$

where $\mathrm{sgn}(x)$ is 1, 0 or -1 according as x is positive, zero or negative. Next, define $\boldsymbol{u}(\boldsymbol{\psi})=\sqrt{p(\boldsymbol{\psi})}\boldsymbol{\psi}$ and denote by $\boldsymbol{U}=[\boldsymbol{U}_1:\cdots:\boldsymbol{U}_N]$ the $n\times N$ matrix with columns formed from successive values of $\boldsymbol{u}(\boldsymbol{\psi})$, $\boldsymbol{\psi}\in\Delta$. Then

$$\boldsymbol{U}\boldsymbol{U}'=\sum_{j=1}^{N}\boldsymbol{U}_j\boldsymbol{U}_j'=\sum_{\boldsymbol{\psi}\in\Delta}\boldsymbol{u}(\boldsymbol{\psi})\boldsymbol{u}'(\boldsymbol{\psi})=\boldsymbol{I}_n.$$

Additionally, let $\boldsymbol{e}_1 = (1, 0, \ldots, 0)', \ldots, \boldsymbol{e}_p = (0, 0, \ldots, 1)'$ be the standard orthonormal basis in $\mathbb{R}^p$ and let

$$p_i = \left(\frac{\alpha_i}{b_{ii} + \alpha_i}\right)^2, \quad \boldsymbol{\gamma}_i = (1/\sqrt{b_{ii}})\boldsymbol{e}_i, \quad i = 1, \ldots, p,$$

Clearly $\boldsymbol{\gamma}_i \in \Gamma$, $i = 1, \ldots, p$, since $\boldsymbol{\gamma}_i' \tilde{\boldsymbol{B}} \boldsymbol{\gamma}_i = 1$. Furthermore, let $\boldsymbol{Z}$ be an observed discrete random vector with the probability mass function given by

$$P_i(\boldsymbol{Z} = \boldsymbol{z}) = \begin{cases} p_i & \text{if } \boldsymbol{z} = \boldsymbol{D}\boldsymbol{C}^{-1}\boldsymbol{\gamma}_i, \\ p_{ij} & \text{if } \boldsymbol{z} = d_j \boldsymbol{U}_j, \\ q_{ij} & \text{if } \boldsymbol{z} = -d_j \boldsymbol{U}_j, \\ 1 - \left[p_i + \sum_{j=1}^{N} (p_{ij} + q_{ij})\right] & \text{if } \boldsymbol{z} = \boldsymbol{0}, \end{cases} \tag{5.9}$$

where we define

$$\boldsymbol{C} := (\boldsymbol{I}_p - \kappa^{-1}\boldsymbol{\Lambda}^{-1}) = \operatorname{diag}\left(\frac{\alpha_1}{b_{11} + \alpha_1}, \ldots, \frac{\alpha_p}{b_{pp} + \alpha_p}\right).$$

We choose the numbers $p_{ij} \geqslant 0$, $q_{ij} \geqslant 0$ and $d_j > 0$, $i = 1, \ldots, n$, $j = 1, \ldots, N$, so that

$$\mathbb{E}_{P_i} \boldsymbol{Z} = \boldsymbol{D}\boldsymbol{\gamma}_i \quad \text{and} \quad \mathbb{E}_{P_i} \boldsymbol{Z}\boldsymbol{Z}' = \boldsymbol{I}_n + \boldsymbol{D}\boldsymbol{\gamma}_i(\boldsymbol{D}\boldsymbol{\gamma}_i)'. \tag{5.10}$$

Since

$$\boldsymbol{C}^{-1}\boldsymbol{\gamma}_i = \boldsymbol{C}^{-1}\boldsymbol{e}_i/\sqrt{b_{ii}} = \frac{b_{ii} + \alpha_i}{\alpha_i}\boldsymbol{\gamma}_i \quad \text{and} \quad p_i = \left(\frac{\alpha_i}{b_{ii} + \alpha_i}\right)^2,$$

it follows that

$$p_i(\boldsymbol{D}\boldsymbol{C}^{-1}\boldsymbol{\gamma}_i) = \frac{\alpha_i}{b_{ii} + \alpha_i}\boldsymbol{D}\boldsymbol{\gamma}_i \quad \text{and} \quad p_i(\boldsymbol{D}\boldsymbol{C}^{-1}\boldsymbol{\gamma}_i)(\boldsymbol{D}\boldsymbol{C}^{-1}\boldsymbol{\gamma}_i)' = \boldsymbol{D}\boldsymbol{\gamma}_i(\boldsymbol{D}\boldsymbol{\gamma}_i)'.$$

Therefore, condition (5.10) is satisfied whenever

$$\sum_{j=1}^{N} (p_{ij} - q_{ij}) d_j \boldsymbol{U}_j = \frac{b_{ii}}{b_{ii} + \alpha_i}\boldsymbol{D}\boldsymbol{\gamma}_i, \quad \sum_{j=1}^{N} (p_{ij} + q_{ij}) d_j^2 \boldsymbol{U}_j \boldsymbol{U}_j' = \boldsymbol{I}_n.$$

An obvious candidate for solving the above system of equations is given by

$$(p_{ij} - q_{ij}) d_j = \frac{b_{ii}}{b_{ii} + \alpha_i}\boldsymbol{U}_j'\boldsymbol{D}\boldsymbol{\gamma}_i, \quad (p_{ij} + q_{ij}) d_j^2 = 1, \quad j = 1, \ldots, N, \tag{5.11}$$

because $\sum_{j=1}^{N} \boldsymbol{U}_j \boldsymbol{U}_j' = \boldsymbol{I}_n$. Furthermore, based on the structure of the vectors $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_N$ and the fact that the matrix $\boldsymbol{D} = [d_{ij}] \in \mathbb{R}^{n \times p}$ has all entries zero

except for $d_{11} = \cdots = d_{pp} = 1$, it follows that for each $j = 1, \ldots, N$ and each $i = 1, \ldots, p$, we have

$$\boldsymbol{U}_j' \boldsymbol{D} \boldsymbol{\gamma}_i = \boldsymbol{U}_j' \boldsymbol{D} \left(\frac{1}{\sqrt{b_{ii}}} \boldsymbol{e}_i \right) = \frac{1}{\sqrt{b_{ii}}} \boldsymbol{U}_j' \boldsymbol{w}_i = c \sqrt{p(\boldsymbol{\psi}_j)} \, \frac{b_{ii} + \alpha_i}{b_{ii}} \operatorname{sgn}(\boldsymbol{U}_j' \boldsymbol{D} \boldsymbol{e}_i).$$

Here, $\boldsymbol{w}_i \in \mathbb{R}^n$ is a vector with all coordinates zero except the ith which is 1, and $\boldsymbol{\psi}_j$ is the element of Δ that corresponds to the vector $\boldsymbol{U}_j$. Therefore, it is evident that a solution of (5.11) is given by

$$d_j = \frac{1}{c \sqrt{p(\boldsymbol{\psi}_j)}}, \quad p_{ij} + q_{ij} = \frac{1}{d_j^2}, \quad p_{ij} - q_{ij} = \frac{\operatorname{sgn}(\boldsymbol{U}_j' \boldsymbol{D} \boldsymbol{e}_i)}{d_j^2}. \tag{5.12}$$

Note that this solution defines a probability distribution, because

$$p_{ij} = \frac{1 + \operatorname{sgn}(\boldsymbol{U}_j' \boldsymbol{D} \boldsymbol{e}_i)}{2 d_j^2} \geqslant 0, \quad q_{ij} = \frac{1 - \operatorname{sgn}(\boldsymbol{U}_j' \boldsymbol{D} \boldsymbol{e}_i)}{2 d_j^2} \geqslant 0$$

and, by (5.8), we have

$$\begin{aligned} p_i + \sum_{j=1}^{N} (p_{ij} + q_{ij}) &= p_i + \sum_{j=1}^{N} c^2 p(\boldsymbol{\psi}_j) = p_i + c^2 \\ &= \frac{\alpha_i^2}{(b_{ii} + \alpha_i)^2} + \max_{1 \leqslant i \leqslant p} \frac{b_{ii}}{(b_{ii} + \alpha_i)^2} = \frac{(\kappa \lambda_i - 1)^2}{(\kappa \lambda_i)^2} + \max_{1 \leqslant i \leqslant p} \frac{1}{(\kappa \lambda_i)^2 b_{ii}} \leqslant 1, \end{aligned}$$

where the last inequality follows from condition (C2).

Let the prior π choose a distribution P of $\boldsymbol{Z}$, and hence the corresponding distribution from $\mathcal{P}$ of the error $\boldsymbol{\epsilon} = \boldsymbol{Z} - E_P \boldsymbol{Z}$, in such a way that

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

where $\overline{P}_i$ is the distribution of $-\boldsymbol{Z}$ when $\boldsymbol{Z}$ is distributed according to P_i, $i = 1, \ldots, n$. Then $\overline{P}_i(\boldsymbol{Z} = \boldsymbol{z}) = P_i(\boldsymbol{Z} = -\boldsymbol{z})$, $\boldsymbol{z} \in \mathbb{R}^p$, $i = 1, \ldots, p$. Therefore, by (5.12), the π-Bayes estimator $\widehat{\boldsymbol{\gamma}}_\pi(\boldsymbol{z}) = E(\boldsymbol{\gamma} \mid \boldsymbol{Z} = \boldsymbol{z})$ of $\boldsymbol{\gamma}$ has the form

$$\begin{aligned} \widehat{\boldsymbol{\gamma}}_\pi(\boldsymbol{z}) &= \frac{\sum_{i=1}^n \frac{\alpha_i}{2} \left[P_i(\boldsymbol{Z} = \boldsymbol{z}) - P_i(\boldsymbol{Z} = -\boldsymbol{z}) \right] \boldsymbol{\gamma}_i}{\sum_{i=1}^n \frac{\alpha_i}{2} \left[P_i(\boldsymbol{Z} = \boldsymbol{z}) + P_i(\boldsymbol{Z} = -\boldsymbol{z}) \right]} \\ &= \begin{cases} \boldsymbol{0} & \text{if } \boldsymbol{z} = \boldsymbol{0}, \\ \boldsymbol{\gamma}_i & \text{if } \boldsymbol{z} = \boldsymbol{D} \boldsymbol{C}^{-1} \boldsymbol{\gamma}_i, \\ \sum_{j=1}^{p} \alpha_j \operatorname{sgn}(\boldsymbol{U}_j' \boldsymbol{D} \boldsymbol{\gamma}_i) \boldsymbol{\gamma}_j & \text{if } \boldsymbol{z} = d_j \boldsymbol{U}_j, \\ -\sum_{j=1}^{p} \alpha_j \operatorname{sgn}(\boldsymbol{U}_j' \boldsymbol{D} \boldsymbol{\gamma}_i) \boldsymbol{\gamma}_j & \text{if } \boldsymbol{z} = -d_j \boldsymbol{U}_j. \end{cases} \end{aligned}$$

To prove that $\widehat{\gamma}_\pi = \widehat{\gamma}_{ML}$, we first observe that for any $\boldsymbol{z} \in \mathbb{R}^n$,

$$\widehat{\gamma}_{ML}(\boldsymbol{z}) = (\boldsymbol{I}_p - \kappa^{-1}\boldsymbol{\Lambda}^{-1})\boldsymbol{D}'\boldsymbol{z} = \boldsymbol{C}\boldsymbol{D}'\boldsymbol{z} = \left(\sum_{i=1}^{p} \frac{\alpha_i}{b_{ii}+\alpha_i}\boldsymbol{e}_i\boldsymbol{e}_i'\right)\boldsymbol{D}'\boldsymbol{z}.$$

This implies that for each $1 \leqslant j \leqslant N$,

$$\begin{aligned}
\widehat{\gamma}_{ML}(d_j\boldsymbol{U}_j) &= d_j\boldsymbol{C}\boldsymbol{D}'\boldsymbol{U}_j d_j \sum_{i=1}^{p} \frac{\alpha_i}{b_{ii}+\alpha_i}\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{D}'\boldsymbol{U}_j \\
&= d_j \sum_{i=1}^{p} \frac{\alpha_i}{b_{ii}+\alpha_i}\boldsymbol{e}_i(\boldsymbol{U}_j'\boldsymbol{D}\boldsymbol{e}_i) \\
&= d_j \sum_{i=1}^{p} \frac{\alpha_i}{b_{ii}+\alpha_i}\boldsymbol{e}_i\left(c\sqrt{p(\boldsymbol{\psi}_j)}\,\mathrm{sgn}(\boldsymbol{U}_j'\boldsymbol{D}\boldsymbol{e}_i)\frac{b_{ii}+\alpha_i}{\sqrt{b_{ii}}}\right) \\
&= \sum_{j=1}^{p} \alpha_j \mathrm{sgn}(\boldsymbol{U}_j'\boldsymbol{D}\boldsymbol{\gamma}_i)\boldsymbol{\gamma}_j = \widehat{\gamma}_\pi(d_j\boldsymbol{U}_j)
\end{aligned}$$

and clearly $\widehat{\gamma}_{ML}(-d_j\boldsymbol{U}_j) = \widehat{\gamma}_\pi(-d_j\boldsymbol{U}_j)$. Moreover, $\widehat{\gamma}_{ML}(\boldsymbol{0}) = \widehat{\gamma}_\pi(\boldsymbol{0})$ and since $\boldsymbol{D}'\boldsymbol{D} = \boldsymbol{I}_p$, we also find that for each $1 \leqslant i \leqslant p$,

$$\widehat{\gamma}_{ML}(\boldsymbol{D}\boldsymbol{C}^{-1}\boldsymbol{\gamma}_i) = \boldsymbol{C}\boldsymbol{D}'\boldsymbol{D}\boldsymbol{C}^{-1}\boldsymbol{\gamma}_i = \boldsymbol{\gamma}_i = \widehat{\gamma}_\pi(\boldsymbol{D}\boldsymbol{C}^{-1}\boldsymbol{\gamma}_i).$$

From these equalities we conclude that the π-Bayes rule $\widehat{\gamma}_\pi$ is equivalent to $\widehat{\gamma}_{ML}$.

To complete the proof, we first compute the risk function of $\widehat{\gamma}_{ML}$. Since $\widehat{\gamma}_{ML}(\boldsymbol{Z}) = \boldsymbol{C}\boldsymbol{D}'\boldsymbol{Z}$ and $\boldsymbol{D}'\boldsymbol{D} = \boldsymbol{I}_p$, it follows that the risk function of $\widehat{\gamma}_{ML}$ is

$$\begin{aligned}
&R(\widehat{\gamma}_{ML}, \boldsymbol{\gamma}, P) = \mathbb{E}_P\big([\boldsymbol{C}\boldsymbol{D}'\boldsymbol{Z} - \boldsymbol{\gamma}]'\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}[\boldsymbol{C}\boldsymbol{D}'\boldsymbol{Z} - \boldsymbol{\gamma}]\big) \\
&= \mathbb{E}_P\big([\boldsymbol{C}\boldsymbol{D}'(\boldsymbol{Z}-\boldsymbol{D}\boldsymbol{\gamma})-(\boldsymbol{I}-\boldsymbol{C}\boldsymbol{D}'\boldsymbol{D})\boldsymbol{\gamma}]'\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}[\boldsymbol{C}\boldsymbol{D}'(\boldsymbol{Z}-\boldsymbol{D}\boldsymbol{\gamma})-(\boldsymbol{I}-\boldsymbol{C}\boldsymbol{D}'\boldsymbol{D})\boldsymbol{\gamma}]\big) \\
&= \mathbb{E}_P\big([\boldsymbol{Z} - \boldsymbol{D}\boldsymbol{\gamma}]'\boldsymbol{D}\boldsymbol{C}'\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}\boldsymbol{C}\boldsymbol{D}'[\boldsymbol{Z} - \boldsymbol{D}\boldsymbol{\gamma}]\big) + \boldsymbol{\gamma}'(\boldsymbol{I} - \boldsymbol{C})'\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}(\boldsymbol{I} - \boldsymbol{C})\boldsymbol{\gamma} \\
&= \mathrm{tr}(\boldsymbol{D}\boldsymbol{C}'\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}\boldsymbol{C}\boldsymbol{D}') + \boldsymbol{\gamma}'(\boldsymbol{I} - \boldsymbol{C})'\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}(\boldsymbol{I} - \boldsymbol{C})\boldsymbol{\gamma} \\
&= \mathrm{tr}(\boldsymbol{D}\boldsymbol{C}'\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}\boldsymbol{C}\boldsymbol{D}') + \kappa^{-2}\boldsymbol{\gamma}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1}\boldsymbol{\gamma} \\
&= \mathrm{tr}(\boldsymbol{D}\boldsymbol{C}'\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}\boldsymbol{C}\boldsymbol{D}') + \kappa^{-2}\boldsymbol{\gamma}'\tilde{\boldsymbol{B}}\boldsymbol{\gamma}.
\end{aligned}$$

When the parameter space is restricted to $\boldsymbol{\gamma} \in \Gamma = \{\boldsymbol{\gamma} : \boldsymbol{\gamma}'\tilde{\boldsymbol{B}}\boldsymbol{\gamma} \leqslant 1\}$, the maximal risk of $\widehat{\gamma}_{ML}$ is given by

$$\sup_{P\in\mathcal{P}} R(\widehat{\gamma}_{ML}, \boldsymbol{\gamma}, P) = \mathrm{tr}(\boldsymbol{D}\boldsymbol{C}'\boldsymbol{\Lambda}\tilde{\boldsymbol{B}}\boldsymbol{\Lambda}\boldsymbol{C}\boldsymbol{D}') + \kappa^{-2}.$$

This maximal value of the risk is attained for each distribution P_i in the support of the prior π, because $\boldsymbol{\gamma}_i'\tilde{\boldsymbol{B}}\boldsymbol{\gamma}_i = 1$ when $\boldsymbol{\gamma}_i = (1/\sqrt{b_{ii}})\boldsymbol{e}_i$, $i = 1, \ldots, p$. The minimaxity of $\widehat{\gamma}_{ML}$ among all estimators then follows from Lehmann and Casella (1998), Corollary 1.6, p. 311. ∎

Acknowledgements. We are grateful to the referees for their valuable comments and suggestions, which greatly improved this paper.

REFERENCES

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

H. Drygas (1993), *Reparametrization methods in linear minimax estimation,* in: Statistical Sciences and Data Analysis, Proc. Third Pacific Area Statistical Conference (Makuhari, 1991), VSP, Utrecht, 87–95.

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

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

J. Kuks and W. Olman (1971), *Minimax linear estimation of regression coefficients I*, Izv. Akad. Nauk Estonsk. SSR 20, 480–482 (in Russian).

J. Kuks and W. Olman (1972), *Minimax linear estimation of regression coefficients II*, Izv. Akad. Nauk Estonsk. SSR 21, 66–72 (in Russian).

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

E. L. Lehmann and G. Casella (1998), *Theory of Point Estimation*, Springer.

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

J. Pilz (1991), *Bayesian estimation and Experimental Design in Linear Regression Models*, Wiley, New York.

C. R. Rao (1973), *Linear Statistical Inference And Its Applications*, Wiley, New York.

C. R. Rao and H. Toutenburg (1995), *Linear Models, Least Squares and Alternatives*, Springer, Heidelberg.

P. Stahlecker (1987), *A priori Information und Minimax-Schätzung im linearen Regressionsmodell*, Math. Systems Econom. 108, Athenäum, Frankfurt.

P. Stahlecker and G. Trenkler (1993), *Minimax estimation in linear regression with singular covariance structure and convex polyhedral constraints*, J. Statist. Plann. Inference 13, 297–318.

G. Trenkler and P. Stahlecker (1987), *Quasi minimax estimation in the linear regression model*, Statistics 36, 185–196.

M. Wilczyński (2007), *Minimax estimation in linear regression with ellipsoidal constraints*, J. Statist. Plann. Inference 137, 79–86.

Maciej Kowal, Maciej Wilczyński
Faculty of Pure and Applied Mathematics
Wroclaw University of Science and Technology
Wrocław, Poland
E-mail: maciej.kowal@pwr.edu.pl
maciej.wilczynski@pwr.edu.pl

Received 25.7.2025;
accepted 9.1.2026