Schwerpunktprogramm der Deutschen Forschungsgemeinschaft

ANWENDUNGSBEZOGENE OPTIMIERUNG UND STEUERUNG

On optimal estimators for linear parameters in mixed models — another approach

by

Johannes Gladitz

Report No. 109 Dezember 1988

Institut für Mathematik

UNIVERSITÄT AUGSBURG

Memminger Straße 6

D-8900 Augsburg

On optimal estimators for linear parameters in mixed models — another approach

By Johannes Gladitz*

Institute of Sociology and Social Policy, Academy of Sciences of the GDR, Berlin

Summary. A weak generalization of the mixed linear model and the use of Moore-Penrose inverses lead to a simple and more compact structure of the best inhomogeneous linear unbiased estimator (BILUE) of a linear estimable parameter vector. That is similar to the well-known structure of the linear Bayes estimator within a Bayesian model. A restricted minimax optimality is proved. The resulting simple presentation of the optimal risk opens the door to an approach to optimal experimental design in mixed models.

AMS 1980 subject classification: 62F10, 62J05.

Key words and phrases: Mixed linear model, Moore-Penrose inverses, best inhomogeneous linear unbiased estimator, quadratic risk, restricted minimax property.

1. INTRODUCTION

Let us consider an ordinary linear model. Further, we assume that additional information on a part of the unknown regression parameters is available, which restrict the variability region of these components. If we express this knowledge in our model by a probability distribution on the parameter space (prior distribution), then we obtain a so-called mixed model. It contains the ordinary as well as the Bayesian linear model as limiting cases. However, also the often considered models with (deterministic) restrictions on the parameter space (e.g. Gaffke and Heiligers (1988)) are included using a corresponding prior distribution. Therefore, the mixed models including all limiting cases are one of the most applicated models for analysis and prediction of cause-response relations in the natural, technical, and social sciences.

Originally, mixed models were considered only in connection of variance analytical investigations, especially for the estimation of variance components. Already 25 years ago,

* The paper was written during the author's stay at the University of Augsburg (FRG), Institute of Mathematics, under a contract with the Deutsche Forschungsgemeinschaft.

Henderson (1963) directed his interest also to the estimation of the (not directly observable) realizations of the random effects and constructed corresponding estimators.

These investigations were continued later, e.g. by Harville (1978, 1979) and Bunke (1977). All these authors used in their model representation a splitted design matrix concerning the fixed and the random part in the parameter vector. This block representation of the design led to relatively complicated estimators and to a risk function, which was closed for any approach to optimal experimental design.

We want to try to attack this problem by a simple trick.

2. THE MIXED MODEL

Let A be an arbitrary $(r \times s)$-matrix. Then we denote by $\mathcal{R}(A)$ the range of A. The starting point of our consideration is the following assumption:

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

where

X is a known $(n \times k)$ design matrix, not necessary of full rank,

ε is a n-dimensional random vector with

$$E(\varepsilon|\beta) = 0, \qquad D(\varepsilon|\beta) = \Omega > 0,$$

independent of a special $\beta \in \mathbb{R}^k$ with known positive definite (p.d.) $(n \times n)$ matrix Ω (E and D denote the expectation and the covariance matrix respectively), and

β is a k-dimensional vector which is orthogonally splitted by a known projector

$$P_1 \qquad (P_2 := I_k - P_1)$$

(I_k identity matrix of order k) into

$$\beta = \beta_1 + \beta_2 \qquad \text{with} \qquad \beta_1 = P_1\beta, \qquad \beta_2 = P_2\beta,$$

where β_2 is fixed but unknown and β_1 is random with a distribution independent of β_2 with

$$E\beta_1 = b, \qquad D\beta_1 = T \geq 0 \text{ positive semidefinite (p.s.d.)}$$

where b and T are known parameters (prior parameters).

Without loss of generality we can assume that

$$b \in \mathcal{R}(P_1) \quad \text{and} \quad \mathcal{R}(T) = \mathcal{R}(P_1).$$

With the special choice of

$$P_1 = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & \ddots & & & & \vdots \\ \vdots & & 1 & & & \vdots \\ \vdots & & & 0 & & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 \end{pmatrix},$$

model (1) includes the mixed models considered in the literature. With $P_1 = 0$ we obtain the usual linear and with $P_1 = I_k$ the Bayesian model, which is formally equivalent to the random coefficient regression model. In the following we use the denotation $\mathcal{L} := \mathcal{R}(P_2)$. The joint distribution of (β, ε) depends on the parameters β_2, b, T, Ω and a parameter κ, which determines uniquely that distribution. We denote it by $P_\pi^{\beta_2}$, where

$$\pi(\vartheta, \kappa), \quad \vartheta = (b, T, \Omega) \quad \text{and} \quad \beta_2 \in \mathcal{L}.$$

The parameter π varies within a set Π, which contains all parameters π fulfilling our assumption (1), the existence of the first and second moments of $P_\pi^{\beta_2}$ included. Then the parameter space Π has the form

$$\Pi = \Theta \times \mathcal{K}, \quad \vartheta \in \Theta, \quad \kappa \in \mathcal{K}.$$

Actually Π also depends on the projector P_1. It is assumed, however, that P_1 has to be choosen from structural or other considerations before the parameter identification , what should be possible in general. Within the usual mixed model frame the determination of P_1, for instance, might be the decision, on which components of β additional information is available or which components are assumed to be fixed.

Our goal is to estimate (or to predict) a realization of a linear parameter $\gamma = C\beta$ after the observation of Y, where C is a known $(r \times k)$ matrix of rank r.

3. LINEAR ESTIMABILITY AND OPTIMALITY

The power of an estimator δ is measured by a quadratic risk function

$$\begin{aligned} R^{\beta_2}(\pi, \delta) &= E_\pi^{\beta_2} \|\gamma - \delta(Y)\|_H^2 \\ &= \operatorname{tr} H M_\pi^{\beta_2}(\delta) \end{aligned} \tag{2}$$

or by the matrix risk corresponding to the Löwner semiordering

$$M_\pi^{\beta_2}(\delta) = E_\pi^{\beta_2} \left[\gamma - \delta(Y)\right] \left[\gamma - \delta(Y)\right]^\top, \tag{3}$$

where $H \geq 0$ denotes a given p.s.d. weighting matrix of order r and $E_\pi^{\beta_2}$ is the expectation corresponding to $P_\pi^{\beta_2}$. Let further

$$\mathcal{D} = \{\delta : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^r, \mathcal{B}^r)\}$$

be the set of all possible estimators ($\mathcal{B}^n$-σ-algebra of Borel sets), and

$$\mathcal{D}_{i\ell} = \left\{\delta \in \mathcal{D} : \delta(y) = a + Ly,\ a \in \mathbb{R}^r,\ y \in \mathbb{R}^n,\ L \in \mathcal{M}_{r\times n}\right\}$$

be the set of inhomogeneous linear estimators ($\mathcal{M}_{r\times k}$ denotes the set of all real valued $(r \times k)$ matrices).

Definition 1. For a given $\pi \in \Pi$ the estimator $\delta \in \mathcal{D}$ is called to be *unbiased*, if

$$\delta \in \mathcal{D}_u^\pi = \left\{\delta \in \mathcal{D} : E_\pi^{\beta_2}\left(\gamma - \delta(Y)\right) = 0 \quad \text{for all } \beta_2 \in \mathcal{L}\right\}. \qquad \square$$

Definition 2. The parameter $\gamma = C\beta$ is called to be *linear estimable* in $\pi \in \Pi$, if

$$\mathcal{D}_{i\ell u}^\pi := \mathcal{D}_{i\ell} \cap \mathcal{D}_u^\pi \neq \emptyset. \qquad \square$$

Remark 1. γ is linear estimable, iff an $L \in \mathcal{M}_{r\times n}$ exists such that

$$P_2(C - LX)^\top = 0. \tag{4}$$

Obviously

$$E_\pi^{\beta_2}(C\beta - a - LY) = 0 \qquad \text{for all } \beta_2 \in \mathcal{L}$$

is equivalent to

$$(C - LX)b + (C - LX)\beta_2 = a \qquad \text{for all } \beta_2 \in \mathcal{L},$$

and this holds for fixed b only, if

$$(C - LX)\beta_2 = 0 \qquad \text{for all } \beta_2 \in \mathcal{L},$$

what is the same as (4). $\square$

With Remark 1, the concept of linear estimability depends on π only by the parameter b. From (4) we obtain for the ordinary linear model ($P_1 = 0$) the well-known estimability condition $C = LX$.

With $P_1 = I_k$ (Bayesian model), Condition (4) is not restrictive. Any linear parameter γ is estimable. The missing information from observations is in the estimation replaced by the corresponding prior information.

From Remark 1 we further obtain that for any inhomogeneous linear unbiased estimator $\delta = a + LY$, it necessarily holds

$$a = (C - LX)b \tag{5}$$

and

$$(C - LX)\beta_2 = 0 \qquad \text{for all } \beta_2 \in \mathcal{L}.$$

Therefore, the set $\mathcal{D}^{\pi}_{i\ell u}$ can be written as

$$\mathcal{D}_{i\ell u} = \{\delta \in \mathcal{D} : \delta = (C - LX)b + Ly,\ L \in \mathcal{M}_{r\times n} \text{ and } P_2(C - LX)^{\top} = 0\}, \tag{6}$$

which only depends on the known b, the reason, why it can be omitted.

Definition 3. The estimator $\delta_1 \in \mathcal{D}$ is called to be *better* than $\delta_2 \in \mathcal{D}$ for a given $\pi \in \Pi$

(i) in the sense of (2), if
$\mathcal{R}^{\beta_2}(\pi, \delta_1) \leq \mathcal{R}^{\beta_2}(\pi, \delta_2)$ for all $\beta_2 \in \mathcal{L}$, and
$\mathcal{R}^{\beta_2}(\pi, \delta_1) \neq \mathcal{R}^{\beta_2}(\pi, \delta_2)$ for at least one $\beta_2 \in \mathcal{L}$,

(ii) in the sense of (3), if
$M^{\beta_2}_{\pi}(\delta_2) - M^{\beta_2}_{\pi}(\delta_1) \geq 0$ (p.s.d.) for all $\beta_2 \in \mathcal{L}$, and
$M^{\beta_2}_{\pi}(\delta_2) \neq M^{\beta_2}_{\pi}(\delta_1)$ for at least one $\beta_2 \in \mathcal{L}$. □

Remark 2. The following known relation consists between the both risks: If there exists an optimal estimator corresponding to (2), which is independent of the special choice of $H \geq 0$, then it is optimal too in the sense of (3) and vice versa. □

In the following we use the risk (2). For a given $\pi \in \Pi$, in general, there does not exist an optimal estimator within the "big" class $\mathcal{D}$. Therefore, one considers the restricted classes $\mathcal{D}^{\pi}_{u}$ or even $\mathcal{D}_{i\ell u}$.

Definition 4. $\delta^{\pi} \in \mathcal{D}^{\pi}_{u}$ is called to be a *best unbiased estimator (BUE)* for γ, if

$$\inf_{\delta\in\mathcal{D}^{\pi}_{u}} \mathcal{R}^{\beta_2}(\pi, \delta) = \mathcal{R}^{\beta_2}(\pi, \delta^{\pi}) \quad \text{for all } \beta_2 \in \mathcal{L}.$$ □

For $\delta \in \mathcal{D}_{ilu}$, the risks (2) and (3) are independent of $\kappa \in \mathcal{K}$ and $\beta_2 \in \mathcal{L}$. In that case we use the denotation

$$\mathcal{R}(\vartheta, \delta) := \mathcal{R}^{\beta_2}(\pi, \delta).$$

Definition 5. $\delta_\vartheta \in \mathcal{D}_{ilu}$ is called to be a *best inhomogeneous linear unbiased estimator (BILUE)* for γ, if

$$\inf_{\delta \in \mathcal{D}_{ilu}} \mathcal{R}(\vartheta, \delta) = \mathcal{R}(\vartheta, \delta_\vartheta). \qquad \square$$

4. THE OPTIMAL ESTIMATOR

Let now γ be a linear estimable parameter, i.e.

$$\mathcal{R}(P_2 C^\top) \subseteq \mathcal{R}(P_2 X^\top), \tag{7}$$

then the following matrix properties are true:

Lemma 1. *Under the assumptions (1) and (7) it holds*

$$\mathcal{R}(C^\top) \subseteq \mathcal{R}(T^+ + Q), \tag{8}$$

where $Q := X^\top \Omega^{-1} X$ *and* T^+ *denotes the Moore-Penrose inverse of* T.

Proof. We split up the range $\mathcal{R}(C^\top)$ orthogonally into the direct sum

$$\mathcal{R}(C^\top) = \mathcal{R}(P_1 C^\top) \oplus \mathcal{R}(P_2 C^\top).$$

Because of $\mathcal{R}(T^+) = \mathcal{R}(T)$, $\mathcal{R}(Q) = \mathcal{R}(X^\top)$, $T^+ \geq 0$, and $Q \geq 0$, we obtain

$$\begin{aligned} \mathcal{R}(T^+ + Q) = \mathcal{R}(T \vdots X^\top) &= \mathcal{R}(T \vdots P_2 X^\top) \\ &= \mathcal{R}(P_1) \oplus \mathcal{R}(P_2 X^\top), \end{aligned}$$

using $\mathcal{R}(P_1) = \mathcal{R}(T)$ from (1). With $\mathcal{R}(P_1 C^\top) \subseteq \mathcal{R}(P_1)$ and (7) it follows (8). $\square$

Lemma 2. *Under the assumptions of Lemma 1, the matrix identity holds:*

$$C(T^+ + Q)^+ T^+ = C - C(T^+ + Q)^+ Q. \tag{9}$$

Proof. The identity (9) is equivalent to

$$C(T^+ + Q)^+ (T^+ + Q) = C,$$

which is, however, a consequence of Lemma 1. $\square$

With the following theorem we obtain a BILUE for γ, which is independent of $H \geq 0$ and $\kappa \in \mathcal{K}$. There we use certain cylinder sets

$$\Pi_\vartheta = \{\pi \in \Pi : \pi = (\vartheta, \kappa),\ \kappa \in \mathcal{K}\}, \quad \vartheta \in \Theta, \tag{10}$$

which devide the parameter space Π into equivalence classes.

Theorem 1. *Let $\vartheta = (b, T, \Omega) \in \Theta$ be given. Under the assumptions (1) and (7) the estimator*

$$\delta_\vartheta = a^* + L^* Y \tag{11}$$

with

$$a^* = (C - L^* X)b,$$
$$L^* = C(T^+ + Q)^+ X^\top \Omega^{-1}$$

is BILUE for γ for any $\pi \in \Pi_\vartheta$ with a risk

$$\mathcal{R}(\vartheta, \delta_\vartheta) = \operatorname{tr} HC(T^+ + Q)^+ C^\top. \tag{12}$$

Proof. First of all we see that $\delta_\vartheta \in \mathcal{D}_u^\pi$, $\pi \in \Pi_\vartheta$. For all $\pi \in \Pi_\vartheta$ we have

$$\begin{aligned} E_\pi^{\beta_2}(\gamma - \delta_\vartheta) &= C\beta_2 - C(T^+ + Q)^+ Q\beta_2 \\ &= C(T^+ + Q)^+ T^+ \beta_2 \\ &= 0 \qquad \text{for all } \beta_2 \in \mathcal{L}, \end{aligned} \tag{13}$$

using Lemma 2 and $\mathcal{L} \perp \mathcal{R}(T^+)$ from (1). If we show that for any $\delta = a + LY \in \mathcal{D}_{i\ell u}$ the mixed terms in the risk are vanishing, i.e.

$$\operatorname{tr} HE_\pi^{\beta_2}(\gamma - \delta_\vartheta)(\delta_\vartheta - \delta)^\top = \operatorname{tr} HE_\pi^{\beta_2}(\delta_\vartheta - \delta)(\gamma - \delta_\vartheta)^\top = 0$$

for all $\beta_2 \in \mathcal{L}$ and $\pi \in \Pi_\vartheta$, then the first part of the theorem is proved. Then we have namely

$$\begin{aligned} R(\vartheta, \delta) &= R(\vartheta, \delta_\vartheta) + E_\pi^{\beta_2} \|\delta_\vartheta - \delta\|_H^2 \\ &\geq R(\vartheta, \delta_\vartheta). \end{aligned}$$

Using (13) and $\widetilde{\beta}_1 := \beta_1 - b$ we obtain

$$\gamma - \delta_\vartheta = (C - L^* X)\widetilde{\beta}_1 - L^* \varepsilon \tag{14}$$

and from $\delta \in \mathcal{D}_{i\ell u}$ (compare (6)) it follows

$$\delta_\vartheta - \delta = (L^* - L)(X\widetilde{\beta}_1 + \varepsilon).$$

Therefore, using (9) it holds independently of $\beta_2 \in \mathcal{L}$

$$\begin{aligned} E_\pi^{\beta_2}(\gamma - \delta_\vartheta)(\delta_\vartheta - \delta)^\top &= C(T^+ + Q)^+ T^+ T X^\top (L^* - L)^\top - C(T^+ + Q)^+ X^\top (L^* - L)^\top \\ &= C(T^+ + Q)^+ (T^+ T - I_k) X^\top (L^* - L)^\top \\ &= 0, \end{aligned}$$

because the matrix $(I_k - T^+ T)$ is a projector onto $\mathcal{L}$ and with $\delta_\vartheta, \delta \in \mathcal{D}_{i\ell u}$ it holds with (6)

$$P_2 X^\top (L^* - L)^\top = 0.$$

Thus, the first part of our theorem is proved.

Considering (9) and (14), we obtain for the risk

$$\begin{aligned} R(\vartheta, \delta_\vartheta) &= E_\pi^{\beta_2} \|(C - L^* X)\widetilde{\beta}_1 - L^* \varepsilon\|_H^2 \\ &= \operatorname{tr} H \left[(C - L^* X) T (C - L^* X)^\top + L^* \Omega L^{*T} \right] \\ &= \operatorname{tr} H C (T^+ + Q)^+ C^\top \end{aligned}$$

for all $\pi \in \Pi_\vartheta$ independently of $\beta_2 \in \mathcal{L}$. □

From Remark 2 it follows that δ_ϑ is also BILUE w.r.t. the matrix risk (3) with

$$M_\pi^{\beta_2}(\delta_\vartheta) = M_\vartheta(\delta_\vartheta) = C(T^+ + Q)^+ C^\top,$$

which has a more clear and simple structure than the usual presentations in the literature (e.g. Bunke (1977)). The structure of this optimal matrix risk is closely connected to the so-called Bayesian information matrix (e.g. Pilz (1983)). It might be the starting point for the development of optimal experimental designs in mixed models.

Remark 3. Using Lemma 2 we obtain for δ_ϑ from (11) the form of representation

$$\delta_\vartheta = C(T^+ + Q)^+ (T^+ b + Q\widehat{\beta}), \tag{15}$$

where

$$\widehat{\beta} = Q^+ X^\top \Omega^{-1} Y,$$

which is just the Aitken estimator for β in the case of regular X. Quite similar to the Bayesian linear model (cf. Hartigan (1969), Lindley and Smith (1972), Bunke and Gladitz (1974)), the structure of the optimal estimator δ_ϑ for the mixed model has proved to be something like a quasi-convex combination between the prior mean Cb and the Aitken estimator $C\widehat{\beta}$. For regular T, i.e. $P_1 = I_k$, the Moore-Penrose inverses turn over to regular inverses and δ_ϑ becomes the well-known linear Bayes estimator. For the case that additional information is not available (i.e. $T^+ = 0$), we obtain from (15) the usual BLUE $\delta_\vartheta = C\widehat{\beta}$. □

Remark 4. If the κ determines for the joint distribution of (β, ε) just a normal distribution with corresponding parameters ϑ and β_2, then δ_ϑ is even BUE. This property follows, as usual, from the fact that using a quadratic loss, the BUE is just the conditional expectation $E_\pi^{\beta_2}(\gamma|Y)$, which is for the normal distribution a linear function in Y. □

Remark 5. Analogous to Rao (1965) we now consider the case that also b is assumed to be unknown. Then the corresponding unbiasedness condition for linear estimators takes the form

$$E_\pi^{\beta_2}(\gamma - a - LY) = 0 \qquad \text{for all } \beta_2 \in \mathcal{L},\ b \in \mathcal{L}^\perp,$$

what means that $a = 0$ and $C = LX$. Therefore we obtain

$$\mathcal{R}(C^\top) \subseteq \mathcal{R}(X^\top) \tag{16}$$

as a necessary and sufficient condition for linear estimability for the parameter γ. We now show that under these conditions the estimator $\widehat{\gamma} = C\widehat{\beta}$ is BILUE for γ and

$$R^{\beta_2}(\pi, \widehat{\gamma}) = \operatorname{tr} HCQ^+C^\top \qquad \text{for all } \beta_2 \in \mathcal{L},\ b \in \mathcal{L}^\perp. \tag{17}$$

Because of (16) it holds

$$\gamma - \widehat{\gamma} = CQ^+X^\top\Omega^{-1}\varepsilon =: \widehat{L}\varepsilon, \tag{18}$$

and therefore, $E_\pi^{\beta_2}(\gamma - \widehat{\gamma}) = 0$ for all $\beta_2 \in \mathcal{L}$, $b \in \mathcal{L}^\perp$. Let L be an arbitrary $(r \times n)$ matrix with $C = LX$, then we have

$$\begin{aligned} E_\pi^{\beta_2}\|\gamma - LY\|_H^2 &= E_\pi^{\beta_2}\|\gamma - \widehat{L}Y\|_H^2 + E_\pi^{\beta_2}\|(\widehat{L} - L)Y\|_H^2 \\ &\geq E_\pi^{\beta_2}\|\gamma - \widehat{\gamma}\|_H^2, \end{aligned}$$

because the mixed terms are vanishing

$$\begin{aligned} &\operatorname{tr} H\widehat{L}(E_\pi^{\beta_2}\varepsilon\varepsilon^\top)(\widehat{L} - L)^\top + \operatorname{tr} H(\widehat{L} - L)(E_\pi^{\beta_2}\varepsilon\varepsilon^\top)\widehat{L}^\top \\ &= \operatorname{tr} HCQ^+(\widehat{L}X - LX)^\top + \operatorname{tr} H(\widehat{L}X - LX)Q^+C^\top \\ &= 0, \end{aligned}$$

taking into consideration the estimability condtion $\widehat{L}X = LX = C$.

The risk (17) follows from (18).

Therefore, if we only know the covariance structure of β_1 without knowledge of the location, this information does not result in an improvement of the generalized least squares estimator. □

5. THE RESTRICTED MINIMAX OPTIMALITY

Indeed, the restriction to the class of linear estimators or to the normal distribution are strong mathematical limitations. Although this assumptions in many practical situations are not so far from the reality (often the parameter κ is assumed to be in a "neighbourhood" of the normal distribution) one should be glad if some robustness properties of the used estimator could be proved concering to a possible violation of the normality assumption. For that we consider the

Definition 6. Let $\Pi^* \subset \Pi$, $\mathcal{D}^* \subseteq \mathcal{D}$ certain given subsets of Π and $\mathcal{D}$ respectively. An estimator $\delta^* \in \mathcal{D}^*$ is called to be *restricted strong minimax* w.r.t. Π^* within the class $\mathcal{D}^*$, if

$$\inf_{\delta \in \mathcal{D}^*} \sup_{\pi \in \Pi^*} R^{\beta_2}(\pi, \delta) = \sup_{\pi \in \Pi^*} R^{\beta_2}(\pi, \delta^*) \quad \text{for all } \beta_2 \in \mathcal{L} \tag{19}$$

(cf. Bunke (1964), Cogburn (1967), Wind (1973)). □

For our optimal estimator δ_ϑ we prove the following robustness property (analogous to Wind (1973)):

Theorem 2. *Let $\vartheta \in \Theta$ be given. Then the estimator δ_ϑ is restricted strong minimax w.r.t. Π_ϑ within $\mathcal{D}_u^{\pi_G}$, where $\pi_G = (\vartheta, \kappa_G)$ and κ_G denotes the normal distribution.*

Proof. Because of Theorem 1, the risk $R^{\beta_2}(\pi, \delta_\vartheta)$ is constant for all $\pi \in \Pi_\vartheta$ and $\beta_2 \in \mathcal{L}$. Further there exists an element $\pi_G \in \Pi_\vartheta$, for which δ_ϑ is BUE; i.e. δ_ϑ is the best estimator within the set $\mathcal{D}_u^{\pi_G}$. Assuming there would exist an element $\delta^* \in \mathcal{D}_u^{\pi_G}$ with $\delta^* \neq \delta_\vartheta$ and

$$\sup_{\pi \in \Pi_\vartheta} R^{\beta_2}(\pi, \delta^*) < R(\vartheta, \delta_\vartheta) \quad \text{for one } \beta_2 \in \mathcal{L},$$

then, because of $\pi_G \in \Pi_\vartheta$, it also holds

$$R^{\beta_2}(\pi_G, \delta^*) < R(\vartheta, \delta_\vartheta),$$

what is in contradiction to

$$R(\vartheta, \delta_\vartheta) = R^{\beta_2}(\pi_G, \delta_\vartheta) \leq R^{\beta_2}(\pi_G, \delta) \quad \text{for all } \delta \in \mathcal{D}_u^{\pi_G},\ \beta_2 \in \mathcal{L}.$$ □

With Theorem 2 we have shown that δ_ϑ minimizes—with respect to Π_ϑ—the least favourable risk within the set of unbiased estimators $\mathcal{D}_u^{\pi_G}$, in fact uniformly in $\beta_2 \in \mathcal{L}$.

If we use a weaker optimality concept than (19), a minimax property for δ_ϑ can be proved even in the class of all estimators $\mathcal{D}$.

Definition 7. Let $\Pi^* \subset \Pi$, $\mathcal{D}^* \subseteq \mathcal{D}$. An estimator $\delta^* \in \mathcal{D}^*$ is called to be *restricted weak minimax* w.r.t. Π^* within $\mathcal{D}^*$, if

$$\inf_{\delta\in\mathcal{D}^*} \sup_{\beta_2\in\mathcal{L}} \sup_{\pi\in\Pi^*} R^{\beta_2}(\pi,\delta) = \sup_{\beta_2\in\mathcal{L}} \sup_{\pi\in\Pi^*} R^{\beta_2}(\pi,\delta^*). \qquad \square$$

From the strong minimax optimality, it follows obviously the weak optimality. For the proof of the next theorem we need the following

Definition 8. Let $\alpha > 0$ be a given constant, ξ a certain prior distribution on $\mathcal{L}$ and $\pi \in \Pi$. An estimator $\delta_\alpha \in \mathcal{D}$ is called to be *α-Bayes* w.r.t. ξ, if

$$\int_{\mathcal{L}} R^{\beta_2}(\pi,\delta_\alpha)\xi(d\beta_2) \leq \int_{\mathcal{L}} R^{\beta_2}(\pi,\delta)\xi(d\beta_2) + \alpha \quad \text{for all } \delta \in \mathcal{D}. \qquad \square$$

Theorem 3. *For any given $\vartheta \in \Theta$ the estimator δ_ϑ is restricted weak minimax* w.r.t. Π_ϑ *in the set $\mathcal{D}$ (of all! estimators).*

Proof. First of all, we show for δ_ϑ the weak minimax optimality for the normal distribution π_G. We then obtain the assertion of the theorem from the property of the normal distribution to be in a certain sense a least favourable distribution within Π_ϑ.

The risk $R^{\beta_2}(\pi_G,\delta_\vartheta)$ is constant for all $\beta_2 \in \mathcal{L}$. Using A 3.43 in Bunke and Bunke (1986), it suffices to show that for any $\alpha > 0$, there exists a prior distribution ξ_α on $\mathcal{L}$, such that δ_ϑ is α-Bayes. Further, we denote by "mixed risk" the expression

$$r_{\pi_G}(\xi_\alpha,\delta) = \int_{\mathcal{L}} R^{\beta_2}(\pi_G,\delta)\xi_\alpha(d\beta_2),\ \delta \in \mathcal{D}. \tag{20}$$

Now we choose by

$$\xi_\alpha = N_k(0,\tau^{-1}(\alpha)P_2)$$

a normal prior distribution for β_2 such that β_2 and (β_1,ε) are stochastically independent, and $\tau > 0$ be a constant yet to be chosen depending on α, T, and Q. Here, P_2 has to be chosen positive semidefinite, e.g. $(I_k - T^+T) = P_2$. Then we have for β a regular normal distribution

$$\beta \sim N_k(b, T + \tau^{-1}P_2),$$

and β und ε are uncorrelated. Therefore, an optimal estimator $\delta_\vartheta^{\xi_\alpha}$ w.r.t. (20) exists for γ in $\mathcal{D}$ with

$$\delta_\vartheta^{\xi_\alpha} = C\big[(T+\tau^{-1}P_2)^{-1} + Q\big]^{-1}\big[(T+\tau^{-1}P_2)^{-1}b + Q\widehat{\beta}\big].$$

Because of $\mathcal{R}(T)\perp\mathcal{R}(P_2) = \mathcal{L}$ and $b \in \mathcal{R}(T)$ we have

$$(T+\tau^{-1}P_2)^{-1} = T^+ + \tau P_2.$$

Therefore, $\delta_\vartheta^{\xi_\alpha}$ takes the form

$$\delta_\vartheta^{\xi_\alpha} = C(T^+ + \tau P_2 + Q)^{-1}(T^+ b + Q\widehat{\beta})$$

with the mixed risk

$$r_{\pi_G}(\xi_\alpha, \delta_\vartheta^{\xi_\alpha}) = \operatorname{tr} HC(T^+ + \tau P_2 + Q)^{-1} C^\top,$$

and it holds

$$r_{\pi_G}(\xi_\alpha, \delta_\vartheta^{\xi_\alpha}) \le r_{\pi_G}(\xi_G, \delta_\theta) = R(\vartheta, \delta_\vartheta).$$

With the denotation $A := T^+ + Q$, let a_i, $i = 1, \dots, k$, be the characteristic roots of A. Then the projector P_2 can be splitted orthogonally into a part P_2^A projecting onto the range $\mathcal{R}(P_2 A) = \mathcal{R}(P_2 X^T)$ and the remaining $P_2 - P_2^A$.

Therefore, we obtain for the Bayesian risk

$$\begin{aligned} r_{\pi_G}(\xi_\alpha, \delta_\vartheta^{\xi_\alpha}) &= \operatorname{tr} HC\left(A + \tau P_2^A + \tau(P_2 - P_2^A)\right)^{-1} C^\top \\ &= \operatorname{tr} HC(A + \tau P_2^A)^+ C^\top + \frac{1}{\tau} \operatorname{tr} HC(P_2 - P_2^A) C^\top \\ &= \operatorname{tr} HC(A + \tau P_2^A)^+ C^\top, \end{aligned}$$

because from the estimability condition $\mathcal{R}(P_2 C^\top) \subseteq \mathcal{R}(P_2 X^\top)$ it follows $P_2 C^\top = P_2^A C^\top$. Now, the following relations hold

$$\begin{aligned} r_{\pi_G}(\xi_\alpha, \delta_\vartheta) - r_{\pi_G}(\xi_\alpha, \delta_\alpha^{\xi_\alpha}) &= \operatorname{tr} C^\top HC\left[A^+ - (A - \tau P_2^A)^+\right] \\ &\le \lambda_{\max} C^\top HC \operatorname{tr}\left(A^+ - (A + \tau P_2^A)^+\right) \\ &\le \lambda_{\max} C^\top HC \left[\sum\nolimits_{i:a_i>0} \frac{1}{a_i} - \sum\nolimits_{i:a_i>0} \frac{1}{a_i + \tau}\right] \\ &= \lambda_{\max} C^\top HC \sum\nolimits_{i:a_i>0} \frac{\tau}{a_i(a_i + \tau)} \\ &\le k\lambda_{\max} C^\top HC \frac{\tau}{\lambda_{\min}^{>0} A(\lambda_{\min}^{>0} A + \tau)}, \end{aligned}$$

where $\lambda_{\min}^{>0}$ denotes the smallest latent root greater than zero.

Let now

$$\alpha := k\lambda_{\max} C^\top HC \frac{\tau}{\lambda_{\min}^{>0} A(\lambda_{\min}^{>0} A + \tau)},$$

then we obtain for a given $\alpha > 0$ the constant τ by

$$\tau(\alpha) = \alpha \, \frac{(\lambda_{\min}^{>0} A)^2}{k\lambda_{\max} C^\top HC - \alpha \lambda_{\min}^{>0} A}. \tag{21}$$

There exists a positive number α_0, such that for all $\alpha \leq \alpha_0$ the term (21) becomes nonnegative. Therefore, for any $0 < \alpha < \alpha_0$ it holds

$$r_{\pi_G}(\xi_\alpha, \delta_\vartheta) \leq r_{\pi_G}(\xi_\alpha, \delta_\vartheta^{\xi_\alpha}) + \alpha.$$

Using the property of $\delta_\vartheta^{\xi_\alpha}$ to be a Bayesian estimator, we obtain

$$r_{\pi_G}(\xi_\alpha, \delta_\vartheta) \leq r_{\pi_G}(\xi_\alpha, \delta) + \alpha \quad \text{for all } \delta \in \mathcal{D}.$$

For $\alpha > \alpha_0$, the $\tau(\alpha_0)$ does it obviously. Therefore, we have shown: For any $\alpha > 0$, there exists a prior distribution ξ_α on $\mathcal{L}$ such that δ_ϑ has the property to be α-Bayes. Thus, using A 3.43 in Bunke and Bunke (1986), it follows that δ_ϑ is minimax in $\mathcal{D}$ for π_G, i.e.

$$\inf_{\delta \in \mathcal{D}} \sup_{\beta_2 \in \mathcal{L}} R^{\beta_2}(\pi_G, \delta) = \sup_{\beta_2 \in \mathcal{L}} R^{\beta_2}(\pi_G, \delta_\vartheta).$$

From the property of $R^{\beta_2}(\pi, \delta_\vartheta) = R(\vartheta, \delta_\vartheta)$ to be constant for all $\beta_2 \in \mathcal{L}$ and $\pi \in \Pi_\vartheta$, we obtain

$$\begin{aligned} \sup_{\beta_2 \in \mathcal{L}} R^{\beta_2}(\pi_G, \delta_\vartheta) &= \sup_{\beta_2 \in \mathcal{L}} \sup_{\pi \in \Pi_\vartheta} R^{\beta_2}(\pi, \delta_\vartheta) \\ &\geq \inf_{\delta \in \mathcal{D}} \sup_{\beta_2 \in \mathcal{L}} \sup_{\pi \in \Pi_\vartheta} R^{\beta_2}(\pi, \delta). \end{aligned}$$

Since $\pi_G \in \Pi_\vartheta$, we have on the other hand

$$\inf_{\delta \in \mathcal{D}} \sup_{\beta_2 \in \mathcal{L}} R^{\beta_2}(\pi_G, \delta) \leq \inf_{\delta \in \mathcal{D}} \sup_{\beta_2 \in \mathcal{L}} \sup_{\pi \in \Pi_\vartheta} R^{\beta_2}(\pi, \delta),$$

what therefore means that

$$\inf_{\delta \in \mathcal{D}} \sup_{\beta_2 \in \mathcal{L}} \sup_{\pi \in \Pi_\vartheta} R^{\beta_2}(\pi, \delta) = R(\vartheta, \delta_\vartheta).$$ □

REFERENCES

Bunke, O. (1964). Bedingte Strategien in der Spieltheorie: Existenzsätze und Anwendung auf statistische Entscheidungsproblem. *Transact. 3rd Prague Conf. Inf. Th., Statist. Decision Functions, Random Processes.*

Bunke, O. (1977). Mixed models, empirical Bayes and Stein estimators. Math. Operationsforsch. Statist., Ser. Statistics **1**, 55–68.

Bunke, H. and Bunke, O. (1986). *Statistical Inference in Linear Models.* Vol. I. Wiley, New York.

Bunke, H. and Gladitz, J. (1974). Empirical linear Bayes decision rules for a sequence of linear models with different regressor matrices. *Math. Operationsforsch. Statist.* **5**, 235–244.

Cogburn, R. (1967). Stringent solutions to statistical decision problems. *Ann. Math. Statist.* **38**, 447–463.

Gaffke, N. and Heiligers, B. (1987) Bayes-, admissible and linear estimators in linear models with restricted parameter space. Preprint Nr. 42, Universität Augsburg, Institut für Mathematik.

Hartigan, J. (1969). Linear Bayes methods. *J. Roy. Statist. Soc. B* **31**, 442–454.

Harville, D.A. (1978). Alternative formulations and procedures for the two-way mixed model. *Biometrics* **34**, 441–454.

Harville, D.A. (1979). Some useful representations for constrained mixed model estimation. *J. Amer. Statist. Assoc.* **74**, 200–206.

Henderson, C.R. (1963). Selection index and exptected genetic advance. In: *Statistical Genetics and Plant Breeding* (W.D. Hanson and H.S. Robinson, eds.). National Academy of Sciences, National Research Council, Washington. Publ. 982, 141–163.

Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimation for the linear model. *J. Roy. Statist. Soc. B* **34**, 1–18.

Pilz, J. (1983). *Bayesian Estimation and Experimental Design in Linear Regression Models.* Teubner-Texte zur Mathematik, Vol. 55, Teubner, Leipzig.

Rao, C.R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52**, 447–458.

Wind, S. (1973). An empirical Bayes approach to multiple linear regression. *Ann. Statist.* **1**, 93–103.

Dr. Johannes Gladitz
Institut für Soziologie und Sozialpolitik
AdW der DDR
Otto–Nuschke–Str. 10/11
Berlin
DDR–1086