ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XV, 4 (1977)

S. GNOT, W. KLONECKI and R. ZMYŚLONY (Wrocław)

BEST LINEAR PLUS QUADRATIC UNBIASED ESTIMATION OF PARAMETERS IN MIXED LINEAR MODELS

1. Introduction. In papers [2]-[5] Seely and Zyskind have proved a number of theorems on unbiased estimation when the choice of estimators is restricted to a finite-dimensional linear space. They also presented exemplification of the developed theory to the class of linear estimators $a'y$ in fixed linear models and to the class of quadratic estimators $y'Ay$ in random and mixed linear models. In this paper we show that the theory of Seely and Zyskind may be applied to mixed linear models when the choice of estimators is restricted to the class Γ of linear plus quadratic estimators $y'Ay+a'y$. For providing an inner product representation for such estimators we introduce a vector space $\mathscr{A}$ consisting of all $(2n\times 2n)$-matrices $\{A, a\}$ of form (1), where A is a symmetric $(n\times n)$-matrix, while $D(a)$ is a diagonal $(n\times n)$-matrix, with the inner product $\operatorname{tr}AB+a'b$ for all $\{A, a\}$, $\{B, b\}\in\mathscr{A}$. For the considered observable random vector y we form a random element $\{yy', y\}\in\mathscr{A}$ and discuss, in this setup, questions of estimability and provide the covariance operator of $\{yy', y\}$ under normality. The main results are Theorems 2 and 3. Both these theorems give necessary and sufficient conditions (NASC) for every estimable function to have a uniformly minimum variance unbiased estimator in Γ.

2. Preliminaries. First, in this section we recall, for the sake of completeness, a number of definitions and results given by Seely [2] and by Seely and Zyskind [5].

Let $\mathscr{A}$ be a Euclidean vector space endowed with an inner product denoted by $(\cdot,\cdot)$. Let $\{P_\theta\}$, where $\theta\in\omega$, be a family of probability measures associated with a measurable space $\{\mathscr{U},\mathscr{S}\}$ and, moreover, let Y be a random vector from $\mathscr{U}$ into $\mathscr{A}$. Assume that the expected value of $(\boldsymbol{A},\boldsymbol{Y})$ exists for every $\boldsymbol{A}\in\mathscr{A}$ and every $\theta\in\omega$. Under the above assumptions, for every $\theta\in\omega$ there exists an element $\mu_\theta\in\mathscr{A}$, called the *expected vector* of Y, such that $\mathrm{E}(\boldsymbol{A},\boldsymbol{Y})=(\boldsymbol{A},\mu_\theta)$ for every $\theta\in\omega$ and $\boldsymbol{A}\in\mathscr{A}$. Let

$$\mathscr{E}=\operatorname{sp}\{\mu_\theta:\ \theta\in\omega\}.$$

Occasionally, μ_θ is denoted by μ or by $\mathrm{E}\boldsymbol{Y}$.

Every function g from ω into the real line, say R^1, for which there exists an element $\boldsymbol{A} \in \mathscr{A}$ such that $\mathrm{E}(\boldsymbol{A}, \boldsymbol{Y}) = g$, is called *$\mathscr{A}$-estimable* (or, shortly, *estimable*) and $(\boldsymbol{A}, \boldsymbol{Y})$ is called its *linear unbiased estimator.*

If $\operatorname{cov}\{(\boldsymbol{A}, \boldsymbol{Y}), (\boldsymbol{B}, \boldsymbol{Y})\}$ exists for every $\boldsymbol{A}, \boldsymbol{B} \in \mathscr{A}$, for every $\theta \in \omega$ there exists a linear operator Σ_θ from $\mathscr{A}$ into $\mathscr{A}$, called the *covariance operator*, such that

$$\operatorname{Cov}\{(\boldsymbol{A}, \boldsymbol{Y}), (\boldsymbol{B}, \boldsymbol{Y})\} = (\boldsymbol{A}, \Sigma_\theta \boldsymbol{B}) \quad \text{for every } \boldsymbol{A}, \boldsymbol{B} \in \mathscr{A}.$$

The operator Σ_θ is determined uniquely. It is symmetric and positive-semidefinite. Let

$$\mathsf{S} = \operatorname{sp}\{\Sigma_\theta \colon \theta \in \omega\}.$$

An element $\boldsymbol{A} \in \mathscr{A}$ is called an *S-min element for an estimable function g* if $\mathrm{E}(\boldsymbol{A}, \boldsymbol{Y}) = g$ and if the inequality

$$\operatorname{Var}_\theta(\boldsymbol{A}, \boldsymbol{Y}) \leqslant \operatorname{Var}_\theta(\boldsymbol{B}, \boldsymbol{Y})$$

holds for every $\boldsymbol{B}$ such that $\mathrm{E}(\boldsymbol{B}, \boldsymbol{Y}) = g$. An element $\boldsymbol{A} \in \mathscr{A}$ is called an *S-min element* if it is an S-min element for $\mathrm{E}(\boldsymbol{A}, \boldsymbol{Y})$.

The following theorem, which is due to Lehmann and Scheffé, is fundamental in the theory of linear estimation in Euclidean vector spaces.

THEOREM (E. Lehmann and H. Scheffé). *$\boldsymbol{A} \in \mathscr{A}$ is an S-min element if and only if*

$$(\boldsymbol{A}, \Sigma_\theta \boldsymbol{A}_0) = 0$$

for every $\boldsymbol{A}_0 \in \mathscr{E}^\perp$ and every $\theta \in \omega$, where $\mathscr{E}^\perp$ is the orthogonal complement of $\mathscr{E}$.

If g is an estimable function and if there exists an S-min element for g, say $\boldsymbol{A}$, then the estimator $(\boldsymbol{A}, \boldsymbol{Y})$ is called *$\bar{\mathscr{A}}$-best for g* or the *best estimator for g*. Here $\bar{\mathscr{A}} = \{(\boldsymbol{A}, \boldsymbol{Y}) \colon \boldsymbol{A} \in \mathscr{A}\}$. The following corollary is easily deduced from the above theorem.

COROLLARY 1. *If S contains the identity operator, then for each $\mathscr{A}$-estimable function there exists an $\bar{\mathscr{A}}$-best estimator if and only if $\mathscr{E}$ is an invariant subspace of each element in S.*

Finally, we need the notion of a quadratic subspace of a vector space $\mathscr{A}$ of matrices of dimensions, say, $n \times n$ introduced by Seely in [3].

A subspace $\mathscr{A}_0$ of $\mathscr{A}$ such that $A \in \mathscr{A}_0$ implies $A^2 \in \mathscr{A}_0$ is called a *quadratic subspace* of $\mathscr{A}$.

In Section 6 we frequently use the following property of a quadratic subspace:

LEMMA 1. *Let $\mathscr{A}_0$ be a subspace of $\mathscr{A}$ and let $\mathscr{M}$ be an arbitrary spanning set for $\mathscr{A}_0$. Then $\mathscr{A}_0$ is a quadratic subspace if and only if $R, S, T \in \mathscr{M}$ implies $RST + TSR \in \mathscr{A}_0$.*

3. The problem. Let y be an n-vector of random variables with a linear structure,

$$y = X\beta + U_1\xi_1 + \ldots + U_k\xi_k,$$

where X is a given $(n \times p)$-matrix, β is a p-vector of unknown parameters, U_i for $i = 1, \ldots, k$ is a given $(n \times c_i)$-matrix, and ξ_i is a c_i-vector of normally distributed random variables with zero mean values and covariance matrix $\sigma_i^2 I$. Further, ξ_i and ξ_j are uncorrelated for $i \neq j$. In this setup,

$$\eta(\beta) = X\beta \quad \text{and} \quad V(\sigma) = \sum_{i=1}^{k} \sigma_i^2 V_i$$

are the vector of means and the covariance matrix of y, respectively, $\beta \in R^p$, while $\sigma = (\sigma_1^2, \ldots, \sigma_k^2) \in \Omega \subset R^k$ and $V_i = U_i U_i'$ for $i = 1, \ldots, k$.

Finally, we assume that

$$\mathscr{V} = \mathrm{sp}\{V_1, \ldots, V_k\} = \mathrm{sp}\{V(\sigma)\colon\ \sigma \in \Omega \subset R^k\} \quad \text{and} \quad I \in \mathscr{V},$$

where I is the identity matrix.

Now, let Γ be the set of all estimators of the form $y'Ay + a'y$ and let $\mathscr{G}$ be the entire collection of Γ-estimable functions, i.e. $g \in \mathscr{G}$ if and only if there exist a matrix A and a vector a such that

$$\mathrm{E}(y'Ay + a'y) = g.$$

The problem is to find necessary and sufficient conditions for every function in $\mathscr{G}$ to have a uniformly minimum variance unbiased estimator in Γ.

4. The setup. In order to bring the problem into the general framework in which the theory presented in Section 2 is applicable, we let $\mathscr{A}$ to be the space of all $(2n \times 2n)$-matrices of the form

$$\begin{pmatrix} A & 0 \\ 0 & D(a) \end{pmatrix}, \tag{1}$$

where A is a symmetric $(n \times n)$-matrix, while $D(a)$, $a = (a_1, \ldots, a_n)' \in R^n$, is a diagonal matrix with diagonal elements $a_1, \ldots, a_n$. Throughout the paper, matrix (1) is denoted by $\{A, a\}$ or, shortly, by $\boldsymbol{A}$. Clearly, with the usual definition of addition and multiplication by a scalar, $\mathscr{A}$ is a linear vector space and

$$(\boldsymbol{A}, \boldsymbol{B}) = \mathrm{tr}AB + a'b$$

defines an inner product.

Next, let

$$\boldsymbol{Y} = \{yy', y\}, \tag{2}$$

where y is the random vector defined in Section 3.

Note that, in this setup, a parametric function g, i.e. a function from ω into R^1, according to definition is $\mathscr{A}$-estimable if there exists an element $\boldsymbol{A} \in \mathscr{A}$ such that $\mathrm{E}(\boldsymbol{A}, \boldsymbol{Y}) = g$. In other words, g is estimable if there exists an unbiased estimator for g of the form $y'Ay + a'y$.

THEOREM 1. *Suppose that y is a normally distributed n-vector with mean value η and covariance matrix V. Then*

$$\mathrm{E}\boldsymbol{Y} = \{\eta\eta' + V, \eta\} \tag{3}$$

and Σ maps $\{A, a\}$ into the element $\{B, b\}$, where

$$B = 2(VAV + VA\eta\eta' + \eta\eta' AV) + Va\eta' + \eta a' V, \qquad b = 2VA\eta + Va. \tag{4}$$

Proof. The first assertion is easily verified by noting that $\mathrm{E}yy' = \eta\eta' + V$. In order to verify the second assertion, we have to show that, for every $\boldsymbol{A}$ and $\boldsymbol{B}$ in $\mathscr{A}$,

$$\mathrm{Cov}\{(\boldsymbol{A}, \boldsymbol{Y}), (\boldsymbol{B}, \boldsymbol{Y})\} = (\boldsymbol{A}, \Sigma\boldsymbol{B}) \quad \text{and} \quad (\Sigma\boldsymbol{A}, \boldsymbol{B}) = (\boldsymbol{A}, \Sigma\boldsymbol{B}).$$

By the assumption that y is normally distributed, we easily obtain

$$\mathrm{Cov}\{(\boldsymbol{A}, \boldsymbol{Y}), (\boldsymbol{B}, \boldsymbol{Y})\} = 2\,\mathrm{tr}\,AVBV + 4\eta' BVA\eta + 2\eta' BVa + 2\eta' AVb + b'Va.$$

On the other hand, using (4) we see that

$$(\Sigma\boldsymbol{A}, \boldsymbol{B}) = 2\,\mathrm{tr}\,AVBV + 4\eta' BVA\eta + 2\eta' BVa + 2\eta' AVb + b'Va.$$

Then, by symmetry,

$$(\boldsymbol{A}, \Sigma\boldsymbol{B}) = 2\,\mathrm{tr}\,BVAV + 4\eta' AVB\eta + 2\eta' AVa + 2\eta' BVa + a'Vb.$$

These results imply that the second assertion follows immediately by noting that $\mathrm{tr}\,RS = \mathrm{tr}\,RS'$ if $R = R'$.

The next lemma describes the structure of the subspace $\mathscr{E}$ in the considered case.

Let $x_1, \ldots, x_p$ be the columns of the $(n \times p)$-matrix X, let

$$B_{ii} = x_i x_i' \qquad \text{for } 1 \leqslant i \leqslant p,$$
$$B_{ij} = x_i x_j' + x_j x_i' \qquad \text{for } 1 \leqslant i < j \leqslant p$$

and, moreover, write

$$\boldsymbol{V}_i = \{V_i, 0\} \quad \text{for } 1 \leqslant i \leqslant p,$$
$$\boldsymbol{B}_{ij} = \{B_{ij}, 0\} \quad \text{for } 1 \leqslant i \leqslant j \leqslant p,$$
$$\boldsymbol{x}_i = \{0, x_i\} \quad \text{for } 1 \leqslant i \leqslant p.$$

LEMMA 2. *If y is the random vector defined in Section 3 and $\boldsymbol{Y}$ is defined by (2), then*

$$\mathscr{E} = \mathrm{sp}\{\boldsymbol{V}_1, \ldots, \boldsymbol{V}_p, \boldsymbol{B}_{11}, \ldots, \boldsymbol{B}_{pp}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_p\}.$$

This lemma follows immediately from (3) and the assumption that β is ranging over R^p.

5. $\mathscr{A}$-estimable functions. Noting that

$$\mathrm{E}(\boldsymbol{A}, \boldsymbol{Y}) = \sum_{i=1}^{k} \sigma_i^2 \operatorname{tr}(AV_i) + \beta' X' A X \beta + a' X \beta,$$

where $\boldsymbol{A} = \{A, a\}$, it is easily established that an $\mathscr{A}$-estimable function g can be presented in the form

$$g = c'\sigma + d'\gamma + e'\beta,$$

where

$$c = (c_1, \ldots, c_k)' \in R^k, \quad d = (d_{11}, \ldots, d_{pp})' \in R^{p(p-1)/2},$$

$$\gamma = (\beta_1^2, \beta_1\beta_2, \ldots, \beta_{p-1}\beta_p, \beta_p^2)' \in R^{p(p-1)/2}, \quad e = (e_1, \ldots, e_p)' \in R^p.$$

Now, applying Theorem 3 of Seely [3] we obtain

LEMMA 3. *The function $c'\sigma + d'\gamma + e'\beta$ is $\mathscr{A}$-estimable if and only if*

$$e \in R(X'X) \quad \textit{and} \quad (c_1, \ldots, c_k, d_{11}, \ldots, d_{pp})' \in R(\boldsymbol{W}),$$

where

$$\boldsymbol{W} = \begin{pmatrix} \operatorname{tr} V_1 V_1, \ldots, & \operatorname{tr} V_1 V_k, & \operatorname{tr} V_1 B_{11}, \ldots, & \operatorname{tr} V_1 B_{pp} \\ \operatorname{tr} V_k V_1, \ldots, & \operatorname{tr} V_k V_k, & \operatorname{tr} V_k B_{11}, \ldots, & \operatorname{tr} V_k B_{pp} \\ \operatorname{tr} B_{11} V_1, \ldots, & \operatorname{tr} B_{11} V_k, & \operatorname{tr} B_{11} B_{11}, \ldots, & \operatorname{tr} B_{11} B_{pp} \\ \operatorname{tr} B_{pp} V_1, \ldots, & \operatorname{tr} B_{pp} V_k, & \operatorname{tr} B_{pp} B_{11}, \ldots, & \operatorname{tr} B_{pp} B_{pp} \end{pmatrix}.$$

It may be interesting that $c'\sigma + d'\gamma + e'\beta$ is $\mathscr{A}$-estimable if and only if both $c'\sigma + d'\gamma$ and $e'\beta$ are estimable. This is a consequence of the fact that $\boldsymbol{x}_i \boldsymbol{V}_i = \boldsymbol{x}_i \boldsymbol{B}_{ij} = 0$ for $1 \leqslant i \leqslant j \leqslant p$. From Lemma 3 it follows that the parametric function $c'\sigma$ is $\mathscr{A}$-estimable if and only if

$$c \in R\big(\operatorname{tr}\big((V_i - PV_iP)(V_j - PV_jP)\big)\big) \quad \text{for } i, j = 1, \ldots, k,$$

where P is the orthogonal projection on $R(X)$ with respect to the usual inner product.

6. Main results. To formulate the NASC for every function in $\mathscr{G}$ to have a uniformly minimum variance unbiased estimator in Γ we need some additional notation.

Let $\mathscr{B} = \operatorname{sp}\{B_{11}, \ldots, B_{pp}\}$ and let $\mathscr{C} = \mathscr{V} + \mathscr{B}$. Note that $\{A, a\} \in \mathscr{E}$ if and only if $a \in R(X)$ and $A \in \mathscr{C}$.

THEOREM 2. *For each $\mathscr{A}$-estimable function there exists an $\bar{\mathscr{A}}$-best estimator if and only if P commutes with each $V \in \mathscr{V}$ and $\mathscr{C}$ is a quadratic subspace of the space $\mathscr{A}^s$ of all $(n \times n)$ symmetric matrices.*

Proof. First note that if Σ is a covariance operator which maps $\{A, a\}$ into $\{B, b\}$, where B and b are defined by (4), then $\mathscr{E}$ is an invariant space of Σ if and only if

$$a \in R(X) \quad \text{and} \quad A \in \mathscr{C} \tag{5}$$

implies

$$b \in R(X) \quad \text{and} \quad B \in \mathscr{C}. \tag{6}$$

Note also that from the assumption that $I \in \mathscr{V}$ it follows that the identity operator belongs to

$$\mathsf{S} = \mathrm{sp}\{\Sigma_{\beta,\sigma}\colon \beta \in R^p, \sigma \in \Omega \subset R^k\}.$$

Thus, in view of Corollary 1, it is enough to show that (5) implies (6) for every $\Sigma \in \mathsf{S}$ if and only if

(7) P commutes with each V_i for $i = 1, \ldots, k$,

and

(8) $\mathscr{C}$ is a quadratic subspace of $\mathscr{A}^s$.

Suppose that (5) implies (6). Let $\{A, a\} = \{I, 0\} \in \mathscr{E}$. From (4) it follows that, for every $V \in \mathscr{V}$ and every $\eta \in R(X)$,

$$V\eta \in R(X) \tag{9}$$

and

$$VV + V\eta\eta' + \eta\eta' V \in \mathscr{C}. \tag{10}$$

Clearly, (9) implies (7). Next, from (10) we conclude that, for $i, j = 1, \ldots, k$,

$$V_i V_j + V_j V_i \in \mathscr{C} \quad \text{and} \quad V_i \eta\eta' + \eta\eta' V_i \in \mathscr{C}.$$

These relations and the fact that $\mathscr{B}$ is a quadratic subspace imply that $\mathscr{C}$ is a quadratic subspace.

Now, let $a \in R(X)$ and let $A \in \mathscr{C}$. From (7) we conclude that $Va \in R(X)$. Using the decomposition $A = PAP + MAM$ we easily find that

$$VA\eta = V(PAP + MAM)\eta = VPAP\eta = PVAP\eta, \quad \text{where } M = I - P.$$

Hence it follows immediately that $VA\eta \in R(X)$. By (4) we have

$$b = 2VA\eta + Va \in R(X).$$

On the other hand, from (7) and (8) it follows that, for $V \in \mathscr{V}$ and for $\eta \in R(X)$,

$$VAV + VA\eta\eta' + \eta\eta' AV \in \mathscr{C} \quad \text{and} \quad Va\eta' + \eta a' V \in \mathscr{B}.$$

Hence

$$B = 2(VAV + VA\eta\eta' + \eta\eta' AV) + Va\eta' + \eta a' V \in \mathscr{C},$$

which completes the proof of Theorem 2.

Another set of NASC for every $\mathscr{A}$-estimable function to have an $\bar{\mathscr{A}}$--best estimator provides the following theorem:

THEOREM 3. *For every $\mathscr{A}$-estimable function there exists an $\bar{\mathscr{A}}$-best estimator if and only if P commutes with every $V \in \mathscr{V}$, and*

$$\mathscr{X}_1 = \mathrm{sp}\{V_1 - PV_1P, \ldots, V_k - PV_kP\}$$

is a quadratic subspace.

Proof. If P commutes with every V_i for $i = 1, \ldots, k$, then

$$\mathscr{X}_1 = \mathrm{sp}\{MV_1, \ldots, MV_k\}.$$

Now observe that $AB = 0$ if $A \in \mathscr{B}$ and $B \in \mathscr{X}_1$, and that $\mathscr{C}$ is a direct sum of $\mathscr{B}$ and $\mathscr{X}_1$, i.e. $\mathscr{C} = \mathscr{B} \oplus \mathscr{X}_1$. However, this shows that if $\mathscr{C}$ or $\mathscr{X}_1$ is a quadratic subspace, then both of them are quadratic subspaces.

It is interesting that NASC identical as in Theorem 3 appear in a theorem of Kleffe and Pincus [1] for the existence of an $\bar{\mathscr{A}}$-best estimator for every $\mathscr{A}$-estimable function of the form $c'\sigma$.

References

[1] J. Kleffe and R. Pincus, *Bayes and best quadratic unbiased estimators for parameters of the covariance matrix in a normal linear model*, Math. Operationsforschung Statist. 5 (1974), p. 43-67.

[2] J. Seely, *Linear spaces and unbiased estimation*, Ann. Math. Statist. 41 (1970), p. 1725-1734.

[3] — *Linear spaces and unbiased estimation — application to the mixed linear model*, ibidem 41 (1970), p. 1735-1748.

[4] — *Quadratic subspaces and completeness*, ibidem 42 (1971), p. 710-721.

[5] — and G. Zyskind, *Linear spaces and minimum variance unbiased estimation*, ibidem 42 (1971), p. 691-703.

INSTITUTE OF IMMUNOLOGY AND EXPERIMENTAL THERAPY
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
51-617 WROCŁAW

Received on 5. 12. 1975

S. GNOT, W. KLONECKI i R. ZMYŚLONY (Wrocław)

NAJLEPSZA LINIOWA I KWADRATOWA ESTYMACJA PARAMETRÓW W MIESZANYCH MODELACH LINIOWYCH

STRESZCZENIE

W pracach [2]-[5] Seely i Zyskind rozwinęli teorię nieobciążonej estymacji, gdy klasa rozważanych estymatorów $\bar{\mathscr{A}}$ jest skończenie wymiarową przestrzenią liniową. Twierdzenia dotyczące estymacji liniowej dla stałych modeli liniowych oraz estymacji kwadratowej dla losowych i mieszanych modeli liniowych otrzymuje się wówczas, jako wnioski, przez wyspecjalizowanie przestrzeni $\bar{\mathscr{A}}$. W tej pracy pokazuje się, że przez odpowiednie wyspecjalizowanie przestrzeni liniowej $\bar{\mathscr{A}}$ otrzymuje się również twierdzenia dotyczące estymacji funkcjami postaci $y'Ay+a'y$, gdzie A jest macierzą symetryczną. Główne wyniki zawarte są w twierdzeniach 2 i 3, gdzie określone są (przy założeniu, że rozkład wektora y jest normalny) warunki konieczne i dostateczne na to, aby dla każdej funkcji, mającej nieobciążony estymator postaci $y'Ay+$ $+a'y$, istniał *najlepszy nieobciążony estymator* w klasie estymatorów $y'Ay+a'y$ (tzn. estymator nieobciążony o jednostajnie najmniejszej wariancji).