The Annals of Statistics
1980, Vol. 8, No. 3, 679–681

A NEW BOUND FOR THE EUCLIDEAN NORM OF THE DIFFERENCE BETWEEN THE LEAST SQUARES AND THE BEST LINEAR UNBIASED ESTIMATORS[1]

By J. K. Baksalary and R. Kala

Academy of Agriculture in Poznań

A new bound is established for the Euclidean norm of the difference between the least squares estimator and the best linear unbiased estimator of the vector of expectations in the general linear model. The bound is valid regardless of the rank of the dispersion matrix and is expressed in substantially simpler terms than the bounds given earlier by Haberman and by Baksalary and Kala.

1. Statement of the problem. Let the triplet

$$(\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}) \tag{1}$$

denote the general linear model in which $\mathbf{y}$ is an $n \times 1$ observable random vector with expectation $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and with dispersion matrix $D(\mathbf{y}) = \mathbf{V}$; $\mathbf{X}$ is a known $n \times p$ matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters and $\mathbf{V}$ is an $n \times n$ nonnegative definite matrix, known or known except for a positive scalar multiplier. Further, let $\hat{\boldsymbol{\mu}}$ stand for the best linear unbiased estimator (BLUE) of $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, and let $\boldsymbol{\mu}^*$ stand for the least squares estimator (LSE) of $\boldsymbol{\mu}$. The LSE is independent of $\mathbf{V}$ and has the representation

$$\boldsymbol{\mu}^* = \mathbf{P}\mathbf{y},$$

where $\mathbf{P} = \mathbf{X}\mathbf{X}^+$ is the orthogonal projector on the column space of $\mathbf{X}$, $\mathbf{X}^+$ denoting the Moore-Penrose inverse of $\mathbf{X}$.

Under the assumption that the dispersion matrix of the model (1) is nonsingular, in which case the BLUE of $\boldsymbol{\mu}$ may be written

$$\hat{\boldsymbol{\mu}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^+ \mathbf{X}'\mathbf{V}^{-1}\mathbf{y},$$

Haberman [5] has given a bound for an arbitrary norm of the vector $\boldsymbol{\mu}^* - \hat{\boldsymbol{\mu}}$, which, however, is inapplicable unless an additional condition involving the matrices $\mathbf{P}$ and $\mathbf{V}$ is satisfied. For the case of the Euclidean vector norm, denoted by $\|\cdot\|_2$, Haberman's result has recently been strengthened by Baksalary and Kala [2]. The bound derived in [2],

$$\|\boldsymbol{\mu}^* - \hat{\boldsymbol{\mu}}\|_2 \leqslant \left((\nu_0)^{\frac{1}{2}}/\lambda_0\right)\|\mathbf{y} - \boldsymbol{\mu}^*\|_2, \tag{2}$$

Received June 1978; revised January 1979.

[1]This research was supported by the Polish Academy of Sciences under Grant MRI.1-2/1/5.

AMS 1970 *subject classifications.* Primary 62J05.

Key words and phrases. Linear model, least squares estimator, best linear unbiased estimator, Euclidean vector norm, spectral matrix norm.

where ν_0 is the largest eigenvalue of $\mathbf{PV}^{-1}(\mathbf{I}-\mathbf{P})\mathbf{V}^{-1}\mathbf{P}$ and λ_0 is the smallest nonzero eigenvalue of $\mathbf{PV}^{-1}\mathbf{P}$, is not only valid for any linear model with nonsingular $\mathbf{V}$ but is, in addition, sharper than the bound of Haberman.

Delete now the assumption that the dispersion matrix of the model (1) is nonsingular. The BLUE of $\boldsymbol{\mu}$ admits then a representation (Rao and Mitra [6])

$$\hat{\boldsymbol{\mu}} = \mathbf{X}(\mathbf{X}'\mathbf{T}^{+}\mathbf{X})^{+}\mathbf{X}'\mathbf{T}^{+}\mathbf{y} \tag{3}$$

where

$$\mathbf{T} = \mathbf{V} + \mathbf{XX}'.$$

It can easily be verified that the arguments used in [2] to establish the inequality (2) are, without much change, applicable also to the present situation, leading, in consequence, to the result stating that

$$\|\boldsymbol{\mu}^* - \hat{\boldsymbol{\mu}}\|_2 \leqslant \big((\nu)^{\frac{1}{2}}/\lambda\big)\|\mathbf{y} - \boldsymbol{\mu}^*\|_2, \tag{4}$$

where ν is the largest eigenvalue of $\mathbf{PT}^{+}(\mathbf{I}-\mathbf{P})\mathbf{T}^{+}\mathbf{P}$ and λ is the smallest nonzero eigenvalue of $\mathbf{PT}^{+}\mathbf{P}$.

A weak point of the bounds specified in (2) and (4) is that they utilize the inverse of $\mathbf{V}$ or the Moore-Penrose inverse of $\mathbf{T}$, respectively. In the next section we will present a new bound for $\|\boldsymbol{\mu}^* - \hat{\boldsymbol{\mu}}\|_2$ which involves neither of these inverses and is valid for any linear model whatsoever.

2. Result.

The new bound is revealed in the following.

THEOREM. *Let $\boldsymbol{\mu}^*$ and $\hat{\boldsymbol{\mu}}$ be the* LSE *and the* BLUE, *respectively, of $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ in the general linear model* $(\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V})$. *Then*

$$\|\boldsymbol{\mu}^* - \hat{\boldsymbol{\mu}}\|_2 \leqslant \big((\gamma)^{\frac{1}{2}}/\delta\big)\|\mathbf{y} - \boldsymbol{\mu}^*\|_2, \tag{5}$$

where γ is the largest eigenvalue of $\mathbf{PV}(\mathbf{I}-\mathbf{P})\mathbf{VP}$ and δ is the smallest nonzero eigenvalue of $(\mathbf{I}-\mathbf{P})\mathbf{V}(\mathbf{I}-\mathbf{P})$. Moreover, with the convention that $\gamma^{\frac{1}{2}}/\delta = 0$ if $(\mathbf{I}-\mathbf{P})\mathbf{V} = \mathbf{0}$, in which case $\gamma = \delta = 0$, the right-hand side of (5) *is equal to zero if and only if $\boldsymbol{\mu}^* = \hat{\boldsymbol{\mu}}$.*

PROOF. Using Albert's [1] formula (4), the BLUE of $\boldsymbol{\mu}$ may be written

$$\hat{\boldsymbol{\mu}} = \mathbf{Py} - \mathbf{PV}(\mathbf{I}-\mathbf{P})\big[(\mathbf{I}-\mathbf{P})\mathbf{V}(\mathbf{I}-\mathbf{P})\big]^{+}(\mathbf{I}-\mathbf{P})\mathbf{y}. \tag{6}$$

(The reader interested in a relationship between the representations of $\hat{\boldsymbol{\mu}}$ given in (3) and (6) is referred to Baksalary and Kala [3]). Since $\mathbf{Py} = \boldsymbol{\mu}^*$, it follows from (6) that

$$\|\boldsymbol{\mu}^* - \hat{\boldsymbol{\mu}}\|_2 = \|\mathbf{PV}(\mathbf{I}-\mathbf{P})\big[(\mathbf{I}-\mathbf{P})\mathbf{V}(\mathbf{I}-\mathbf{P})\big]^{+}(\mathbf{y} - \boldsymbol{\mu}^*)\|_2. \tag{7}$$

It is known that the matrix norm corresponding to the Euclidean vector norm is the spectral norm defined as

$$\|\mathbf{A}\|_2 = \max\big\{\alpha^{\frac{1}{2}} : \alpha \in \sigma(\mathbf{AA}')\big\}, \tag{8}$$

where $\sigma(\cdot)$ denotes the spectrum of the matrix argument. Since this norm is consistent and multiplicative (see, e.g., Ben-Israel and Greville [4], pages 33–34), it follows from (7) that

$$\|\boldsymbol{\mu}^* - \hat{\boldsymbol{\mu}}\|_2 \leqslant \|\mathbf{PV}(\mathbf{I} - \mathbf{P})\|_2 \|[(\mathbf{I} - \mathbf{P})\mathbf{V}(\mathbf{I} - \mathbf{P})]^+\|_2 \|\mathbf{y} - \boldsymbol{\mu}^*\|_2. \tag{9}$$

On the other hand, from the spectral representation of the Moore-Penrose inverse, it is clear that for a nonnegative definite matrix $\mathbf{A} \neq \mathbf{0}$,

$$\|\mathbf{A}^+\|_2 = [\min\{\alpha : \alpha \in \sigma(\mathbf{A}), \alpha > 0\}]^{-1}. \tag{10}$$

The application of (8) and (10) to (9) proves (5). To establish the remainder of the theorem, we refer to the result of Zyskind and Martin [7] which states that

$$\boldsymbol{\mu}^* = \hat{\boldsymbol{\mu}} \Leftrightarrow \mathbf{VP} = \mathbf{PV}. \tag{11}$$

Observing that the condition on the right-hand side of (11) can equivalently be written $\mathbf{VP} = \mathbf{PVP}$, it follows immediately that $\boldsymbol{\mu}^* = \hat{\boldsymbol{\mu}}$ if and only if $\gamma = 0$, thus completing the proof.

It should be noted, in the conclusion, that examples can be constructed to show that the bounds in (2) and (4) are not sharper then the bound in (5), nor, unfortunately, is the reverse true. Nevertheless, the computational simplicity and unrestricted applicability motivate the recommendation of (5) as a substantially more convenient tool in investigating the problem "BLUE versus LSE".

Acknowledgement. The authors are grateful to the referees for helpful suggestions.

REFERENCES

[1] ALBERT, A. (1973). The Gauss-Markov theorem for regression models with possibly singular covariances. *SIAM J. Appl. Math.* **24** 182–187.

[2] BAKSALARY, J. K. and KALA, R. (1978a). A bound for the Euclidean norm of the difference between the least squares and the best linear unbiased estimators. *Ann. Statist.* **6** 1390–1393.

[3] BAKSALARY, J. K. and KALA, R. (1978b). Relationships between some representations of the best linear unbiased estimator in the general Gauss-Markoff model. *SIAM J. Appl. Math.* **35** 515–520.

[4] BEN-ISRAEL, A. and GREVILLE, T. N. E. (1974). *Generalized Inverses: Theory and Applications.* Wiley, New York.

[5] HABERMAN, S. J. (1975). How much do Gauss-Markov and least squares estimates differ? A coordinate-free approach. *Ann. Statist.* **3** 982–990.

[6] RAO, C. R. and MITRA, S. K. (1971). Further contributions to the theory of generalized inverses of matrices and its applications. *Sankhyā, Ser. A* **33** 289–300.

[7] ZYSKIND, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Statist.* **38** 1092–1109.

DEPARTMENT OF MATHEMATICAL
AND STATISTICAL METHODS
ACADEMY OF AGRICULTURE
60-637 POZNAŃ
WOJSKA POLSKIEGO 28
POLAND