The Annals of Statistics
1978, Vol. 6, No. 6, 1390–1393

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

BY J. K. BAKSALARY AND R. KALA

Academy of Agriculture, Poznań, Poland

Haberman's bound for a norm of the difference between the least squares and the best linear unbiased estimators in a linear model with nonsingular covariance structure is examined in the particular case when a vector norm involved is taken as the Euclidean one. In this frequently occurring case, a new substantially improved bound is developed which, furthermore, is applicable regardless of any additional condition.

1. Statement of the problem. Let the triplet

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

denote a linear model in which $\mathbf{y}$ is an $n \times 1$ observable random vector having $\mathbf{X}\boldsymbol{\beta}$ (henceforth signed by $\boldsymbol{\mu}$) as its expectation and $\mathbf{V}$ as its dispersion matrix; $\mathbf{X}$ is an $n \times p$ known matrix of rank r, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\mathbf{V}$ is an $n \times n$ positive definite symmetric matrix, known or known except for a positive scalar multiplier. Further, let $\mathbf{P}$ and $\mathbf{Q}$ be two projectors on Ω, the column space of $\mathbf{X}$, the first of which projects orthogonally under the Euclidean inner product signed by $(\cdot, \cdot)$, while the second projects orthogonally under the inner product defined as

$$((\mathbf{u}, \mathbf{v})) = (\mathbf{u}, \mathbf{V}^{-1}\mathbf{v})\,. \tag{2}$$

It is well known that

$$\boldsymbol{\mu}^* = \mathbf{P}\mathbf{y} \tag{3}$$

and

$$\hat{\boldsymbol{\mu}} = \mathbf{Q}\mathbf{y} \tag{4}$$

are the least squares and the best linear unbiased estimators of $\boldsymbol{\mu}$, respectively.

Much attention has been paid in the literature to the problem of the equality of (3) and (4), the practical importance of which is evident from the computational advantage of $\mathbf{P}$ over $\mathbf{Q}$. The interest of most authors dealing with this problem has been focused on developing criteria for the above-mentioned equality which, however, provide only categorical ascertainments and, in particular, give the negative answer in case of a great as well as of a small departure of $\boldsymbol{\mu}^*$ from $\hat{\boldsymbol{\mu}}$. In this light a method due to Haberman (1975, Section 3), based on a

Received December 1976; revised September 1977.

AMS 1970 *subject classifications.* Primary 62J05, 62J10.

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

bound for a norm of the difference between these two estimators, appears to be more informative for the problem and, therefore, more appropriate for practical applications. The basic result for this method is the following.

If

$$\|\mathbf{P} - \mathbf{P}\mathbf{V}^{-1}\mathbf{P}\| < 1 , \tag{5}$$

then

$$\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\| \leqq \frac{\|\mathbf{P}\mathbf{V}^{-1}(\mathbf{I} - \mathbf{P})\|}{1 - \|\mathbf{P} - \mathbf{P}\mathbf{V}^{-1}\mathbf{P}\|} \|\mathbf{y} - \boldsymbol{\mu}^*\| , \tag{6}$$

where the sign $\|\cdot\|$ *is used (according to the context) for a vector norm or for the corresponding matrix norm.*

A weak point of Haberman's method is that the applicability of the bound in (6) is restricted to those cases only for which condition (5) is satisfied. This is a serious disadvantage, particularly as the unfulfillment of (5) does not preclude the possibility of $\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|$ being small enough to suggest the use of the least squares estimator. This can be shown by an example in which

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} a^{-1} & 0 & 0 \\ 0 & b^{-1} & 0 \\ 0 & 0 & c^{-1} \end{pmatrix},$$

where $a, b, c > 0$. Then

$$\boldsymbol{\mu}^{*\prime} = \left(\frac{y_1 + y_2}{2} \quad \frac{y_1 + y_2}{2} \quad y_3 \right), \tag{7}$$

$$\hat{\boldsymbol{\mu}}' = \left(\frac{ay_1 + by_2}{a + b} \quad \frac{ay_1 + by_2}{a + b} \quad y_3 \right), \tag{8}$$

$$\mathbf{P} - \mathbf{P}\mathbf{V}^{-1}\mathbf{P} = \frac{1}{4} \begin{pmatrix} 2 - a - b & 2 - a - b & 0 \\ 2 - a - b & 2 - a - b & 0 \\ 0 & 0 & 4 - 4c \end{pmatrix}.$$

From (7) and (8) it is seen that, regardless of the value of c, $\boldsymbol{\mu}^* = \hat{\boldsymbol{\mu}}$ whenever $a = b$. But, on the other hand, the spectral radius of the matrix $\mathbf{P} - \mathbf{P}\mathbf{V}^{-1}\mathbf{P}$, which is a lower bound for all its multiplicative norms (see, e.g., Ben-Israel and Greville (1975), page 36), is

$$\rho(\mathbf{P} - \mathbf{P}\mathbf{V}^{-1}\mathbf{P}) = \max \{|1 - (a + b)/2|, |1 - c|\} ,$$

and hence it is evident that, for any $c \geqq 2$ and $a = b$, the formula (6) is not applicable, although both of the estimators are equal.

The purpose of this paper is to answer the question if it is possible to construct a bound similar to that in (6) which, however, could be applied to any linear model of the form (1), irrespective of any condition of the type (5). This problem will be studied in the special case when the vector norm involved is chosen as the Euclidean norm, which is the one most frequently employed in practice.

As is known, the corresponding matrix norm is then the spectral norm. Both these norms will be signed by $\|\cdot\|_2$.

2. Solution. Assume that neither Ω nor $\Omega^{\perp}$, the orthogonal (under the Euclidean inner product) complement of Ω, is the trivial linear subspace $\{\mathbf{0}\}$, for then it follows from Haberman's (1975) Theorem 2 that $\boldsymbol{\mu}^*$ coincides with $\hat{\boldsymbol{\mu}}$ and, therefore, the problem becomes uninteresting.

The key to answering the question stated above is the fact that the projector $\mathbf{Q}$ that yields $\hat{\boldsymbol{\mu}}$ is unchangeable under the multiplication of $\mathbf{V}^{-1}$, defining inner product (2), by any scalar $\alpha > 0$. Replacing $\mathbf{V}^{-1}$ in formulae (5) and (6) by $\alpha\mathbf{V}^{-1}$, the problem can be formulated as follows: to find the value of $\alpha > 0$ such that

$$\frac{\alpha\|\mathbf{P}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{P})\|_2}{1-\|\mathbf{P}-\alpha\mathbf{P}\mathbf{V}^{-1}\mathbf{P}\|_2}$$

is minimized, subject to the constraint that $\|\mathbf{P}-\alpha\mathbf{P}\mathbf{V}^{-1}\mathbf{P}\|_2 < 1$. Since the spectral norm of a symmetric matrix is equal to its spectral radius, we have

$$\|\mathbf{P}-\alpha\mathbf{P}\mathbf{V}^{-1}\mathbf{P}\|_2 = \max\{\kappa_1, -\kappa_n\},$$

where κ_1 and κ_n are the largest and the smallest eigenvalues of $\mathbf{P}-\alpha\mathbf{P}\mathbf{V}^{-1}\mathbf{P}$, respectively. But

$$\begin{aligned}\kappa_1 &= \sup_{\|\mathbf{u}\|_2=1}(\mathbf{u}, (\mathbf{P}-\alpha\mathbf{P}\mathbf{V}^{-1}\mathbf{P})\mathbf{u})\\ &= \sup_{\|\mathbf{u}\|_2=1}(\mathbf{P}\mathbf{u}, (\mathbf{I}-\alpha\mathbf{V}^{-1})\mathbf{P}\mathbf{u})\\ &= \sup_{\mathbf{u}\in\Omega,\|\mathbf{u}\|_2=1}(\mathbf{u}, (\mathbf{I}-\alpha\mathbf{V}^{-1})\mathbf{u})\\ &= 1-\alpha m,\end{aligned}$$

where

(9) $$m = \inf_{\mathbf{u}\in\Omega,\|\mathbf{u}\|_2=1}(\mathbf{u}, \mathbf{V}^{-1}\mathbf{u}) > 0.$$

Similarly,

$$\kappa_n = 1-\alpha M,$$

where

$$M = \sup_{\mathbf{u}\in\Omega,\|\mathbf{u}\|_2=1}(\mathbf{u}, \mathbf{V}^{-1}\mathbf{u}) \geqq m.$$

Therefore

$$\|\mathbf{P}-\alpha\mathbf{P}\mathbf{V}^{-1}\mathbf{P}\|_2 = \max\{1-\alpha m, -1+\alpha M\},$$

and hence it is seen that $\|\mathbf{P}-\alpha\mathbf{P}\mathbf{V}^{-1}\mathbf{P}\|_2 < 1$ whenever $0 < \alpha < 2/M$.

Thus the problem reduces to finding $0 < \alpha < 2/M$ to minimize

(10) $$\frac{\alpha N}{1-\max\{1-\alpha m, -1+\alpha M\}},$$

where $N = \|\mathbf{P}\mathbf{V}^{-1}(\mathbf{I}-\mathbf{P})\|_2$ or, we note incidentally,

(11) $$N = \sup_{\mathbf{u}\in\Omega,\|\mathbf{u}\|_2=1}\sup_{\mathbf{v}\in\Omega^{\perp},\|\mathbf{v}\|_2=1}(\mathbf{u}, \mathbf{V}^{-1}\mathbf{v}).$$

The minimum of (10) is achieved for any $0 < \alpha \leqq 2/(M+m)$ and equals N/m. This observation leads to the following.

THEOREM. *Let $\boldsymbol{\mu}^*$ and $\hat{\boldsymbol{\mu}}$, respectively, be the least squares and the best linear unbiased estimators of $\boldsymbol{\mu} \equiv \mathbf{X}\boldsymbol{\beta}$ in the model $(\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, and let $\Omega \neq \{\mathbf{0}\}$ and $\Omega^{\perp} \neq \{\mathbf{0}\}$. Then*

$$\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|_2 \leqq (N/m)\|\mathbf{y} - \boldsymbol{\mu}^*\|_2, \tag{12}$$

where N and m are given in (11) *and* (9), *or, equivalently,*

$$\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|_2 \leqq (\nu_1^{\frac{1}{2}}/\lambda_r)\|\mathbf{y} - \boldsymbol{\mu}^*\|_2,$$

where ν_1 is the largest eigenvalue of $\mathbf{PV}^{-1}(\mathbf{I} - \mathbf{P})\mathbf{V}^{-1}\mathbf{P}$, while λ_r is the smallest eigenvalue of $\mathbf{PV}^{-1}\mathbf{P}$, excluding zeros.

It is now interesting to compare the bound obtained in this paper with the special case (when the Euclidean vector norm is taken) of that given in (6). This comparison is possible only if (6) is applicable, that is if $M < 2$. Since the bound in (6) is the value of (10) for $\alpha = 1$, and the bound in (12) is the minimum of (10), the latter cannot be weaker. Moreover, from the proof of the theorem it follows that in the case of $M < 2 < M + m$, the bound in (12) is stronger.

Finally note that if $\boldsymbol{\mu}^* = \hat{\boldsymbol{\mu}}$, then, by Haberman's (1975) Theorem 2, the bound in (12) is equal to zero. Since this occurs irrespective of any additional condition, the present theorem may be viewed as an improvement (in the sense of more detailed information) of all other criteria for the equality of the considered two estimators, whereas such a statement could not be referred to the result of Haberman, as has been indicated by the example of Section 1.

Acknowledgment. The authors thank the referee for his comment which resulted in a more clear and efficient proof of the theorem.

REFERENCES

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

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

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