Biometrika (1975), **62**, 1, *p.* 121
Printed in Great Britain

(10) The inefficiency of least squares

BY PETER BLOOMFIELD AND GEOFFREY S. WATSON
Department of Statistics, Princeton University, New Jersey

SUMMARY

Two criteria are set up to judge the relative performance of the least squares estimator and the best linear unbiased estimator of β in the linear model $y = X\beta + u$, where $E(u) = 0$, $E(uu') = \Gamma$. The matrices X and Γ are found so that the relative performance of least squares is worst. Both criteria give the same least favourable situation: when $\mathcal{M}(X)$ is any one of the 2^k manifolds $\mathcal{M}(\gamma_1 \pm \gamma_n, \ldots, \gamma_k \pm \gamma_{n-k+1})$, where $\Gamma\gamma_i = f_i\gamma_i$ and $f_1 \leqslant \ldots \leqslant f_n$ are fixed, $\mathcal{M}(.)$ denoting the subspace spanned by the columns of the relevant matrix. The case where all f_i may be chosen in a preassigned interval is also discussed. The practical implications of the various results are mentioned.

Some key words: Efficiency; Least squares; Regression.

1. INTRODUCTION

This paper is concerned with the relative performances of the least squares estimator, b, and best linear unbiased estimator, $\hat{\beta}$, of β in the linear model

$$y = X\beta + u, \tag{1·1}$$

where y, which is $n \times 1$ and X which is $n \times k$ are observed and the $n \times 1$ error vector u has $E(u) = 0$, $E(uu') = \text{var}(u) = \Gamma$. It will, except in § 5, be assumed that rank $(X) = k$ and rank $(\Gamma) = n$. For our purposes we assume $n \geqslant 2k$. Then

$$b = (X'X)^{-1}X'y, \quad \hat{\beta} = (X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}y \tag{1·2}$$

$$\text{var}(b) = (X'X)^{-1}X'\Gamma X(X'X)^{-1}, \quad \text{var}(\hat{\beta}) = (X'\Gamma^{-1}X)^{-1}, \tag{1·3}$$

ellu

and

$$\text{var}(b) - \text{var}(\hat{\beta}) \quad \text{is nonnegative-definite.} \tag{1·4}$$

There is no loss of generality in supposing that

$$X'X = I_k. \tag{1·5}$$

The circumstances in which $b = \hat{\beta}$ are well known; the columns of X must span the same subspace as the columns of ΓX, that is $\mathcal{M}(X) = \Gamma\mathcal{M}(X)$. For the history, and for results in the arbitrary rank case, see Watson (1967, 1972). Here b is as 'good' as the 'best' $\hat{\beta}$. How 'bad' can b be relative to $\hat{\beta}$? The earliest reference to this problem is Tukey (1948). One needs a measure of relative performance; two are suggested below. For any given X and Γ, the measure will then be computed. In practice, X will be known but rarely Γ. Thus it would be of interest to have an attainable lower bound to the measure for X fixed and all Γ in some class. For mathematical reasons, it is also of importance to find lower bounds for Γ fixed and all X subject to (1·5). Finally, various combinations could be considered.

The basic criterion is (1·4) but there is no unique way to order nonnegative matrices. Thus all criteria are somewhat arbitrary. If generalized variances are used to define the 'efficiency' $\mathscr{E}$ of b relative to $\hat{\beta}$, then

$$\mathscr{E} = \frac{|\mathrm{var}(\hat{\beta})|}{|\mathrm{var}(b)|} = \{|X'\Gamma X|\,|X'\Gamma^{-1}X|\}^{-1}, \tag{1·6}$$

on using (1·3) and (1·5). Clearly $\mathscr{E} \leqslant 1$. We will give the lower bound under various circumstances. Our other criterion arose from trying to express the feeling that b will be bad when $\mathscr{M}(X)$ and $\Gamma\mathscr{M}(X)$ are dissimilar. When they are identical $\Gamma X = XA$, A nonsingular, so that $X'\Gamma X = X'XA = A$ and A is symmetric. Hence $\Gamma XX'$ is symmetric, so that the commutator $C = XX'\Gamma - \Gamma XX'$ is null. Thus we will expect b to be bad when $K = \mathrm{tr}(C'C)$ is large.

We will find that the circumstances that make $\mathscr{E}$ a minimum make K a maximum. While the two criteria are apparently different, there is a connexion, since determinantal manipulations yield the identity $|X'\Gamma X|\,|X'\Gamma^{-1}X| = |\Gamma + C|/|\Gamma|$.

The spectral decomposition of Γ will be defined by

$$\Gamma = \sum_{i=1}^{n} f_i \gamma_i \gamma_i' = GFG', \quad \gamma_i'\gamma_j = \delta_{ij}, \tag{1·7}$$

$$G = [\gamma_1, \ldots, \gamma_n], \quad F = \mathrm{diag}(f_1, \ldots, f_n), \tag{1·8}$$

with $0 < f_1 \leqslant \ldots \leqslant f_n$. We may consider maximizing $\mathscr{E}^{-1}$ and K under all combinations of

X fixed, X variable, subject to $X'X = I_k$, (1·9 a, b)

G fixed, G variable, subject to $G'G = I_n$, (1·10 a, b)

F fixed, F variable, subject to $f_* < f_1, \ldots, f_n < f^*$. (1·11 a, b)

It will turn out that the essential problems to solve come from the two combinations:

(1·9 b), (1·10 a), (1·11 a), that is X variable, Γ fixed; (1·12)

(1·9 a), (1·10 a), (1·11 b), that is X fixed, F variable but with fixed eigenvectors. (1·13)

In §§ 2 and 4 we solve (1·12) and §§ 3 and 5 we solve (1·13) for $\mathscr{E}^{-1}$ and K, respectively. The reductions of the other problems to these are mentioned as they arise.

In § 6 we discuss the results, especially as they relate to applications. We also consider the effect of relaxing the full rank assumptions on X and Γ.

In §§ 2 and 4, the method of Lagrangian multipliers is used, and it is convenient to show here that it will lead us to the maximum of $\mathscr{E}^{-1}$ and K on the X manifold specified by $X'X = I_k$. Since $\mathscr{E}^{-1}$ and K are polynomials in the elements of X, they are continuously differentiable, so that it is only necessary to show that the manifold is smooth, i.e. that the gradients of the $\frac{1}{2}k(k+1)$ functions

$$f_{ij}(X) = \sum_{l=1}^{n} x_{li} x_{lj} \quad (i \leqslant j,\ i, j = 1, \ldots, k)$$

are linearly independent. For $i \leqslant j$,

$$\frac{\partial f_{ij}(X)}{\partial x_{lm}} = \delta_{im} x_{lj} + \delta_{jm} x_{li},$$

so that linear dependence means there is a nonnull triangular matrix $A = ((a_{ij}))$, $a_{ij} = 0$ $(i > j)$, such that, for all l and m,

$$\sum_{i,j} a_{ij}(\delta_{im} x_{lj} + \delta_{jm} x_{li}) = 0,$$

that is $AX' + A'X' = 0$. But $(A+A')X'$ being null, when $X'X = I_k$, implies that $A+A'$ is null. Since A is triangular, A is null, which is a contradiction. Hence we shall be able to find the extremes among the stationary values.

2. The efficiency lower bound for variable X

Since $\mathscr{E}^{-1} = |X'\Gamma X|\,|X'\Gamma^{-1}X|$ and $\Gamma = GFG'$ we set

$$Y = G'X, \quad \mathscr{E}^{-1} = |Y'FY|\,|Y'F^{-1}Y|.$$

Because $Y'Y = X'GG'X = I_k$, $X = GY$, there is no difference between the problems (i) X variable, G fixed, F fixed, and (ii) X fixed, G variable, F fixed; always $X'X = I_k$. Thus we speak in terms of the first, i.e. (1·12).

Suppose that X maximizes $\mathscr{E}^{-1}$. Then XH, where H is $k \times k$ orthogonal, also achieves the same maximum. Thus we may seek a simple form for the maximizing X and obtain all others by an orthogonal transformation.

We maximize $\log \mathscr{E}^{-1}$ by the Lagrangian method using matrix differentiation (Rao, 1973, pp. 72, 73). The multipliers for the $\frac{1}{2}k(k+1)$ conditions can be written as an upper triangular matrix Λ. The stationary points are found by equating to the zero matrix the derivative, with respect to X, of $\log|X'\Gamma X| + \log|X'\Gamma^{-1}X| - 2\,\mathrm{tr}\,(X'X\Lambda)$. Thus

$$\Gamma X(X'\Gamma X)^{-1} + \Gamma^{-1}X(X'\Gamma^{-1}X)^{-1} - X(\Lambda+\Lambda') = 0. \tag{2·1}$$

Premultiplying (2·1) by X', we find that $2I_k = X'X(\Lambda+\Lambda') = \Lambda+\Lambda'$, so that (2·1) becomes

$$\Gamma X(X'\Gamma X)^{-1} + \Gamma^{-1}X(X'\Gamma^{-1}X)^{-1} = 2X. \tag{2·2}$$

If now (2·2) is premultiplied by $X'\Gamma$ we have

$$X'\Gamma^2 X(X'\Gamma X)^{-1} = 2X'\Gamma X - (X'\Gamma^{-1}X)^{-1}. \tag{2·3}$$

The right-hand side of (2·3) is symmetric, so that transposing (2·3) proves that $X'\Gamma^2 X$ and $(X'\Gamma X)^{-1}$ commute. Thus $X'\Gamma^2 X$ and $X'\Gamma X$ commute and so may be diagonalized by the same $k \times k$ orthogonal matrix. Application of this transformation to (2·3) shows that $X'\Gamma^{-1}X$ is also diagonalized. By the second paragraph of this section we may assume then that, for the X matrices making $\mathscr{E}$ stationary, $X'\Gamma X$ and $X'\Gamma^{-1}X$ are diagonal. Writing now $X = [x_1, \ldots, x_k]$, we have

$$\mathscr{E}^{-1} = \prod_{i=1}^{k} x_i'\Gamma x_i x_i'\Gamma^{-1} x_i. \tag{2·4}$$

Equation (2·2) may now be written $(i = 1, \ldots, k)$

$$\frac{\Gamma x_i}{x_i'\Gamma x_i} + \frac{\Gamma^{-1}x_i}{x_i'\Gamma^{-1}x_i} = 2x_i. \tag{2·5}$$

Premultiplication of (2·5) by Γ shows that the linear manifold spanned by x_i and Γx_i is closed with respect to Γ and hence is spanned by at most two eigenvectors of Γ. Furthermore, the corresponding eigenvalues are the roots a_i and b_i of the quadratic

$$\frac{f^2}{x_i'\Gamma x_i}+\frac{1}{x_i'\Gamma^{-1}x_i} = 2f, \tag{2·6}$$

so that

$$a_i+b_i = 2x_i'\Gamma x_i, \quad a_i b_i = \frac{x_i'\Gamma x_i}{x_i'\Gamma^{-1}x_i}. \tag{2·7}$$

Hence the contribution of x_i to the product (2·4) is

$$x_i'\Gamma x_i x_i'\Gamma^{-1}x_i = \frac{(a_i+b_i)^2}{4a_i b_i}.$$

If x_i and Γx_i are parallel, x_i is an eigenvector, (2·6) has a repeated root equal to the eigenvalue and the contribution is unity.

The k manifolds associated with $(x_1, \Gamma x_1), \ldots, (x_k, \Gamma x_k)$ are mutually orthogonal because $X'X$ and $X'\Gamma X$ are diagonal. All such sets of manifolds may be found by making k selections without replacement of one or two eigenvectors at a time. If $n > 2k$, it is clearly advantageous and possible to take pairs each time.

Then the remaining problem is to choose two disjoint subsets $(a_1, \ldots, a_k)$, $(b_1, \ldots, b_k)$ from $(f_1, \ldots, f_n)$, where $0 < f_1 \leqslant \ldots \leqslant f_n$, to maximize

$$\prod_{l=1}^{k} \frac{(a_l+b_l)^2}{a_l b_l}.$$

This may be solved by a nontrivial but elementary combinatorial argument. Thus we find, for all X subject to $X'X = I_k$,

$$\mathscr{E} \geqslant \prod_{l=1}^{k} \frac{4f_l f_{n-l+1}}{(f_l+f_{n-l+1})^2}. \tag{2·8}$$

This lower bound was conjectured by J. Durbin and given a proof incorrect for $k > 1$ by Watson (1955). Further discussion and some unattainable lower bounds for this case were given by Watson (1967). Hannan (1970, p. 422) showed that when estimating a linear combination $c'\beta$, the inequality

$$\frac{\text{var}\,(c'\hat{\beta})}{\text{var}\,(c'b)} \geqslant \frac{4f_1 f_n}{(f_n+f_1)^2}$$

holds. This bound also arises as the lower bound to the efficiency when $k = 1$. It is easily generalized to $k' < k$ linear combinations.

We see (2·8) also holds for fixed X, $X'X = I_k$ and all G. It holds in fact for all X of rank k, all G provided that F is fixed and $f_1 \leqslant \ldots \leqslant f_n$. The relation of these last conditions is considered in § 6.

3. The efficiency lower bound for variable bounded eigenvalues

We now search for the minimum of $\mathscr{E}$, when X and G are fixed but the eigenvalues f_i are varied in a closed interval $[f_*, f^*]$. Without loss of generality, $G = I_n$. We may also drop here the convention that $f_1 \leqslant \ldots \leqslant f_n$; in what follows f_1 is any eigenvalue of $\Gamma = F$.

If X is partitioned by its first row and the remainder so that $X' = [r_1, R_2]$ and F is correspondingly partitioned as diag (f_1, F_2), then

$$|X'FX| = |R_2 F_2 R_2'| \{1 + f_1 r_1'(R_2 F_2 R_2')^{-1} r_1\},$$

provided that $R_2 F_2 R_2'$ is nonsingular. But this is a consequence of the nonsingularity of $X'FX$. A similar result holds for $|X'F^{-1}X|$, so that

$$|X'F^{-1}X| = |R_2 F_2^{-1} R_2'| \{1 + f_1^{-1} r_1'(R_2 F_2^{-1} R_2')^{-1} r_1\}.$$

Thus

$$\mathscr{E}^{-1} = a(1 + bf_1)(1 + cf_1^{-1}), \tag{3·1}$$

where a, b and c are positive quantities not depending on f_1. Thus (3·1) shows that $\mathscr{E}^{-1}$ is maximized for fixed $f_2, \ldots, f_n$ by setting $f_1 = f_*$ or f^*. Hence all eigenvalues will be at the ends of the interval $[f_*, f^*]$.

If all the f_i are set equal to f_* or f^*, then $\mathscr{E}^{-1}$ can be written

$$|f_* A + f^* B| \, |f_*^{-1} A + f^{*-1} B|, \tag{3·2}$$

where $X'X = \Sigma r_i r_i' = A + B = I_k$. Thus A and B can be simultaneously diagonalized; let their eigenvalues be denoted by α_j and β_j, respectively ($j = 1, \ldots, k$), where $\alpha_j + \beta_j = 1$, with α_j and $\beta_j \geqslant 0$. Thus (3·2) reduces to

$$\prod_{j=1}^{k} (f_* \alpha_j + f^* \beta_j)(f_*^{-1} \alpha_j + f^{*-1} \beta_j) \tag{3·3}$$

which is bounded above by $\{\tfrac{1}{4}(f_* + f^*)^2/(f_* f^*)\}^k$. This bound will only be attainable if it is possible to partition the rows of X into two sets so that $\alpha_j = \beta_j = \tfrac{1}{2}$ ($j = 1, \ldots, k$). This means that $A = B = \tfrac{1}{2} I_k$. In general choice of rows so that this happens is impossible. The actual lower bound to $\mathscr{E}$ is the reciprocal of the largest attainable value of (3·3) over all partitions of the rows of X.

If X is also not fixed, it could be chosen so that $\alpha_j = \beta_j = \tfrac{1}{2}$ ($j = 1, \ldots, k$). In this case $\mathscr{E}$ achieves its lower bound, the reciprocal of the upper bound of (3·3). For reference we state

$$\inf \mathscr{E} \geqslant \{4 f_* f^*/(f_* + f^*)^2\}^k, \tag{3·4}$$

where the infimum is over $f_* \leqslant f_i \leqslant f^*$ ($1 = 1, \ldots, n$).

Comparing (3·4) and (2·8), we see that the bound (3·4) is the minimum of the (2·8) bound when all the f_i lie in (f_*, f^*). The result for $k = 1$ may be deduced from an inequality derived by Tukey (1948).

4. THE LARGEST COMMUTATOR FOR VARIABLE X

We now wish to maximize $\operatorname{tr}(CC')$, where $C = XX'\Gamma - \Gamma XX'$, over all X such that $X'X = I_k$, again assuming that $f_1 \leqslant \ldots \leqslant f_n$, with the f_i's fixed. As in § 2, this is the same as X fixed, $X'X = I_k$, F fixed and all G. If X maximizes $\operatorname{tr}(CC')$ so does XH, where H is $k \times k$ orthogonal, since C is unchanged. The Lagrangian multipliers for the $\frac{1}{2}k(k+1)$ restrictions will be summarized by an upper triangular matrix, Λ. Since $C' = -C$ and

$$\operatorname{tr}(C^2) = 2\operatorname{tr}\{(X'\Gamma X)^2\} - 2\operatorname{tr}(X'\Gamma^2 X)$$

we differentiate

$$\operatorname{tr}(X'\Gamma^2 X) - \operatorname{tr}\{(X'\Gamma X)^2\} - \operatorname{tr}(X'X\Lambda). \tag{4·1}$$

The resulting matrix equation is

$$\Gamma^2 X - 2\Gamma XX'\Gamma X = XS, \tag{4·2}$$

where S is the symmetric matrix $S = \frac{1}{2}(\Lambda + \Lambda')$.

Premultiplication of (4·2) first by X' and then by $X'\Gamma^{-1}$ leads to the two equations

$$S = X'\Gamma^2 X - 2(X'\Gamma X)^2, \quad -X'\Gamma X = X'\Gamma^{-1}XS. \tag{4·3}$$

Since S is symmetric, (4·3) shows that $X'\Gamma^{-1}X$ and S commute and so may be diagonalized by the same $k \times k$ orthogonal matrix H. By virtue of a remark above, we may assume the diagonality to obtain a solution and generate all others by orthogonal transformation.

On writing $X = [x_1, \ldots, x_k]$, (4·2) premultiplied by Γ^{-1} becomes, for $i = 1, \ldots, k$,

$$\Gamma x_i - S_{ii}\Gamma^{-1}x_i = 2x_i'\Gamma x_i x_i. \tag{4·4}$$

Premultiplication of (4·4) by x_i' verifies that $S_{ii} = -(x_i'\Gamma x_i)/(x_i'\Gamma^{-1}x_i)$, so that (4·4) is identical to (2·5) whose solutions we know. To obtain

$$\operatorname{tr}(CC') = 2\Sigma x_i'\Gamma^2 x_i - 2\Sigma(x_i'\Gamma x_i)^2$$

we have (2·7) and the deduction from (2·5),

$$\frac{x_i'\Gamma^2 x_i}{x_i'\Gamma x_i} + \frac{1}{x_i'\Gamma^{-1}x_i} = 2x_i'\Gamma x_i.$$

Hence $x_i'\Gamma^2 x_i - (x_i'\Gamma x_i)^2 = \frac{1}{4}(a_i - b_i)^2$.

Thus to maximize $\operatorname{tr}(C'C)$ we need to maximize $\Sigma(a_i - b_i)^2$, where the sum is over k distinct pairs of eigenvalues. This is a problem similar to, but simpler than, the corresponding one in § 2, so the details will be omitted.

Thus, we have shown that, for fixed Γ,

$$\sup_{X'X = I_k} \operatorname{tr}(C'C) = \tfrac{1}{4}\sum_{l=1}^{k}(f_l - f_{n-l+1})^2. \tag{4·5}$$

In fact, as with (2·8), the supremum is really over all X of rank k and with fixed F. In (2·8) we are interested in the relationship of each term in the product and unity. But

$$\frac{(f_l + f_{n-l+1})^2}{4f_l f_{n-l+1}} - 1 = \frac{(f_l - f_{n-l+1})^2}{4f_l f_{n+l+1}}.$$

Then (4·5) is a sum of these differences weighted by $f_l f_{n-l+1}$.

5. The largest commutator for variable eigenvalues

To maximize $K = \operatorname{tr}(C'C)$ for fixed X and G by choosing the eigenvalues f_i, no longer ordered, in the closed interval $[f_*, f^*]$, we may suppose $G = I_n$. Thus we maximize

$$K = \operatorname{tr}\{(X'F^2X)\} - \operatorname{tr}\{(X'FX)^2\}, \tag{5·1}$$

which, if we write $X' = [r_1, \ldots, r_n]$, becomes

$$K = \Sigma_i f_i^2 r_i' r_i - \Sigma_i \Sigma_j f_i f_j (r_i' r_j)^2. \tag{5·2}$$

Then

$$\frac{\partial K}{\partial f_1} = 2f_1 r_1' r_1 (1 - r_1' r_1) - \sum_{j \neq 1} f_j (r_1' r_j)^2,$$

$$\frac{\partial^2 K}{\partial f_1^2} = 2r_1' r_1 (1 - r_1' r_1).$$

Since $X'X = I_k$, $0 \leqslant r_i' r_i \leqslant 1$, X could be completed to be an orthogonal matrix all of whose rows have unit length. Thus any stationary values correspond to minima so that K will be a maximum if all the f_i are either f_* or f^*.

For any allocation of f_* and f^* to the rows of X,

$$X'FX = f_* A + f^* B, \quad X'F^2X = f_*^2 A + f^{*2} B,$$

where $A + B = I_k$, as in (3·2). From (5·1), and after diagonalizing, $K = (f^* - f_*)^2 \Sigma \alpha_j (1 - \alpha_j)$. Thus

$$K \leqslant \tfrac{1}{4} k (f^* - f_*)^2, \tag{5·3}$$

this upper bound being attained if and only if every $\alpha_j = \frac{1}{2}$, that is $A = B = \frac{1}{2} I_k$. It would be attained if X were free. Conversely, comparing (5·3) and (4·5), it is clear that the bound in (5·3) is the maximum of the bound (4·5) when the eigenvalues may be freely placed on $[f_*, f^*]$.

6. Discussion

We first discuss the results obtained and then consider their meaning for applications.

In §§ 2 and 4, had some of the columns of X been eigenvectors, they would have been kept fixed while the remaining columns were varied. The results are the same with a decreased k and a decreased set of eigenvalues to choose from.

The restriction $n \geqslant 2k$ has been assumed because, in practice, n is much larger than k. If $k < n < 2k$ one is forced to take some simple eigenvectors; the result then has been given by Knott (1975).

The assumptions that X and Γ are of full rank are more serious. If either fail, the definition (1·6) of the efficiency fails. The commutator, however, still makes sense. For we know that, quite generally, the least squares estimator of the estimable part of β, i.e. that part lying in the row-space of X, $\mathscr{M}(X)$, is optimal if $\Gamma \mathscr{M}(X) \subset \mathscr{M}(X)$. Then if rank $(X) = k$, so that the assumption (1·5) that $X'X = I_k$ may be made, the justification and minimization of $C = XX'\Gamma - \Gamma XX'$ remain unchanged.

If rank $(X) = k' < k$, let P be the orthogonal projector onto $\mathscr{M}(X)$, and consider the commutator $C = P\Gamma - \Gamma P$. The optimality condition may be written $\Gamma P = PA$ so that $P'\Gamma' = P\Gamma = A'P$. Thus $P'\Gamma P = P'PA = PA$, while the second equation gives $P'\Gamma P = A'P$.

Thus $PA = A'P$ or $C = P\Gamma - \Gamma P = 0$. Thus minimizing the commutator $C = P\Gamma - \Gamma P$ still makes sense. If Z has k' orthonormal columns which span $\mathscr{M}(X)$, then $P = ZZ'$. Hence in the general case the commutator argument holds with $k' = k$. Thus if $0 \leqslant f_1 \leqslant \ldots \leqslant f_n$ and rank $(X) = k'$, the least favourable regression subspaces for least squares are

$$\mathscr{M}(\gamma_1 \pm \gamma_n, \ldots, \gamma_k \pm \gamma_{n-k+1}), \tag{6·1}$$

when Γ is fixed. If X is fixed and G is variable, (6·1) can be interpreted as the least favourable placement of the eigenvectors relative to X.

When the eigenvalues are allowed to vary in $[f_*, f^*]$, Γ is necessarily nonsingular and we have to worry only about the case rank $(X) = k' < k$. If we use the commutator $P\Gamma - \Gamma P$ as suggested above, the changes in § 5 are as follows. We now have $A + B = Z'Z$, the orthonormal projector onto $\mathscr{M}(X')$. Then A and B are Gramians with eigenvalues $\alpha_j \geqslant 0$ and $\beta_j \geqslant 0$ and $\alpha_j + \beta_j = 1$ or 0. Thus (5·3) holds with k replaced by k'. The upper bound will be attained if and only if k' of the $\alpha_j = \frac{1}{2} = \beta_j$, the remainder being zero. Then $A = B = \frac{1}{2}Z'Z$. The last remark of § 5 is true here too.

This leads us to a discussion of practical applications. The case just mentioned would happen if an experiment were done twice, with the same regressors but error covariance matrices $f^*\Sigma$ and $f_*\Sigma$, respectively, where Σ is a matrix known up to a scalar.

The most common practical cases are (i) uncorrelated errors with variances f_i and (ii) stationary errors for which the f_i are to be interpreted as equally spaced values of 2π times the spectral density function and $[\gamma_1, \ldots, \gamma_n]$ as its discrete Fourier transform.

In case (i) there are two possibilities. If the f_i are equal in blocks, then situations similar to the replication example just described can arise. If the f_i are distinct, the worst case occurs only when basis vectors for $\mathscr{M}(X)$ are vectors with elements all zero except for two equal entries corresponding to the errors with least and greatest variances, second least and second greatest variances, etc. Such matrices of zeros and ones only occur in design models so our lower bound is unduly pessimistic. Nevertheless, least squares can be very inefficient if f_n/f_1 is very large and it is wise to use a robust regression method.

In case (ii) the worst case occurs when the regressors have power concentrated equally at the frequencies where the density is highest and lowest, etc.

It is often the case that the regressors are largely of low frequency so that it is only possible to get good estimates of the power in the error process at higher frequencies. Thus we will never know whether there are very high and very low values of the spectral density at the low frequencies which will hurt least squares most here.

This work was partially supported by the Office of Naval Research.

REFERENCES

HANNAN, E. J. (1970). *Multiple Time Series*. New York: Wiley.
KNOTT, M. (1975). On the minimum efficiency of least squares. *Biometrika* **62**, 129–32.
RAO, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd edition. New York: Wiley.
TUKEY, J. W. (1948). Approximate weights. *Ann. Math. Statist.* **19**, 91–2.
WATSON, G. S. (1955). Serial correlation in regression analysis. *Biometrika* **42**, 327–41.
WATSON, G. S. (1967). Linear least squares regression. *Ann. Math. Statist.* **38**, 1679–99.
WATSON, G. S. (1972). Prediction and the efficiency of least squares. *Biometrika* **59**, 91–8.

[Received *January* 1974. *Revised September* 1974]