T. R. NO. 62

Technical Report

SOME RESULTS IN FINDING A LOWER BOUND ON THE EFFICIENCY OF LEAST SQUARE ESTIMATES RELATIVE TO BEST LINEAR ESTIMATES IN REGRESSION MODEL

M. Raghavachari

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD

SOME RESULTS IN FINDING A LOWER BOUND ON THE EFFICIENCY OF LEAST SQUARE ESTIMATES RELATIVE TO BEST LINEAR ESTIMATES IN REGRESSION MODEL

M. Raghavachari

T R No 62

Dec., 1974

Indian Institute of Management
Ahmedabad

No. 62

To

Chairman (Research)
IIMA

Technical Report

Title of the report: Some results in finding a lower bound on the efficiency of Least Square estimates relative to best linear estimates in Regression model

Name of the Author: M. Raghavachari

Under which area do you like to be classified? P&QM Area

ABSTRACT (within 250 words)

Consider the usual regression model

$$Y = X\beta + U$$

The standard estimators of β are (i) Least squares estimator and (ii) Best linear estimator. ~~There has been work~~ The paper gives some results on finding an attainable lower bound on the efficiency of least square estimates relative to the best linear estimate. Specifically the paper is an attempt to verify the validity of a conjecture made by G.S. Watson.

Please indicate restrictions if any that the author wishes to place upon this note: none

Date 18.12.74

Signature of the Author

SOME RESULTS IN FINDING A LOWER BOUND ON THE EFFICIENCY OF LEAST SQUARE ESTIMATES RELATIVE TO BEST LINEAR ESTIMATES IN REGRESSION MODEL

by

M.Raghavachari
Indian Institute of Management, Ahmedabad

1. Introduction:

Consider the usual regression model in matrix notation:

$$Y = X\beta + U$$

where Y is nx1, X is nxk, β is kx1 and U is nx1. The X matrix will be regarded as fixed; the error vector U is random with E(U) = 0 and E(UU') = Γ, say. Assume that the ranks of X and Γ are k and n respectively. It is well known that the best linear estimator is given by

$$\hat{\beta} = (X' \Gamma^{-1} X)^{-1} (X' \Gamma^{-1} Y)$$

with

$$\text{Var}(\hat{\beta}) = (X' \Gamma^{-1} X)^{-1}.$$

The least squares estimator b is given by

$$b = (X'X)^{-1} X'Y$$

with

$$\text{Var}(b) = (X'X)^{-1} (X'\Gamma X) (X'X)^{-1}$$

A measure of efficiency of the least squares estimator relative to the best linear estimator is given by the ratio of generalized variances:

$$\text{(1)} \quad \text{eff}(b) = |\text{Var}(\hat{\beta})| \Big/ |\text{Var}(b)| = |X'X|^2 \Big/ |X'\Gamma X|\,|X'\Gamma^{-1}X|.$$

It can be shown, see e.g. G.S. Watson (1967) that $0 \leq \text{eff}(b) \leq 1$ and that the upper bound can be obtained for a particular choice of X. It is of interest to determine an attainable lower bound for eff (b). For

---

Presented at the VIII International Symposium on Mathematical Programming held in Stanford University, Stanford, USA in August 1973.

$k > 1$ and $n > 2k-1$ G.S. Watson (1955 and 1967) suggested **the inequality**

(2) $$\mathrm{eff}(b) \geq \left[4\lambda_1\lambda_n/(\lambda_1+\lambda_n)^2\right]\left[4\lambda_2\lambda_{n-1}/(\lambda_2+\lambda_{n-1})^2\right]\cdots\left[4\lambda_k\lambda_{n-k+1}/(\lambda_k+\lambda\right.$$

where $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigen values of Γ. For $k=1$, (2) has shown to be true. See G.S. Watson (1967). The purpose of this paper is to study the problem for $k > 1$. Some insights into the nature of the problem for general $k \geq 2$ are given. These results may be useful in settling the validity of (2). Other lower bounds for eff(b) have also been proposed by G. Golub (1963) and G.S. Watson (1967).

2. As shown in Section 1, the problem is to

(3) $$\text{minimize } |X'X|^2 / |X'\Gamma X| \,.\, |X'\Gamma^{-1}X|$$

G.S.Watson (1967) has shown that there is no loss of generality in assuming $X'X = I_k$ in (3). Here I_k denotes the identity matrix of order k. The problem (3) is then equivalent to the constrained maximization problem:

Find a X such that $|X'\Gamma X|\;|X'\Gamma^{-1}X|$ is maximum subject to $X'X = I_k$. We can further assume without loss of generality that Γ is a diagonal matrix with elements $\lambda_1, \lambda_2, \cdots \lambda_n$ where the λ's are the eigen values of Γ. This is because there exists an orthogonal matrix C such that $C'\Gamma C = \mathrm{diag}(\lambda_1, \lambda_2, \cdots \lambda_n)$ and $C'\Gamma^{-1}C = \mathrm{diag}(\lambda_1^{-1}, \cdots \lambda_n^{-1})$. Thus the problem is equivalent to:

(4) $$\text{maximize } |X'\Gamma X| \,.\, |X'\Gamma^{-1}X|$$
$$\text{subject to } X'X = I_k$$

with $\Gamma = \mathrm{diag}(\lambda_1, \cdots \lambda_n)$. In what follows we will be concerned with this problem and Γ will be $\mathrm{diag}(\lambda_1, \cdots \lambda_n)$. X will be a nxk matrix with

$$X' = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdot\cdot & \cdot\cdot & \cdots & \cdot\cdot \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{pmatrix}$$

For k=1, the problem reduces to

$$\text{maximize } (\lambda_1 x_{11}^2 + \lambda_2 x_{12}^2 + \dots + \lambda_n x_{1n}^2) \cdot \left(\lambda_1^{-1} x_{11}^2 + \lambda_2^{-1} x_{12}^2 + \dots + \lambda_n^{-1} x_{1n}^2\right)$$

$$\text{subject to } x_{11}^2 + x_{12}^2 + \dots + x_{1n}^2 = 1 .$$

An application of Kantorivich's inequality solves the above problem Several proofs of this inequality have been given in literature. We give below another proof based on theory of Mathematical programming.

Proof for the case k=1: By setting $x_{1j}^2 = y_j$, the problem is to

$$\text{maximize}\left(\lambda_1 y_1 + \lambda_2 y_2 + \dots + \lambda_n y_n\right)\left(\lambda_1^{-1} y_1 + \dots + \lambda_n^{-1} y_n\right)$$

(5) s.t.
$$y_1 + y_2 + \dots + y_n = 1$$
$$y_j \geqslant 0, \quad j = 1,2,\dots n$$

Let $y_1^o, \dots, y_n^o$ be an optimal solution of (5) with $\lambda_1^{-1} y_1^o + \lambda_2^{-1} y_2^o + \dots + \lambda_n^{-1} y_n^o = \delta_o$, Then $y_1^o, \dots y_n^o$ is an optimal solution of the following problem

$$\text{maximize } \lambda_1 y_1 + \lambda_2 y_2 + \dots + \lambda_n y_n$$

(6) s.t.
$$y_1 + y_2 + \dots + y_n = 1.$$
$$\lambda_1^{-1} y_1 + \lambda_2^{-1} y_2 + \dots + \lambda_n^{-1} y_n = \delta_o$$
$$y_j \geqslant 0 \quad j = 1, \dots n.$$

The problem (6) is a linear programming problem with two constraint equations. It is well known from the theory of linear programming that there exists an optimal solution of (6) with utmost two of y_j's positive and the remaining y_j's zero. Let y_i and y_j be the two positive values satisfying

$$y_i + y_j = 1$$
$$\lambda_i^{-1} y_i + \lambda_j^{-1} y_j = \delta_o$$

When $\lambda_i = \lambda_j$, the value of the objective function of (6) equals 1 ; if $\lambda_i \neq \lambda_j$ solving equations (+) we can verify that

$$\lambda_i y_i + \lambda_j y_j = \lambda_i + \lambda_j - \delta_0 \lambda_i \lambda_j$$

so that $\delta_0 (\lambda_i + \lambda_j - \delta_0 \lambda_i \lambda_j)$ will be maximum when $\delta_0 = \frac{1}{2}\left(\frac{1}{\lambda_i} + \frac{1}{\lambda_j}\right)$. This gives $y_i = y_j = \frac{1}{2}$

and the value of the objective function of (6) equals

$$\frac{(\lambda_i + \lambda_j)^2}{4\lambda_i\lambda_j} \quad \text{which is} \geqslant 1. \quad \text{Further}$$

it can be checked that $\max\limits_{1 \le i \le n} \frac{(\lambda_i + \lambda_j)^2}{4\lambda_i \lambda_j} = \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}$

which occurs when $y_1 = y_n = \frac{1}{2}$. This proves completely the case $k = 1$.

General case: $k > 1$. The mathematical problem was given by (4). This can further be reduced to another equivalent problem. This is given in the following lemma.

Lemma: The problem (4) is equivalent to

$$\text{maximize } |X' \Gamma X| \, |X' \Gamma^{-1} X|$$

(7) s. t. $X_i' X_i = 1 \quad i = 1, 2, \ldots k$

where $X_i' = (x_{i1}, x_{i2}, \ldots x_{in})$.

Proof: We have to show that any optimal solution of (7) satisfies $X_i' X_j = 0$, for $i \neq j$ Consider the lagrangian function

$$Q = |X' \Gamma X| \, |X' \Gamma^{-1} X| - \sum_{i=1}^{k} t_i (X_i' X_i - 1).$$

Any optimal solution X· must satisfy $\frac{\partial Q}{\partial X_i'} = 0$ for i = 1, ... k

where as usual $\frac{\partial Q}{\partial X_i'}$ denotes the vector $\left(\frac{\partial Q}{\partial x_{i1}}, \cdots \frac{\partial Q}{\partial x_{in}} \right)$.

We have

$$\frac{\partial \left\{ |X'\Gamma X| \, |X' \Gamma^{-1} X| \right\}}{\partial X_i'} = |X' \Gamma^{-1} X| \frac{\partial |X'\Gamma X|}{\partial X_i'} + |X'\Gamma X| \frac{\partial}{\partial X_i'} |X' \Gamma^{-1} X|$$

Note that

$$(8) \qquad |X'\Gamma X| = \begin{vmatrix} X_1'\Gamma X_1 & X_1'\Gamma X_2 & \cdots\cdots & X_1'\Gamma X_k \\ X_2'\Gamma X_1 & X_2'\Gamma X_2 & \cdots\cdots & X_2'\Gamma X_k \\ - & - & & - \\ - & - & & - \\ X_k'\Gamma X_1 & X_k'\Gamma X_2 & \cdots\cdots & X_k\Gamma X_k \end{vmatrix}$$

It can be shown that

$$(9) \qquad X_i' \frac{\partial |X'\Gamma X|}{\partial X_i'} = \begin{vmatrix} X_1'\Gamma X_1 & X_1'\Gamma X_2 & \cdots\cdots & X_1'\Gamma X_k \\ X_2'\Gamma X_1 & X_2'\Gamma X_2 & \cdots\cdots & X_2'\Gamma X_k \\ 2X_i'\Gamma X_1 & 2X_i'\Gamma X_2 & \cdots\cdots & 2X_i'\Gamma X_k \\ X_k'\Gamma X_1 & X_k'\Gamma X_2 & \cdots\cdots & X_k'\Gamma X_k \end{vmatrix} = 2\,|X'\Gamma X|$$

Also

$$X_j' \frac{\partial |X'\Gamma X|}{\partial X_i'} = \pm \begin{vmatrix} X_1'\Gamma X_1 & X_1'\Gamma X_2 & \cdots & X_1'\Gamma X_i & X_1'\Gamma X_k \\ - & - & - & - & - \\ X_j'\Gamma X_1 & X_j'\Gamma X_2 & & 2X_j'\Gamma X_i & X_j'\Gamma X_k \\ X_j'\Gamma X_1 & X_j'\Gamma X_2 & & X_j'\Gamma X_i & X_j'\Gamma X_k \\ X_k'\Gamma X_1 & X_k'\Gamma X_2 & & X_k'\Gamma X_i & X_k'\Gamma X_k \end{vmatrix}$$

(10)

$$\pm (X_j'\Gamma X_i) \begin{vmatrix} X_1'\Gamma X_1 & X_1'\Gamma X_2 & \cdots & X_1'\Gamma X_{i-1} & X_1'\Gamma X_{i+1} \cdots X_1'\Gamma X_k \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ X_{j-1}'\Gamma X_1 & X_{j-1}'\Gamma X_2 & & X_{j-1}'\Gamma X_{i-1} & X_{j-1}'\Gamma X_{i+1} \quad X_{j-1}'\Gamma \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ X_j'\Gamma X_1 & X_j'\Gamma X_2 & & X_j'\Gamma X_{i-1} & X_j'\Gamma X_{i+1} \quad X_j'\Gamma \\ X_k'\Gamma X_1 & X_k'\Gamma X_2 & & X_k'\Gamma X_{i-1} & X_k'\Gamma X_{i+1} \quad X_k'\Gamma \end{vmatrix}$$

$$- \pm \sum_{l \neq i,j} (X_l'\Gamma X_i) \, |A_l|$$

where in A_l there are two identical columns, namely

$$X_j'\Gamma X_1, \quad \ldots \quad X_j'\Gamma X_{i-1} \quad X_j'\Gamma X_{i+1} \quad \cdots \quad X_j'\Gamma X_k)$$

$= 0$

Both these results (9) and (10) can be verified by expanding the determinant in (8) by the ith row, differentiating and collecting terms

We thus have

$$X_i' \frac{\partial |X'\Gamma X|}{\partial X_i'} = 2|X'\Gamma X|; \quad X_i' \frac{\partial |X'\Gamma^{-1}X|}{\partial X_i'} = 2|X'\Gamma^{-1}X| \quad i = 1, \cdots k$$

$$X_j' \frac{\partial |X'\Gamma X|}{\partial X_i'} = X_j' \frac{\partial |X'\Gamma^{-1}X|}{\partial X_i'} = 0 \quad i \neq j$$

Any optimal solution therefore satisfies

(11) $$X_i' \frac{\partial Q}{\partial X_i'} - t_i X_i' X_i = 0$$

(12) $$X_j' \frac{\partial Q}{\partial X_i'} - t_i X_j' X_i = 0$$

In virtue of (9) and (10), (11) implies $t_i = 2\,|X'\Gamma X|\,|X'\Gamma^{-1}X|$

and (12) implies $X_j' X_i = 0$ for $i \neq j$ which proves the lemma.

The following theorem gives the properties that an optimal solution (7) must satisfy

Let $\lambda_1, \lambda_2, \cdots \lambda_n$ be all distinct

Theorem: There exists an optimal Solution X to (7). Where n-2k columns are all zeros.

Proof: Let $a_{ij}(\Gamma)$ denote the (i, j) element of the matrix $X'\Gamma X$ i.e. $a_{ij}(\Gamma) = X_i'\Gamma X_j$. Similarly let $a_{ij}(\Gamma^{-1})$ denote the (i,j) element of the matrix $(X'\Gamma^{-1}X)$. i.e. $a_{ij}(\Gamma^{-1}) = X_i'\Gamma^{-1}X_j$. Let $A_{ij}(\Gamma)$ denote the cofactor of $a_{ij}(\Gamma)$ in $X'\Gamma X$ and $A_{ij}(\Gamma^{-1})$ denote the cofactor of $a_{ij}(\Gamma^{-1})$ in $X'\Gamma^{-1}X$. Consider the problem (7) and the Lagrangian equations are given by

$$\sum_{j=1}^{k} \left(\lambda_p A_{ij}(\Gamma)\,|X'\Gamma^{-1}X| + \lambda_p^{-1} A_{ij}(\Gamma^{-1})\,|X'\Gamma X| - t_i\,\delta_{ij} \right) x_{ij} = 0$$

$$i = 1,2, \ldots k$$

where δ_{ij} is the usual Kronecker delta. $\quad p = 1,2, \ldots n$

For a given p, the system of k homogeneous equations in k unknowns $x_{1p}, x_{2p}, \ldots x_{kp}$ has a non-trivial solution if and only if the determinant of the coefficient matrix vanishes. This condition is seen to be equivalent to

$$\left| \, |X'\Gamma^{-1}X| \lambda_p \; A_{ij}(\Gamma) + |X'\Gamma X| \lambda_p^{-1} \; A_{ij}(\Gamma^{-1}) - t_i \; I \right| = 0$$

Since $t_i = 2 \; |X'\Gamma X| \; |X' \Gamma^{-1} X|$ by (11), the condition reduces to

$$\left| \lambda_p \; (X'\Gamma X)^{-1} + \lambda_p^{-1} \; (X'\Gamma^{-1}X)^{-1} - 2 \; I \right| = 0 . \; (p = 1, 2, \ldots n)$$

For a given $(X'\Gamma X)^{-1}$ and $(X'\Gamma^{-1}X)^{-1}$, the equation

(14) $$\left| y \; (X'\Gamma X)^{-1} + y^{-1} \; (X'\Gamma^{-1}X)^{-1} - 2 \; I \right| = 0$$

is of degree 2 k in y and the relation holds for atmost (2k) values of λ. Since the λ's are assumed to be all distinct, it follows that $x_{1p}, \ldots x_{kp}$ must be all zero for atleast (n-2k) values of p. This proves the theorem.

The above theorem shows that we can consider X' as a matrix with k rows and 2k columns i.e. we can assume n=2k. Assume for definiteness that $(x_{1p}, \ldots x_{kp})$ are all zero for p = 2k+1 , ... n. Then we have the corollary.

Corollary: Let $\lambda_1, \ldots \lambda_{2k}$ be distinct and consider the problem:

maximize $|X'\Gamma X| \; |X' \Gamma^{-1} X|$

s.t. $\quad X_i' X_i = 1$

with $X_i' = (x_{i1}, \ldots x_{i\,2k})$ and $\Gamma = \text{diag} \; (\lambda_1, \ldots \lambda_{2k})$

For an optimal solution X of (7) we have

$$|X'\Gamma X| \,/\, |X'\Gamma^{-1}X| = \lambda_1 \lambda_2 \ldots \lambda_{2k}$$

$$\sum_{i=1}^{k} X_i'\Gamma X_i = \frac{\lambda_1+\lambda_2+\ldots+\lambda_{2k}}{2}$$

$$\sum_{i=1}^{i} X_i'\,\Gamma^{-1}X_i = \left(\lambda_1^{-1}+\lambda_2^{-1}+\ldots+\lambda_{2k}^{-1}\right)/2$$

Proof: Consider the equation of degree 2k in y given by (14). $\lambda_1, \lambda_2, \ldots \lambda_{2k}$ are the rbots of the equation and the corollary follows immediately by considering the product and sum of the roots.

It can be verified that the values of the x_{ij}'s which yield the Watson's bounds satisfy these properties. The author has not been able to settle the conjecture for the general k. However, the mathematical problem has been reduced to a simpler form as given in the corollary. The constraints $X_i' X_i = 1$, $i = 1,2, \ldots k$ can also be written as $X_i' X_i \leqslant 1$, $i = 1,2, \ldots k$ without changing the problem. The constraint set then becomes a convex set in x_{ij}s. It may be possible to exploit the convexity properties to settle the validity of Watson's bounds. In fact it is well-known that

$$\max_{\substack{X_i' X_i = 1 \\ i=1,2..k}} |X'\Gamma X| = \lambda_{k+1}\lambda_{k+2} \ldots \lambda_{2k}$$

(Note that $\lambda_1 < \lambda_2 < \cdots < \lambda_{2k}$) and the maximum value is attained for $x^2_{i,2k-i+1} = 1$ for $i = 1,2,\ldots k$ and other $x_{ij}^2 = 0$

Similarly

$$\max_{\substack{X_i' X_i = 1 \\ i=1,2,..k}} \left| X' \Gamma^{-1} X \right| = \lambda_1^{-1} \lambda_2^{-1} \ldots \lambda_k^{-1}$$

and is attained for $x_{ii}^2 = 1$ for $i = 1,2, .. k$ and other $x_{ij}^2 = 0$. The midpoint of the line segment joining these two solutions in x_{ij}^2 is given by $x_{ii}^2 = x_{i,2k-i+1}^2 = \frac{1}{2}$ for $i=1,2,..k$ and the other $x_{ij}^2 = 0$. This solution gives the Watson's bound for the objective function $\left| X' \Gamma X \right| \left| X' \Gamma^{-1} X \right|$.

Acknowledgment

The author is grateful to Paul Shaman, Carnegie-Mellon University for many useful discussions.

References

G.S.Watson (1955). Serial Correlation in Regression Analysis, I Biometrika, 42, 327-341.

G.H. Golub (1963). Comparison of the Variance of Minimum Variance and Weighted Least Squares Regression Coefficients. Annals of Mathematical Statistics. Vol 34 No 3, 984-991

G.S.Watson (1967). Linear Least Squares Regression. Ann. Math. Statist. 38, 1679-1699.