

1982

Properties of Shrinkage Estimators in Linear Regression when Disturbances Are not Normal

Aman Ullah

V.K. Srivastava

R. Chandar

Follow this and additional works at: <https://ir.lib.uwo.ca/economicsresrpt>

 Part of the [Economics Commons](#)

Citation of this paper:

Ullah, Aman, V. K. Srivastava, R. Chandar. "Properties of Shrinkage Estimators in Linear Regression when Disturbances Are not Normal." Department of Economics Research Reports, 8212. London, ON: Department of Economics, University of Western Ontario (1982).

20068

ISSN: 0318-725X

ISBN: 0-7714-0351-8

RESEARCH REPORT 8212

PROPERTIES OF SHRINKAGE ESTIMATORS IN
LINEAR REGRESSION WHEN DISTURBANCES
ARE NOT NORMAL

by

A. Ullah*, V. K. Srivastava and
R. Chandar

University of Western Ontario and C.O.R.E.
and
Lucknow University

Abstract

This paper considers a class of recently developed biased estimators of regression coefficients and studies its sampling properties when the disturbances are not normally distributed. It has been found that the conditions of dominance of these estimators over the last squares estimator, under non-normality, are quite different than their well-known dominance conditions under normality. Some implications of the results are also discussed.

May, 1982

*The research support from the NSERC is gratefully acknowledged. The authors are thankful to M. Mouchart for his useful comments. They are also grateful to Arnold Zellner, two referees and an associate editor for useful suggestions.

Department of Economics Library

MAY 21 1982

University of Western Ontario

1. INTRODUCTION

The least squares estimator for the coefficients in a linear regression model is well known for its unbiasedness and minimum variance. If we are prepared to sacrifice the unbiasedness property, the estimation of coefficients can be improved; see, e.g., James and Stein (1961); Judge and Bock (1978); Judge et al. (1980); Ullah et al. (1978) and Zellner and Vandaele (1975). Properties of such improved estimators have generally been analyzed under the assumption that disturbances are normally distributed. Such an assumption, it is well recognized, is often questionable and may have varying effects in a variety of situations; see, e.g., Gnanadesikan (1977). This has motivated us to analyze the properties of improved estimators when disturbances are not normal. Some work in this direction has been reported by Brandwein (1979) and Brandwein and Strawderman (1978, 1980) who considered a class of spherically symmetric distributions for disturbances and obtained conditions for minimaxity of an improved family. Their results again call for a specification of distribution of disturbances. Such is not the case with our set-up. We simply assume that disturbances are small and possess moments of fourth order. Under this general specification, it is found that the conditions for dominance of improved estimators over least squares estimator are quite different as compared with those for normal disturbances. The plan of the paper is as follows. In Section 2 we present the estimators and their properties under the non-normal disturbances. Section 3 provides the proofs of the results in Section 2.

2. ESTIMATORS AND THEIR PROPERTIES

Consider a linear regression model:

$$(2.1) \quad y = X\beta + u$$

where y is a $T \times 1$ vector of observations on the variable to be explained, X is a $T \times p$ matrix, with full column rank, of observations on p explanatory variables, β is a $p \times 1$ vector of regression coefficients and u is a $T \times 1$ vector of disturbances.

It is assumed that the elements of u are independently and identically distributed with first four finite moments as 0 , σ^2 , $\sigma^3 v_1$ and $\sigma^4(v_2 + 3)$ respectively. Thus for all t ($t = 1, 2, \dots, T$) we have

$$(2.2) \quad \begin{aligned} E(u_t) &= 0 & , & & E(u_t^2) &= \sigma^2 \\ E(u_t^3) &= \sigma^3 v_1 & , & & E(u_t^4) &= \sigma^4 (v_2 + 3) \end{aligned}$$

where v_1 and v_2 are Pearson's measures of skewness and kurtosis of the distribution of disturbances. Notice that $v_1 = 0$ is implied by symmetry of distribution while $v_2 = 0$ means that the distribution is mesokurtic.

The ordinary least squares (OLS) estimator of β in (2.1) is given by

$$(2.3) \quad b = (X'X)^{-1}X'y$$

which is unbiased with variance-covariance matrix $\sigma^2(X'X)^{-1}$.

Assuming Q to be a symmetric positive definite matrix, the risk of b is

$$(2.4) \quad E(b-\beta)'Q(b-\beta) = \sigma^2 \operatorname{tr}(X'X)^{-1}Q$$

where "tr" represents the trace of the matrix.

Now consider a general class of shrinkage estimators as

$$(2.5) \quad \hat{\beta} = [I + hD]^{-1}b$$

where D is any known $p \times p$ positive definite matrix and the stochastic scalar h is

$$(2.6) \quad h = \frac{ks}{b'Cb}; \quad s = (y - Xb)'(y - Xb).$$

In (2.6), k is a positive constant and C is a known positive definite matrix.

It may be observed that for $D = I$, $\hat{\beta}$ is a Stein-type estimator and for $D = (X'X)^{-1}$ it is a ridge-type adaptive estimator; for the details about these types of estimators, see Judge and Bock (1978), and Judge et al. (1980).

Before presenting the risk function of $\hat{\beta}$ we introduce the following notations:

$$(2.7) \quad \begin{aligned} M &= X(X'X)^{-1}X' ; \quad \bar{M} = I - M & , \quad N &= (X'X)^{-1}X'(I * M)X \\ n &= T - p & , \quad q &= n(n+2) + v_2 \operatorname{tr}(\bar{M} * \bar{M}) \\ \phi &= \frac{\operatorname{tr}N(X'X)^{-1}QD}{\operatorname{tr}(X'X)^{-1}QD} & , \quad \theta &= \frac{v_2}{n + v_2} \end{aligned}$$

where "*" denotes the Hadamard product of matrices.¹ Note that $I * M = \operatorname{Diag}(m_{tt})$, $t = 1, \dots, T$, where m_{tt} is the t^{th} diagonal element of M .

Some observations about ϕ and θ in (2.7) will be useful for the main results. First we note that

$$(2.8) \quad 0 \leq \eta_* \leq \phi \leq \eta^* \leq 1$$

where η_* and η^* , respectively, are the minimum and the maximum characteristic roots of the matrix N . The result (2.8) will be proved in Section 3.2.

Next, regarding θ in (2.7) we observe that

$$(2.9) \quad \begin{aligned} 0 < \theta < 1, & \quad \text{when } \nu_2 > 0 \\ \theta < 0, & \quad \text{when } \nu_2 < 0 \text{ and } n \geq 2 \\ \theta = 0, & \quad \text{when } \nu_2 = 0. \end{aligned}$$

The inequality $0 < \theta < 1$ is obvious. When $\nu_2 < 0$, we note from a result in Rao (1973, p. 57) that $2 + \nu_2 \geq 0$. Hence $\theta < 0$ holds for all $n \geq 2$. When $n = 1$, $\theta < 0$ provided $-1 < \nu_2 < 0$.²

Using (2.7) and assuming disturbances to be small, we can now present the small-disturbance approximation for the risk of $\hat{\beta}$. These results are asymptotic by nature and require disturbances to be small; for details see Kadane (1971). It will be indicated later in this section that, under the normality of disturbances, the dominance conditions of $\hat{\beta}$ based on its exact risk are the same as those obtained by a small-disturbance approximation [also see Ullah (1980)]. We therefore use a small-disturbance approximation for the non-normal case where the exact expressions of risk would be difficult to obtain.

THEOREM. Under the assumptions (2.2) the risk of $\hat{\beta}$, up to order σ^4 , is

$$(2.10) \quad E(\hat{\beta} - \beta)' Q (\hat{\beta} - \beta) = \sigma^2 r_2 - \sigma^3 r_3 - \sigma^4 r_4$$

where

$$(2.11) \quad r_2 = \text{tr}(X'X)^{-1}Q, \quad r_3 = 2k\nu_1 \frac{\beta' D' Q (X'X)^{-1} X' (I^* \bar{M}) e}{\beta' C \beta};$$

e denotes a column unitary vector, and

$$(2.12) \quad r_4 = -kq \frac{\beta' D' Q D \beta}{(\beta' C \beta)^2} \left[k - \frac{2(n+v_2)}{q} \frac{\beta' C \beta}{\beta' D' Q D \beta} \{ (1 - \theta \phi) \operatorname{tr}(X' X)^{-1} Q D \right. \\ \left. - 2 \frac{\beta' C (X' X)^{-1} Q D \beta}{\beta' C \beta} (1 - \theta \frac{\beta' C N (X' X)^{-1} Q D \beta}{\beta' C (X' X)^{-1} Q D \beta}) \right].$$

The result (2.10) is derived in Section 3.1.³

Observing from (2.4) that the risk of b is

$$(2.13) \quad E(b - \beta)' Q (b - \beta) = \sigma^2 r_2$$

it follows from (2.10) that

$$(2.14) \quad E(\hat{\beta} - \beta)' Q (\hat{\beta} - \beta) - E(b - \beta)' Q (b - \beta) = -\sigma^3 r_3 - \sigma^4 r_4.$$

We shall analyze (2.14) for the cases $v_1 = 0$ and $v_1 \neq 0$ separately.

I. Results for Non-normal Symmetrical Distributions ($v_1 = 0$)

Let us write the expression (2.14) for symmetrical distributions ($v_1 = 0$) of disturbances as

$$(2.15) \quad E(\hat{\beta} - \beta)' Q (\hat{\beta} - \beta) - E(b - \beta)' Q (b - \beta) = -\sigma^4 r_4.$$

Now using (2.8), (2.9) and (2.12) we can easily verify the results in the following corollaries.

COROLLARY 1. For symmetrical leptokurtic disturbances ($v_1 = 0$ and $v_2 > 0$)

$$(2.16) \quad E(\hat{\beta} - \beta)' Q (\hat{\beta} - \beta) - E(b - \beta)' Q (b - \beta) < 0$$

when⁴

$$(2.17) \quad 0 < k \leq \frac{2(n+v_2)}{q} \delta_{*\mu*} [(1 - \theta \phi) d - 2(1 - \theta \eta_*)];$$

$$d = \frac{1}{\mu_*} \operatorname{tr}(X' X)^{-1} Q D > 2 + \Delta_1$$

where δ_* and η_* are the minimum characteristic roots of the matrices $C(D'QD)^{-1}$ and N , respectively, and μ^* is the minimum characteristic root of $(X'X)^{-1}QD$ along with⁵

$$(2.18) \quad \Delta_1 = \frac{2\theta}{1-\theta\phi}(\phi - \eta_*) > 0.$$

COROLLARY 2. For symmetrical platykurtic disturbances ($\nu_1 = 0$ and $\nu_2 < 0$)

$$E(\hat{\beta} - \beta)'Q(\hat{\beta} - \beta) - E(b - \beta)'Q(b - \beta) < 0$$

holds for $n \geq 2$, and for $n = 1$ provided $1 + \nu_2 > 0$, so long as

$$(2.19) \quad 0 < k \leq \frac{2(n+\nu_2)}{q} \delta_* \mu^* [(1 - \theta\phi)d - 2(1 - \theta\eta^*)]; \quad d > 2 + \Delta_2$$

where η^* is the maximum characteristic root of N ,

$$(2.20) \quad \Delta_2 = \frac{2\theta}{1-\theta\phi}(\phi - \eta^*) > 0$$

and the other terms are as given before.

Remarks: The following observations can now be made about the results in Corollaries 1 and 2.

(i) For a normal distribution ($\nu_2 = 0$), the conditions of dominance of $\hat{\beta}$ over b in (2.17) and (2.19) reduce to the following:

$$(2.21) \quad 0 < k \leq \frac{2\delta_* \mu^*}{n+2}(d-2); \quad d = \frac{1}{\mu^*} \text{tr}(X'X)^{-1}QD > 2$$

which tally with the conditions given in Judge and Bock (1978, Chap. 10, p. 234) on the basis of an exact risk function.⁶ When $D = Q^{-1}X'X$ and $C = X'X$, (2.21) compares with the result in Strawderman (1978, Theorem 6).

(ii) Comparing (2.17) and (2.19) with (2.21), it is clear that the conditions of dominance of $\hat{\beta}$ over b for symmetrical leptokurtic and platykurtic distributions are different than the conditions for normal distribution of disturbances.

(iii) When $D = I$ and $C = X'X$, $\hat{\beta}$ is the James and Stein (1961) Stein-type estimator. Its condition of dominance over b , for the symmetrical leptokurtic disturbances, can be written from Corollary 1 as (considering $Q = I$),

$$(2.22) \quad 0 < k \leq \frac{2(n+\nu_2)}{q} [(1 - \theta\phi)d - 2(1 - \theta\eta_*)]; \quad d = \lambda_* \sum_1^p \lambda_i^{-1} > 2 + \Delta_1$$

where Δ_1 is as given in (2.18), $\phi = \text{tr } N(X'X)^{-1} / \text{tr } (X'X)^{-1}$ from (2.7) and λ_* is the minimum of the characteristic roots $\lambda_1, \dots, \lambda_p$ of $X'X$. Similarly we can write the dominance condition for symmetrical platykurtic disturbances from Corollary 2. It is clear that in the non-normal symmetric distribution cases the range of dominance depends on the magnitude as well as direction of ν_2 , and on the data matrix. When $\nu_2 = 0$, (2.22) reduces to the dominance condition in the normal case, viz. $0 < k \leq \frac{2}{n+2}(d - 2); d > 2$, see, for example, Judge and Bock (1978).

These results indicate that the estimator $\hat{\beta}$ is useful compared to the OLS if the degree of collinearity measured here by $d = \lambda_* \sum_1^p \lambda_i^{-1}$ is greater than 2 for the normal case and greater than $2 + \Delta_1$ for the non-normal case. For the non-normal case the usefulness of these estimators will crucially depend on the magnitude of Δ_1 . It is likely that $d > 2 + \Delta_1$ may not be satisfied for the highly collinear data.

The dominance conditions for the ridge-type ($D = (X'X)^{-1}$) estimators, and for other choices of C , D and Q can be easily written from the Corollaries 1 and 2.

(iv) In general, if we choose C , Q and D such that

$$(2.23) \quad C(D'QD)^{-1} = I \quad \text{and} \quad (X'X)^{-1}QD = I$$

then we get estimators $\hat{\beta}$ from (2.6) whose dominance condition for $\nu_2 > 0$ will be

$$(2.24) \quad 0 < k < \frac{2(n+\nu_2)}{q} [(1 - \theta\bar{\eta})p - 2(1 - \theta\eta_*)]; \quad p > 2 + \Delta_1$$

where $\Delta_1 = 2\theta(1 - \theta\bar{\eta})^{-1}(\bar{\eta} - \eta_*) > 0$ from (2.18) and $\bar{\eta}$ is the simple average of the characteristic roots of N . For $\nu_2 < 0$, we replace η_* with η^* in (2.24) and note, from (2.9) that $\theta < 0$ in this case. Further, when $\nu_2 = 0$ (normal disturbances) then the dominance condition reduces to

$$(2.25) \quad 0 < k < \frac{2}{n+2}(p-2), \quad p > 2.$$

It is clear from (2.24) and (2.25) that the range of k as well as the condition on p are free from the data matrix in the normal case, but it is not so in the non-normal case. This restricts the applicability of the shrinkage type estimators for the non-normal case. For example, in the non-normal case we require that the number of exogenous variables, p , be greater than $2 + \Delta_1$, where Δ_1 depends on the data matrix through $\bar{\eta}$. In addition, both the range of k and the condition on p depend upon the shape of distribution through θ . Since Δ_1 is greater than zero, it is clear that in general p would be required to be greater than or equal to four. The exact numbers of p would, of course, depend upon Δ_1 . For the case of multivariate mean vector Brandwein (1979) and Brandwein and Strawderman (1980) have developed shrinkage estimators for $p \geq 4$.

We now look into the choices of C , Q and D which satisfy (2.23). For a given Q in the loss function, $(\hat{\beta} - \beta)'Q(\hat{\beta} - \beta)$, (2.23) gives $C = X'XQ^{-1}X'X$ and $D = Q^{-1}X'X$. Substituting these values of C and D in (2.5) and (2.6) we get the following estimator

$$\hat{\beta}_1 = [I + \frac{ks}{b'X'XQ^{-1}X'X} Q^{-1}X'X]^{-1} b$$

whose dominance condition is given by (2.24). Similarly, for a given choice of D , (2.23) gives $Q = X'XD^{-1}$ and $C = D'X'X$. Using this value of C in (2.6) we get the estimator

$$\hat{\beta}_2 = [I + \frac{ks}{b'D'X'X} D]^{-1} b$$

whose dominance condition is (2.24) for the loss function $(\hat{\beta}_2 - \beta)'X'XD^{-1}(\hat{\beta}_2 - \beta)$.

This result gives the values of Q and C for the choices of $D = I$ and $D = (X'X)^{-1}$ in Stein-type and Ridge-type estimators, respectively. Remember that the choice of D is such that $Q = X'XD^{-1}$ is positive definite.

II. The Case of Skewed Distributions ($v_1 \neq 0$)

Let us write the expression (2.14) as

$$E(\hat{\beta} - \beta)'Q(\hat{\beta} - \beta) - E(b - \beta)'Q(b - \beta) = -\sigma^3 r_3 - \sigma^4 r_4$$

where r_4 is as given in (2.12) and r_3 from (2.11) is

$$r_3 = 2kv_1 \frac{\beta' D' Q (X' X)^{-1} X' (I^* \bar{M}) e}{\beta' C \beta} = 2kn \frac{\beta' D' Q \alpha}{\beta' C \beta}; \quad \alpha = \text{cov}(b, \bar{s})$$

where $\bar{s} = u' \bar{M} u / n = y' \bar{M} y / n$ is the disturbance variance estimator and

$n\alpha = n \text{cov}(b, \bar{s}) = (X' X)^{-1} X' E[u' \bar{M}_1 \cdot u] = v_1 (X' X)^{-1} X' (I^* \bar{M}) e$ is a $p \times 1$ vector

of covariance between b and \bar{s} . We note that if all the elements of

$\alpha = \text{cov}(b, \bar{s})$ and β are of the same (opposite) sign, then r_3 is positive

(negative), otherwise the sign of r_3 is not clear.

Recall that r_4 is positive under the conditions (2.17) when $v_2 > 0$, (2.19) when $v_2 < 0$ and (2.21) when $v_2 = 0$. Thus under the ranges of k implied by these conditions, $\hat{\beta}$ dominates b for the skewed distribution of disturbances, i.e.,

$$E(\hat{\beta} - \beta)'Q(\hat{\beta} - \beta) - E(b - \beta)'Q(b - \beta) < 0,$$

when β and $\alpha = \text{cov}(b, \bar{s})$ are of the same sign. This is a sufficient condition.

Thus we observe from above and the results in I that for the better performance of the class of estimators defined by (2.5) with respect to the least squares estimator, the set of conditions for non-normal distributions of disturbances may be quite different from those for normal distribution.

3. DERIVATIONS OF RESULTS

In this section we derive the risk function of $\hat{\beta}$ given in (2.10) of Section 2 and provide the proof of the inequality in (2.8).

3.1 Risk Function of $\hat{\beta}$

Let us rewrite the model (2.1) as

$$(3.1) \quad y = X\beta + \sigma v \quad (u = \sigma v)$$

so that as σ approaches 0, the disturbance term tends to be small. The elements of v are independently and identically distributed.

Thus from (2.2) we have for all t and t^* ($t, t^* = 1, 2, \dots, T$)

$$(3.2) \quad \begin{aligned} E(v_t) &= 0, & E(v_t v_{t^*}) &= 1 & \text{if } t = t^* \\ & & &= 0 & \text{if } t \neq t^* \\ E(v_t^3) &= v_1, & E(v_t^4) &= v_2 + 3. \end{aligned}$$

Now we have the following lemma:

LEMMA: If A is any nonstochastic matrix of order $T \times T$, then

$$(3.3) \quad E(v' A v) = \text{tr } A$$

$$(3.4) \quad E(v' A v \cdot v) = v_1 (I * A) e$$

$$(3.5) \quad E(v' A v \cdot v v') = v_2 (I * A) + (\text{tr } A) I + A + A'$$

where e is a $T \times 1$ vector with all elements unity.

Proof: Owing to independence and identical distribution of the elements of v , we observe that

$$(3.6) \quad E(v' A v) = \text{tr } A \cdot E(vv')$$

$$= \text{tr } A \cdot I$$

which gives (3.3).

Next, the t^* -th element of $E(v' A v \cdot v)$ is given by

$$(3.7) \quad \sum_{t_1, t_2}^T a_{t_1 t_2} E(v_{t_1} v_{t_2} v_{t^*})$$

$$= a_{t^* t^*} v_1$$

because the expectation term is nonzero only when $t_1 = t_2 = t^*$. This leads us to (3.4).

Similarly, the (t^*, t^{**}) -th element of $E(v' A v \cdot vv')$ is

$$(3.8) \quad \sum_{t_1, t_2}^T a_{t_1 t_2} E(v_{t_1} v_{t_2} v_{t^*} v_{t^{**}})$$

When $t^* = t^{**}$, it is easy to verify that (3.8) is equal to

$$(3.9) \quad \text{tr } A + (v_2 + 2)a_{t^* t^*}$$

and when $t^* \neq t^{**}$, it is equal to

$$(3.10) \quad a_{t^* t^{**}} + a_{t^{**} t^*} .$$

Combining (3.9) and (3.10) we obtain (3.5) from (3.8).

Now for the derivation of the result in (2.10) we first substitute (3.1) in (2.6) and note that $h = ks/b'Cb = k\sigma^2 v' \bar{M}v/b'Cb$ is at least of order σ^2 . Thus, for sufficiently small σ in Kadane's (1971) sense, we can write (2.5) as

$$\hat{\beta} = (I + hD)^{-1}b = (I - hD)b + \dots$$

or

$$\hat{\beta} - \beta = \sigma(X'X)^{-1}X'v - \frac{\sigma^2 kv' \bar{M}v D}{\beta' C\beta + 2\sigma\beta' C(X'X)^{-1}X'v + \sigma^2 v' X(X'X)^{-1}C(X'X)^{-1}X'v} \times$$

$$\{\beta + \sigma(X'X)^{-1}X'v\} + \dots$$

(3.11)

$$= \sigma(X'X)^{-1}X'v - \sigma^2 \frac{kv' \bar{M}v}{\beta' C\beta} \left[1 + 2\sigma \frac{\beta' C(X'X)^{-1}X'v}{\beta' C\beta} + \sigma^2 \frac{v' X(X'X)^{-1}C(X'X)^{-1}X'v}{\beta' C\beta} \right] \times$$

$$D\{\beta + \sigma(X'X)^{-1}X'v\} + \dots$$

Expanding the expression in square brackets and retaining terms to order $O(\sigma^3)$, we find

$$(3.12) \quad \hat{\beta} - \beta = \sigma\xi_1 + \sigma^2\xi_2 + \sigma^3\xi_3$$

where

$$(3.13) \quad \xi_1 = (X'X)^{-1}X'v$$

$$(3.14) \quad \xi_2 = \frac{kv' \bar{M}v}{\beta' C\beta} D\beta$$

$$(3.15) \quad \xi_3 = -\frac{kv' \bar{M}v}{\beta' C\beta} \left[D(X'X)^{-1}X'v - 2 \frac{\beta' C(X'X)^{-1}X'v}{\beta' C\beta} D\beta \right]$$

Thus we have, to order $O(\sigma^4)$,

$$(3.16) \quad E(\hat{\beta} - \beta)' Q(\hat{\beta} - \beta) = \sigma^2 E(\xi_1' Q \xi_1) + 2\sigma^3 E(\xi_1' Q \xi_2) + \sigma^4 E(\xi_2' Q \xi_2 + 2\xi_1' Q \xi_3).$$

Using (3.3), (3.4) and (3.5), it is easy to verify that

$$(3.17) \quad E(\xi_1' Q \xi_1) = E[v' X (X' X)^{-1} Q (X' X)^{-1} X' v] = \text{tr}(X' X)^{-1} Q$$

$$(3.18) \quad E(\xi_1' Q \xi_2) = -\frac{k}{\beta' C \beta} \beta' D' Q (X' X)^{-1} X' \cdot E[v' \bar{M} v \cdot v]$$

$$= -k v_1 \frac{\beta' D' Q (X' X)^{-1} X' (I * \bar{M})}{\beta' C \beta}$$

$$(3.19) \quad E(\xi_2' Q \xi_2) = k^2 \frac{\beta' D' Q D \beta}{(\beta' C \beta)^2} \text{tr} \bar{M} \cdot E[v' \bar{M} v \cdot v v']$$

$$= k^2 [n(n+2) + v_2 \text{tr} \bar{M} (I * \bar{M})] \frac{\beta' D' Q D \beta}{(\beta' C \beta)^2}$$

$$= k^2 v_1 \frac{\beta' D' Q D \beta}{(\beta' C \beta)^2}$$

$$(3.20) \quad E(\xi_1' Q \xi_3) = -\frac{k}{\beta' C \beta} \text{tr} X (X' X)^{-1} Q D \{I - \frac{2}{\beta' C \beta} \beta \beta' C\} (X' X)^{-1} X' \cdot E[v' \bar{M} v \cdot v v']$$

$$= -\frac{k}{\beta' C \beta} [n \text{tr} (X' X)^{-1} Q D + v_2 \text{tr} (X' X)^{-1} Q D \bar{N}]$$

$$- \frac{2}{\beta' C \beta} \{n \beta' C (X' X)^{-1} Q D \beta + v_2 \beta' C \bar{N} (X' X)^{-1} Q D \beta\}$$

where $\bar{N} = (X' X)^{-1} X' (I * \bar{M}) X = I - N$ by using

$$(3.21) \quad (I * \bar{M}) = I - (I * M) .$$

Substituting (3.17) - (3.21) in (3.16), we obtain the expression (2.10) for the risk, to order $O(\sigma^4)$, of $\hat{\beta}$ after a little algebraic manipulation.

3.2 Proof of the Inequality in 2.8

Let us write from equation (2.7)

$$\phi = \frac{\text{tr } N (X'X)^{-1} QD}{\text{tr } (X'X)^{-1} QD} = \frac{\text{tr } NL}{\text{tr } L}$$

where $N = (X'X)^{-1} X' (I_* M) X$; $M = X(X'X)^{-1} X'$, and $L = (X'X)^{-1} QD$. We require to show that (see (2.8))

$$0 \leq \eta_* \leq \phi \leq \eta^* \leq 1$$

where η_* and η^* are the minimum and maximum characteristic roots, respectively, of the matrix N .

First let us note that N is at least positive semidefinite, see Rao (1973, p. 77, problem 32). Further L is positive definite. Thus $\text{tr } NL / \text{tr } L$ is a non-negative quantity, and using Anderson and Gupta (1963, p. 524)

$$(3.22) \quad 0 \leq \eta_* \leq \frac{\text{tr } NL}{\text{tr } L} \leq \eta^*.$$

Now let P be a non-singular matrix such that $P' X' X P = I$ or $X' X = (P P')^{-1}$. Thus $M = X(X' X)^{-1} X' = X P P' X' = Z Z'$ ($Z = X P$) and $N = (X' X)^{-1} X' (I_* M) X = P Z' (I_* Z Z') Z P^{-1}$. Since P is non-singular, characteristic roots of $P^{-1} N P =$ characteristic roots of $N =$ characteristic roots of $Z' (I_* Z Z') Z$. But $Z' Z = I$ and $Z Z'$ is idempotent. Therefore using Poincaré separation theorem [Rao (1973), p. 64] characteristic roots of N are ≤ 1 . Using this in (3.22), the inequality in (2.8) follows.

FOOTNOTES

¹If $A = ((a_{ij}))$ and $B = ((b_{ij}))$ are two matrices with $i, j = 1, \dots, n$, then the Hadamard product is defined as $A \star B = ((a_{ij} b_{ij}))$, see Rao (1973, p. 30).

²When $n = 0$, $s = 0$ and $\hat{\beta}$ from (2.5) is equal to b . This value of n is therefore excluded in our study.

³When the disturbances are normal ($v_2 = 0$), the result in (2.10) compares with that of Ullah and Ullah (1978) for $D = I = Q$ and $C = X'X$. Also see Srivastava and Upadhyaya (1977).

⁴Note that $d > 2$ implies $p \geq 3$.

⁵In obtaining this result we note from Rao (1973, p. 74) that if A and B are two matrices then $\min_{\beta} \left(\frac{\beta' A \beta}{\beta' B \beta} \right)$ is the minimum root of $|A - \lambda B| = 0$.

⁶Notice from Section 3 that the risk function of $(I + hD)^{-1}b$ is identical to the risk function of $(I - hD)b$ up to the order of approximation considered. Judge and Bock have considered the estimators $(I - hD)b$.

REFERENCES

- Anderson, T. W. and S. Dasgupta (1963): "Some Inequalities on Characteristic Roots of Matrices," Biometrika 50, 522-524.
- Brandwein, A. C. (1979): "Minimax Estimation of the Mean of Spherically Symmetric Distributions Under General Quadratic Loss," Journal of Multivariate Analysis, 579-588.
- Brandwein, A. C. and W. E. Strawderman (1978): "Minimax Estimation of Location Parameters for Spherically Symmetric Unimodal Distributions Under Quadratic Loss," The Annals of Statistics, 377-416.
- Brandwein, A. C. and W. E. Strawderman (1980): "Minimax Estimation of Location Parameters for Spherically Symmetric Distributions with Concave Loss," The Annals of Statistics, 279-284.
- Gnanadesikan, R. (1977): Methods for Statistical Data Analysis of Multivariate Observations, John Wiley and Sons, New York.
- James, W. and C. Stein (1961): "Estimation with Quadratic Loss," Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 361-379.
- Judge, G. G. and M. E. Bock (1978): The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics, North-Holland Publishing Co., Amsterdam.
- Judge, G. G., W. E. Griffiths, R. C. Hill and T. C. Lee (1980): The Theory and Practice of Econometrics, John Wiley and Sons, Sussex.
- Kadane, J. B. (1971): "Comparison of k-class Estimators when the Disturbances are Small," Econometrica 39, 723-737.
- Rao, C. R. (1973): Linear Statistical Inference and its Applications, John Wiley and Sons, New York.

- Srivastava, V. K. and S. Upadhyaya (1977): "Properties of Stein-Like Estimators in Regression Model when Disturbances are Small," Journal of Statistical Research 11, 5-21.
- Strawderman, W. E. (1978): "Minimax Adaptive Generalized Ridge Regression Estimators," Journal of the American Statistical Association, 623-627.
- Ullah, A. (1980): "The Exact, Large-Sample and Small-Disturbance Conditions of Dominance of Biased Estimators in Linear Models," Economics Letters, 339-344.
- Ullah, A. and S. Ullah (1978): "Double k-class Estimators of Coefficients in Linear Regression," Econometrica 46, 705-722.
- Ullah, A., H. D. Vinod and K. R. Kadiyala (1981): "A Family of Improved Shrinkage Factors for the Ordinary Ridge Estimator," Economic Studies Quarterly, 164-175.
- Zellner, A. and W. Vandaele (1975): "Bayes-Stein Estimators for k-Means, Regression, and Simultaneous Equations," Studies in Bayesian Econometrics and Statistics (eds. S. E. Fienberg and A. Zellner), North-Holland Co., 627-653.