THE EGYPTIAN STATISTICAL JOURNAL
ISSR, Cairo Univ., Vol.37, No.2, 1993.

DOUBLE f-CLASS PREDICTORS IN LINEAR REGRESSION MODELS*

By: G.A. GHAZAL

This article develops a family of biased predictors, namely the double f-Class, in the linear regression model. Employing the small disturbance approach, I study the bias and mean square error of the double f-Class predictor.

1. INTROUDUCTION

The standard econometric analysis of the linear regression model is concerned with the properties of the linear unbiased estimator (and predictor) that has the smallest variance [see Chow (1983) and Theil (1961), (1962)].

In that approach, therefore, attention is restricted at the outset to linear and unbiased estimators and predictors, even though "better" estimators and predictors might be available. Linearity is desirable because such estimators and predictors are usually easy to compute and because their statistical properties can often be analyzed

---

* The author is grateful to the anonymous refrees for their most helpful comments.

without the need for rather advanced mathematics. The criterion of unbiasedness is attractive both because it seems intuitively desirable that the mathematical expectation of an estimator (or predictor) be equal to what is being estimated (or predicted).

But the price to be paid, in terms of variance, by concentrating on linear unbiased estimators and predictors may be unacceptably high. For example, many empirical investigators deal with a form of this problem by dropping a nonsignificant variable from a regression when a high correlation between it and another variable causes a large standard error for both variables, this is the problem of multicollinearity. Of course, if the original model is "correct", estimates based on the reduced model will be biased

To analyze a wider class of estimators and predictors, therefore, we need a criterion that is not confined to linear unbiased estimation and prediction. An attractive choice is the "mean squared error" or "quadratic loss function".

$$L(\hat{\theta}, \theta) = [(\hat{\theta} - \theta)'(\hat{\theta} - \theta)] \tag{1.1}$$

where $\hat{\theta}$ is an estimater (or predictor) of θ, which may be a scalar or a vector. It is easy to see that

$$E[L(\hat{\theta},\theta)] = [E(\hat{\theta}) - \theta]'[E(\hat{\theta}) - \theta] + cov(\hat{\theta})$$

$$= B^2(\hat{\theta}) + cov(\hat{\theta}) \qquad (1.2)$$

where $B^2(\hat{\theta})$ is the squared bias of $\hat{\theta}$. Under this criterion a biased estimator and predictor will perform better than an unbiased estimator if the resulting reduction in variance is sufficient to offset the increase in squared bias.

The loss function (1-1) may be expressed as $\Sigma(\hat{\theta}_i - \theta_i)^2$ and in this form it is clear that $L(\hat{\theta},\theta)$ depends on units of measurement of the individual θ_i. To avoid this problem, the loss function may be modified by introducing explicit weights a_i - for example, $\Sigma a_i(\hat{\theta}_i - \theta_i)^2$, $a_i \geqslant 0$; more generally, a positive semidefinite matrix A may be included to weight the losses:

$$L(\hat{\theta}, \beta; A) = [(\hat{\theta} - \theta)' A(\hat{\theta} - \theta)] \qquad (1 - 3)$$

In the usual regression, $Y = X\beta + U$, we can, for instance, consider the risk function

$$R(b,\beta, X'X) = E[(b - \beta)' X'X(b - \beta)]$$

$$= E[(Xb - X\beta)' (Xb - X\beta] \qquad (1-4)$$

It may be interpreted as a function of the differences between the predicted values and the expected values of the sample Y. This risk function has been termed a "prediction goal" by Bibby and Toutenburg (1977). Needless to say it is of importance consider the preformance of various estimators under a "forecast goal".

In fact, Stein (1956), and James and Stein (1961) suggested a biased estimator (and hence a biased predictor) for the orthonormal linear statistical model which dominates the least squares estimator (predictor) in the sense that the sum of its component mean squared errors is smaller than that of the former, provided at least three parameters are to be estimated. Extension of the James and Stein estimator and predictor for nonorthogonal regressors have been discussed by Sclove (1968), Bock (1975) and Zellner and Vandaele (1975).

To compare the performance of various estimators under a "forecast goal", we consider, in this paper, the prediction of the dependent variable of the general linear model with the usual nonorthogonal regressors. We develop a family of biased predictors, namely double f-class, by using an operational variant of the minimum predictive mean squared error [see: Vinod and Ullah, 1981, P. 33] which depends on unknown parameters. We present the model and the double f-class predictor in section 2. The f_1 and f_2 in double f-class are taken as arbitrary scalars which could be stochastic or non-stochastic. For $f_2 = 1$ we note that [James and Stein (1961)] Stein-rule estimator (predictor) in the regression context is a member of the double f-class. In section 3 we analyze the approximate bias and mean square error (MSE) of the double f-class predictor for the case in which f_1 is an arbitrary constant and f_2 is a nonnegative constant less than or equal to one. These approximations are true in "small sigma asymptotic sense" discussed by Kadane (1970 - 1971). Next, in section 4 we give proofs of the theorems, stated in section 3.

2. The Model and Predictors

(2-1) The Model and Unbiased Predictor

Let us consider the regression model

$$\underset{(Tx1)}{y} = \underset{(Txk)}{x} \; \underset{(kx1)}{\beta} + \underset{(Tx1)}{u} \qquad \ldots (2\text{-}1)$$

where

y is a Tx1 Vector of obsrvations on the dependent variable,

x is a Txk matrix of observations on k explanatory variable,

β is a kx1 parameter vector, and

u is a Tx1 distrubance vector.

Consider the problem of predicting a single drawing of the dependent variable given the vector of explanatory variables. The actual drawing will be

$$\underset{(1x1)}{y_*} = \underset{(1xk)}{x_*'} \underset{(kx1)}{\beta} + \underset{(1x1)}{u_*} \qquad \ldots(2\text{-}2)$$

where

y_* is the 1x1 scalar (unknown) value of the dependent variable.

x_* is the kx1 vector of (known) future values of the explanatory variables ,

u_* is the 1x1 scalar value of the prediction distrubance .

We state the following conventional assumptions:

Assumption (1)

The matrix x and the vector x_* of explanatory variables are nonstochastic, and the rank of x is K.

Assumption (2)

The sample disturbance vector u and the prediction disturbance u_* are distributed indpendently as multivaride normal, and univariate normal respectively. Therefore, we shall assume.

$$E(u) = \underset{(Tx1)}{0} \quad , \qquad E(u_*) = \underset{(1x1)}{0}$$

$$E(u'u) = \sigma^2 I_T \quad , \qquad E(u_*^2) = \sigma^2 \qquad \ldots(2\text{-}3)$$

$$E(u_* u) = \underset{(Tx1)}{0}$$

The normality of the vector u and the scalar u_* is required for deriving the results in section 3.

Assumption (3)

The sample size T is greater than the total number of regressors k in (2-1).

For a point prediction, we can use the OLS predictor:

$$P^{*} = x_{*}' b \qquad \ldots (2\text{-}4)$$

where b is the OLS estimator

$$b = (x'x)^{-1} x'y \qquad \ldots (2\text{-}5)$$

The predictor in (2-4) is a linear function of y, and its expected value is given by

$$E(p^{*}) = x_{*}' E(b) = E(y_{*}) \qquad \ldots (2\text{-}6)$$

Hence, in the sense that

$$E(p^{*} - y_{*}) = 0 \qquad \ldots (2\text{-}7)$$

p^{*} is unbiased predictor for y_{*}. In fact, among all linear, unbiased predictors, it has the smallest variance for the prediction error. The prediction variance (= MSE) is:

$$\sigma^{2}_{p^{*}} = MSE\ (p^{*}) = E(P^{*} - y_{*})^{2}$$

$$= \sigma^{2}[1+\underline{tr}(x_{*}x_{*}')\ (x'x)^{-1}] \qquad \ldots (2\text{-}8)$$

In practice σ^{2} is replaced by the OLS unbiased estimator

$$S^{2} = \frac{1}{n}\ \hat{u}'\hat{u} = \frac{1}{n}\ y'My \qquad \ldots (2\text{-}9)$$

where

$$n = T-K \qquad \hat{u} = My, \qquad \text{and}$$

$$M = I - x(x'x)^{-1}x' \qquad \ldots (2\text{-}10)$$

(2-2) A Biased Predictor

Consider now a class of linear predictors

$$\underset{(1\times1)}{P} = \underset{(1\times T)}{C'}\ \underset{(T\times1)}{y} \qquad \ldots (2\text{-}11)$$

where C is the arbitrary Tx1 vector. The predictive mean squared error of $\hat{P}$ can be written as:

$$MSE(\hat{P}) = E(\hat{P}-y_*)^2$$

$$=(C'x - x_*')\beta\ \beta'(C'x - x_*')' + \sigma^2 C'C + \sigma^2 \quad \ldots (2\text{-}12)$$

The matrix C for which (2-12) is minimum is

$$C = (x\beta\ \beta'x' + \sigma^2 I)^{-1}\ x\beta\ \beta'x_* \quad \ldots (2\text{-}13)$$

substituting this in (2-11) gives "the minimum mean square error predictor" as

$$\hat{P} = x_*'\beta\ \beta'x'(x\beta\ \beta'x' + \sigma^2 I)\ y \quad \ldots (2\text{-}14)$$

For more details see Bibby and Toutenburg [1977, Ch. 5, PP. 84-96].

An alternative form of (2-14) is

$$\hat{P} = \left(\frac{\beta'x'y}{\sigma^2 + \beta'x'x\beta}\right) x_*'\beta \quad \ldots (2\text{-}15)$$

where $(x\beta\ \beta'x' + \sigma^2 I)^{-1} = \frac{1}{\sigma^2}[I - x\ \beta(\sigma^2 + \beta'x'x\beta)\ \beta'x']$

by using matrix inversion result in Rao [1973, P.33]. Further, the predictor $\hat{P}$ in (2-15) can be written as

$$\hat{P} = \left(\frac{(y-u)'y}{\sigma^2 + (y-u)'(y-u)}\right) x_*'\beta \quad \ldots (2\text{-}16)$$

where we use $x\beta = y-u$.

We note that the predictor $\hat{P}$ depends on unknown values of β and σ^2. Thus we propose an operational variant of (2-16) as

$$\hat{P}_0 = \left(\frac{(y-\hat{u})'y}{(\frac{1}{n})\hat{u}'\hat{u} + (y-\hat{u})'(y-\hat{u})}\right) P^* \quad \ldots (2\text{-}17)$$

where P^*, n and $\hat{u}$ are as given in (2-4) and (2-10), respectively.

Moreover, we can write $\hat{P}_0$ in an alternative form as

$$\hat{P}_0 = \left(1 - \frac{\hat{u}'\hat{u} / n}{y'y - (1 - \frac{1}{n})\, \hat{u}'\hat{u}}\right) P^* \qquad \ldots (2\text{-}18)$$

(2-3) Family of Double f-Class Predictors:

A natural generalization of $\hat{P}_0$ in (2-18) can be found in the following double f-Class predictors:

$$P_{f_1,f_2} = \left(1 - \frac{f_1\, \hat{u}'\hat{u}}{y'y - f_2 \hat{u}'\hat{u}}\right) P^* \qquad \ldots (2\text{-}19)$$

where f_1 and f_2 arbitrary scalars which may be stochastic or nonstochastic. P_{f_1,f_2} in (2-19), represents a family of biased predictors. We shall give in the following section the the moments of the predictor in (2-19) for fixed f_2, when $0 \leq f_2 \leq 1$. It is interesting to note that for $f_1 = 0$, P_{f_1,f_2} reduces to the OLS predictor P^* in (2-4). For the value $k_2=1$, (2-19) can be written as

$$P_{f_1,1} = \left(1 - f_1 \frac{\hat{u}'\hat{u}}{y'y - \hat{u}'\hat{u}}\right) P^* \qquad \ldots (2\text{-}20)$$

we may regard the predictor $P_{f_1,1}$ as the James and Stein-rule (1961) Stein-rule predictor for the nonorthogonal case.

(2-4) Why a Small-Disturbance Approach ?

The sampling error of the predictor (2-19) can be written as

$$P_{f_1,f_2} - y_* = (P^* - y_*) - f_1 \delta\, P^* \qquad \ldots (2\text{-}21)$$

where

$$\delta = \frac{y'My}{y'Ny} \quad \ldots \quad (2-22)$$

M is as defined in (2-10) and N is given by

$$N = I_T - f_2 M \quad \ldots \quad (2-23)$$

As δ is a ratio of two quadratic forms, we can use Lemma (1) in Sawa (1972,P.658) to derive the exact moments of the double f-class predictors P_{f_1,f_2} in (2-19). In order to apply this lemma, the quadratic form y'Ny appearing in the denominator of (2-22) is required to be nonnegative. According to Lemma (2) in Sawa (1972, P. 658) the TxT matrix N in (2-23) is non-negative definite if and only if $f_2 < 1$. Therefore we restrict our analysis to the case where $0 \leq f_2 \leq 1$.

However, in these situations the exact moments often have a complicated mathematical structure so that it is difficult to use them in further studies and comparisons.

On the other hand, it has been shown [see: Sawa (1972), Ullah and Ullah (1978), Vinod, Ullah and Kadiyala (1981)] that the results obtained from the exact moments are identical with those obtained from approximate moments derived by Kadane's (1971) small-σ expansion. A set of rigorous mathematical conditions for this equality can perhaps be developed following the work of Berger (1976) and Casella (1980).

Depending upon these findings we shall use Kadane's (1970,1971) principle of small - σ expansion to derive the approximate formulae for the bias and the MSE of the double f-class predictors P_{f_1,f_2} (2-19) for

$0 \leqslant f_2 \leqslant 1$.

§3. Statement of Results

In this section we shall analyze the approximate bias, and the approximate MSE of the double f-class predictors P_{f_1, f_2} (2-19) for the case in which f_1 is arbitrary constant and f_2 is nonnegative constant less than or equal to one. These approximations are true in "small - σ asymptotic"

The following two theorems can now be stated.

Theorem (1)

Using the small-disturbance asymptotic distribution of the double f-class predictors (2-19), the bias, to order $0(\sigma^4)$, is given by

$$\text{Bias } (P_{f_1, f_2}) = E\,(P_{f_1, f_2} - y_*)$$

$$= -nf_1 \left\{ \frac{\sigma^2}{(\beta' x' x\, \beta)} + [(n + 2) f_2 - T] \frac{\sigma^4}{(\beta' x' x \beta)^2} \right\} (X_*' \beta)$$

when $0 \leqslant f_2 \leqslant 1$. ... (3 - 1)

Theorem (2)

According to small-disturbance asymptotic distribution of the double f-class predictors (2-19), the mean squared error, to order $0(\sigma^6)$, is given by

$$\text{MSE } (P_{f_1, f_2}) = \sigma^2 \left[1 + \underline{tr}\, (x_* x_*')(x'x)^{-1}\right]$$

$$+ \sigma^4 \frac{nf_1}{(\beta' x' x \beta)^2} \{ [4+f_1 (n+2)] (\beta' x_* x_*' \beta) - 2(\beta' x' x \beta) [\underline{tr}(x_* x_*')(x' x)^{-1}] \}$$

$$- \sigma^6 \frac{2nf_1}{(\beta'x'x\beta)^3} \Big[\{4[T-f_2(n+2)]+f_1(n+2)[T+2-f_2(n+4)]\}(\beta'x_*x_*'\beta)$$

$$+ [f_2(n+2)-T-f_1 \frac{(n+2)}{2}](\beta'x'x\beta)[\underline{tr}\ (x_*x_*')(x'x)]^{-1}\Big\} \quad \ldots \quad (3-2)$$

when $0 \leq f_2 \leq 1$.

Next, the two corollaries below may be proved

Corollary(1)

The following results regarding the bias of the double f-class predictor in (2-21) are true for $0 \leq f_2 \leq 1$ and $f_1 \geq 0$.

a) $P_{f_1 f_2}$ is unbiased only for $f_1=0$ and in that case it is the OLS predictor.

b) The direction of the bias is opposite to the sign of $(X_*'\beta)$.

c) As $\sigma \to 0$, the bias terms vanishes.

Corollary (2)

The double f-class predictor P_{f_1,f_2} in (2-21) dominates OLS predictor P^* in (2-4), in the sense that

$$MSE(P_{f_1,f_2}) - MSE(P^*) < 0 \qquad \ldots \qquad (3-3)$$

for

$$f_1 \leq -\frac{2}{n+2} \qquad \ldots \qquad (3-4)$$

and for any f_2 in

$$0 \leq f_2 \leq 1 \qquad \ldots \qquad (3-5)$$

Proof (of corollary (2))

Up to order σ^4 we have, from (3-2),

$$MSE(P_{f_1,f_2}) - MSE(P^*)$$
$$= \frac{\sigma^4 nf_1}{(n+2)} \frac{(\beta' x_* x_*' \beta)}{(\beta' x' x \beta)^2} \left\{ f_1 - \frac{2}{(n+2)} \left[\left(\frac{\beta' x_* x_*' \beta}{\beta' x' x \beta} \right)^{-1} \left(\underline{tr}(x_* x_*')(x'x)^{-1} \right) - 2 \right] \right\} \quad \ldots \quad (3-6)$$

wherein (2-8) has been utilized.

It is clear that the kxk matrix $(x_* x_*')$ is positive semi-definite with (k-1) zero eigenvalues and 1 possitive eigenvalue. i.e.,

$$\lambda_1 = \lambda_2 = \ldots = \lambda_{k-1} = 0 \qquad \text{and} \qquad \lambda_k \neq 0$$

hence

$$\underline{tr}\ (x_* x_*')(x'x)^{-1} = \sum_{i=1}^{k} \lambda_i = \lambda_k \quad \ldots \quad (3-7)$$

Moreover, according to a result in Giri [1977, Theorem 1.7.2, p.17] $0 \leq \beta' x_* x_*' \beta \,/\, \beta' x' x \beta \leq \lambda_k$. Thus

$$\frac{1}{\lambda_k} \leq \left(\frac{\beta' x_* x_*' \beta}{\beta' x' x \beta} \right)^{-1} \leq \infty \quad \ldots \quad (3-8)$$

Using (3-7) and (3-8) in (3-6) the result in the above corollary (2) follows by looking into the condition under which $MSE(P_{f_1 f_2}) - MSE(P^*) < 0$.

4. Proof of Theorems

(4-1) Preliminary

First, let us write

$$A = (x'x)^{-1} x' \quad \ldots \quad (4-1)$$

$$D = \beta' x' x \beta \quad \ldots \quad (4-2)$$

then, using (2-10),(2-23) and (4-1), it is easy to verify that

$$AM = 0 \quad \text{and} \quad Mx = 0 \quad \ldots \quad (4-3)$$

$$MN = (1-f_2)M \quad \ldots \quad (4-4)$$

$$AN = A \quad \ldots \quad (4-5)$$

$$ANx = Ax = I_k \quad \ldots \quad (4-6)$$

$$AA' = (x'x)^{-1} \quad \ldots \quad (4-7)$$

Next, the following notation is useful in the proofs of Theorems 1 and 2 :

$$\underline{tr}\ M = T-K = n \quad \ldots \quad (4-8)$$

$$\underline{tr}\ N = T-f_2\ n \quad \ldots \quad (4-9)$$

$$\underline{tr}\ MN = (1-f_2)\ n \quad \ldots \quad (4-10)$$

Finally, the following lemma will be used repeatedly in the proofs of Theorems 1 and 2.

Lemma (1)

Let f, G and H be symmetric matrices with non-stochastic elements, then

$$E(u'Hu) = \sigma^2(\underline{tr}\ H) \quad \ldots \quad (4-11)$$

$$E(u'Gu.u'Hu) = \sigma^4[(\underline{tr}\ G)(\underline{tr}\ H)+2(\underline{tr}\ GH)] \quad \ldots \quad (4-12)$$

$$E(u'Fu.u'Gu.u'Hu)=\sigma^6[(\underline{tr}\ F)(\underline{tr}\ G)(\underline{tr}\ H)+2(\underline{tr}\ F)(\underline{tr}\ GH)$$
$$+2(\underline{tr}\ G)(\underline{tr}\ FH)+2(\underline{tr}\ H)(\underline{tr}\ FG)+8(\underline{tr}\ FGH)] \quad \ldots \quad (4\text{-}13)$$

The results in (4-11) and (4-12) can be obtained from Kadane (1971), whereas the result in (4-13) can be obtained from Srivastava (1970).

(4-2) Proof of Theorem (1)

Let us write the regression model as

$y = x\beta + \sigma u$ where[1] $\sigma u \sim N(0, \sigma^2 I)$ or $u \sim N(0,I)$.

The sampling error of the double f-class predictors (2-19) can, then, be written as

$$P_{f_1,f_2} - y_* = \sigma x_*' Au - \sigma u_* - f_1 \sigma^2 \frac{U'Mu}{h} \cdot (x_*'\beta + \sigma x_*' Au) \quad \ldots \quad (4\text{-}14)$$

where

$$h = D + 2\sigma\beta'x'u + \sigma^2 u'Nu \quad \ldots \quad (4\text{-}15)$$

Now, for sufficiently small-σ we write h^{-1} and h^{-2} in terms of the following Binomial expansions:

$(1+\lambda)^{-1} = 1 - \lambda + \lambda^2 - \lambda^3 + \ldots$ and $(1-\lambda)^{-2} = 1 - 2\lambda + 3\lambda^2 - 4\lambda^3 + \ldots\ldots.$

Thus we have

$$\frac{1}{h} = \frac{1}{D}\left[1 - \left(\frac{2\sigma\beta'x'u + \sigma^2u'Nu}{D}\right) + \left(\frac{2\sigma\beta'x'u + \sigma^2u'Nu}{D}\right)^2 - \left(\frac{2\sigma\beta'x'u + \sigma^2u'Nu}{D}\right)^3 + \ldots\ldots.\right] \quad \ldots \quad (4\text{-}16)$$

---

1) Similarly, we can write $y_* = x_*'\beta + u_*$ as $y_* = x_*'\beta + \sigma u_*$ where $\sigma u_* \sim N(0, \sigma^2)$ or $u_* \sim N(0,1)$.

and similarly

$$\frac{1}{h^2} = \frac{1}{D^2}\left[1 - 2\left(\frac{2\sigma\beta' x' u + \sigma^2 u' N u}{D}\right) + 3\left(\frac{2\sigma\beta' x' u + \sigma^2 u' N u}{D}\right)^2 - 4\left(\frac{2\sigma\beta' x' u + \sigma^2 u' N u}{D}\right)^3 + \ldots\ldots\ldots\right] \quad \ldots \quad (4\text{-}17)$$

Now using (4-16) in the expression for $(P_{f_1,f_2} - y_*)$, and collecting terms up to σ^4 we obtain

$$P_{f_1,f_2} - y_* = \sigma H_1 + \sigma^2 H_2 + \sigma^3 H_3 + \sigma^4 H_4 \quad \ldots \quad (4\text{-}18)$$

where

$$H_1 = x_*' A u + u_* \quad \ldots \quad (4\text{-}19)$$

$$H_2 = -\frac{f_1}{D}(u' M u)(x_*' \beta) \quad \ldots \quad (4\text{-}20)$$

$$H_3 = -\frac{f_1}{D}(u' M u)(x_*' A u) + \frac{2 f_1}{D^2}(u' M u)(\beta' x' u)(x_*' \beta) \quad \ldots \quad (4\text{-}21)$$

$$H_4 = \frac{2 f_1}{D^2}(u' M u)(\beta' x' u)(x_*' A u) + \frac{f_1}{D^2}(u' M u)(u' N u)(x_*' \beta) - \frac{4 f_1}{D^3}(u' M u)(\beta' x' u)(x_*' \beta) \quad \ldots \quad (4\text{-}22)$$

under the normality assumption of the disturbances we find

$$E(H_1) = 0 \quad \ldots \quad (4\text{-}23)$$

$$E(H_3) = 0 \quad \ldots \quad (4\text{-}24)$$

Moreover, we have

$$E(H_2) = -\frac{n f_1}{D}(x_*' \beta) \quad \ldots \quad (4\text{-}25)$$

$$E(H_4) = \frac{2 f_1 n}{D^2}(x_*' \beta) + \frac{f_1 n}{D^2}[T - f_2(n+2) + 2](x_*' \beta) - \frac{4 f_1}{D^3} n D (x_*' \beta)$$

$$= -\frac{n f_1}{D^2}[(n+2) f_2 - T](x_*' \beta) \quad \ldots \quad (4\text{-}26)$$

wherein the results in (4-11) and (4-12) have been utilized.

Combining (4-18), (4-24), (4-25) and (4-26), the result in Theorem (1) is obtained.

(4-3) Proof of Theorem (2)

Using (4-16) and (4-17) in the expression for $(P_{f_1,f_2} - y_*)^2$ and collecting only those terms which contribute up to order σ^6, we obtain

$$(P_{f_1,f_2} - y_*)^2 = \sigma^2 H_1^2 + 2\sigma^3 H_1 H_2 + \sigma^4(2H_1H_3 + H_2^2) + \sigma^5(2H_1H_4 + 2H_2H_3) + \sigma^6(2H_1H_5 + 2H_2H_4 + H_3^2) \quad \ldots \quad (4\text{-}27)$$

where H_1 to H_4 are as defined in (4-19) to (4-22), and H_5 is given by

$$H_5 = \frac{f_1}{D^2}(u'Mu)(u'Nu)(x_*'Au) - \frac{4f_1}{D^3}(u'Mu)(\beta'x'u)^2(x_*'Au) - \frac{4f_1}{D^3}(u'Mu)(u'Nu)(\beta'x'u)(x_*'\beta) + \frac{8f_1}{D^4}(u'Mu)(\beta'x'u)^3(x_*'\beta) \quad \ldots \quad (4\text{-}28)$$

Now, using the normality of disturbances, it is easy to see that

$$E(H_1H_2) = 0 \quad \ldots \quad (4\text{-}29)$$

$$E(H_1H_4) = 0 \quad \ldots \quad (4\text{-}30)$$

$$E(H_2H_3) = 0 \quad \ldots \quad (4\text{-}31)$$

Next, using the results in (4-11) to (4-13) the other six expectations on the right-hand side of (4-27) can be evaluated as follows:

(i) $E(H_1^2) = [\underline{tr}\ (x_*x_*')(x'x)^{-1}] + 1$ ... (4-32)

wherein (4-19), (4-11) and (4-7) have been used.

(ii) $E(H_1H_3) = -\frac{f_1}{D} n[\underline{tr}\ (x_*x_*')(x'x)^{-1}] + \frac{2f_1}{D^2} n(\beta'x_*x_*'\beta)$ ... (4-33)

wherein (4-12), (4-3), (4-6), (4-7) and (4-8) have been utilized.

(iii) $E(H_2^2) = \frac{f_1^2}{D^2} n(n+2)(\beta'x_*x_*'\beta)$ ... (4-34)

wherein (4-12) and (4-8) have been used.

(iv) $E(H_1H_5) = \frac{f_1}{D^2} E[(u'Mu)(u'Nu)(u'A'x_*x_*'Au)]$

$- \frac{4f_1}{D^3} E[(u'Mu)(u'x\beta\beta'x'u)(u'A'x_*x_*'Au)]$

$- \frac{4f_1}{D^3} E[(u'Mu)(u'Nu)(u'A'x_*\beta'x'u)](x_*'\beta)$

$+ \frac{8f_1}{D^4} E[(u'Mu)(u'x\beta\beta'x'u)(u'A'x_*\beta'x'u)(x_*'\beta)$ ... (4-35)

Herein the terms with expectation zero have been dropped. The expectations of the four terms on the right-hand side of (4-35) are equal to

$E[u'Mu.u'Nu.u'A'x_*x_*'Au] = n[T-f_2(n+2) + 4][\underline{tr}\ (x_*x_*')(x'x)^{-1}]$ ... (4-36) wherein (4-3), (4-5), (4-7), (4-8), (4-9) and (4-10) have been utilized. Similarly for the expectations of the second, third, and fourth terms in (4-35) we have

$$E[u'Mu.u'x\,\beta\beta'x'u.u'A'x_*x_*'Au]$$

$$= nD[\underline{tr}(x_*x_*')(x'x)^{-1}]+2n(\beta'x_*x_*'\beta) \quad \ldots (4\text{-}37)$$

$$E[u'Mu.u'Nu.u'A'x_*\beta'x'u](x_*'\beta)$$

$$= n[\,T-f_2(n+2)+4](\beta'x_*x_*'\beta) \quad \ldots (4\text{-}38)$$

$$E[u'Mu.u'x\,\beta\beta'x'u.u'A'x_*\beta'x'u](x_*'\beta)=3nD(\beta'x_*x_*'\beta) \quad \ldots (4\text{-}39)$$

Using (4-36) to (4-39) in (4-35) we get

$$E(H_1H_5) = \frac{nf_1}{D^2}[\,T-f_2(n+2)][\underline{tr}(x_*x_*')(x'x)^{-1}]$$
$$- \frac{nf_1}{D^3}[\,4T-4f_2(n+2)](\beta'x_*x_*'\beta) \quad \ldots (4\text{-}40)$$

(V) $$E(H_2H_4) = -\frac{2f_1^2}{D^3}E[u'Mu.u'Mu.u'A'x_*\beta'x'u](x_*'\beta)$$
$$- \frac{f_1^2}{D^3}E[u'Mu.u'Mu.u'Nu](\beta'x_*x_*'\beta)$$
$$+ \frac{4f_1^2}{D^4}E[u'Mu.u'Mu.u'x\,\beta\beta'x'u](\beta'x_*x_*'\beta) \quad \ldots (4\text{-}41)$$

The three expectations on the right-hand side of (4-41) can be evaluated on the same lines of (4-36). Thus, we have

$$E[u'Mu.u'Mu.u'A'x_*\beta'x'u](x_*'\beta)=n(n+2)(\beta'x_*x_*'\beta) \quad \ldots (4\text{-}42)$$

$$E[u'Mu.u'Mu.u'Nu](\beta'X_*X_*'\beta)=n.(n+2)(T-nf_2-4f_2+4)(\beta'X_*X_*'\beta) \quad \ldots \quad (4-43)$$

$$E[u'Mu.u'Mu.u'X\beta\beta'X'u](\beta'X_*X_*'\beta)=n(n+2)D(\beta X_*X_*'\beta) \quad \ldots \quad (4-44)$$

using (4-42) to (4-44) in (4-41) we get

$$E(H_2H_4) = -\frac{f_1^2}{D^3} n(n+2)(T-nf_2-4f_2+2)(\beta'X_*X_*'\beta) \quad \ldots \quad (4-45)$$

$$\text{(vi)} \quad E(H_3^2) = \frac{f_1^2}{D^2} E[u'Mu.u'Mu.u'A'X_*X_*'Au]$$

$$- \frac{4f_1^2}{D^3} E[u'Mu.u'Mu.u'A'X_*\beta'X'u](X_*'\beta)$$

$$+ \frac{4f_1^2}{D^4} E[u'Mu.u'Mu.u'X\beta\beta'X'u](\beta'X_*X_*'\beta) \quad \ldots \quad (4-46)$$

The third term on the right-hand side of (4-46) is equal to (4-44), and the other two terms are equal to

$$E[u'Mu.u'Mu.u'A'X_*X_*'Au]=n(n+2)[\underline{tr}(X_*X_*')(X'X)^{-1}] \quad \ldots \quad (4-47)$$

$$E[u'Mu.u'Mu.u'A'X_*\beta'X'u]=n(n+2)(\beta'X_*X_*'\beta) \quad \ldots \quad (4-48)$$

using (4-47), (4-48) and (4-44) in (4-46) we get

$$E(H_3^2) = \frac{f_1^2}{D^2} n(n+2)[\underline{tr}(X_*X_*')(X'X)^{-1}] \quad \ldots \quad (4-49)$$

From (4-27), (4-29), (4-30), (4-31), (4-32), (4-33), (4-34), (4-40), (4-45) and (4-49), we get the result stated in Theorem (2).

REFERENCES

1) Berger, J.O.,(1976), "Tail Minimaxity in Location Vector Problems and Its Applications", Annals of Statistics, PP. 33-50.

2) Bibby, J., and Toutenburg, H. (1977), "Prediction and Improved Estimation in Linear Models", John Wiley & Sons, New York.

3) Bock, M. E. (1975), "Minimax Estimators of the Meam of a Multivariate Distribution", The Annals of Statistics, PP. 209-218

4) Casella, G., (1980), "Minimax Ridge Regression Estimation", The Annals of Statistics, PP.1036-1056.

5) Chow, G. C., (1983), "Econometrics", Mc Graw-Hill, New York.

6) Giri, N. C., (1977), "Multivariate Statistical Inference" Academic Press, Inc., New York.

7) Goldberger, A. S., (1962), "Best Linear Unbiased Prediction in the Generalized Linear Regression Model", Journal of the American Statistical Association, PP. 369-375.

8) James, W., and C. Stein (1961), "Estimation with Quadratic Loss., "Proceeding of the Fourth Berkèley Symposium on Mathematical Statistics and Probability. Berkeley: University of California press, PP. 361-379.

9) Kadane, J.B., (1970), "Testing Overidentifying Restrictions When the Disturbances are Small", Journal of the American Statistical Association, PP.182-185.

10) Kadane, J.B., (1971), "Comparison of K-Class Estimators When the Disturbances are small", Econometrica, PP.723-738.

11) Rao, C.R., (1973), Linear Statistical Inference and Its Applications", John Wiley & Sone, New York.

12) Sawa, T., 1972, Finite Sample Properties of K-Class Estimators", Econometrica, 1972, PP. 653-680.

13) Sclove, S.L. (1968), "Improved Estimators for Coefficients in Linear Regression, Journal of the American Statistical Association, PP. 599-606 .

14) Srivastava, V.K., 1970, "The Efficiency of Estimating Unrelated Regression Equations, Annals of the Institute of Statistical Mathematics PP. 483-493.

15) Stein, C (1956), "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution", Proceeding of the Third Berkeley Symposium on Mathematical Statistics and Probability, PP. 197-206.

16) Theil, H., 1961, "Economic Forecasts and Policy", North-Holland Publishing Co., Amsterdam.

17) Theil, H., 1966, "Applied Economic Forecasting", North-Holland Publishing Co., Amsterdam.

18) Ullah, A. and Ullah, S., 1978,"Double K-Class estimators of Coefficients in Linear Regression", Econometrica, PP. 705-722.

19) Vinod, H.D., and Ullah, A., 1981, "Recent Advances in Regression Methods", Marcel Dekker, Inc., New York.

20) Vinod, H.D., Ullah, A. and Kadiyala, K., 1981, "Evaluation of the Mean Squared Error of Certain Generalized Ridge Estimators Using Confluent Hypergeometric Functions", Sankhya.

21) Zellner, A., and W. Vandaele (1975), "Bayes-Stein Estimators for K-Means, Regression, and Simultaneous Equations Models", in Studies in Bayesian Econometrics and Statistics, ed. by S.E. Fienberg and A. Zellner. Amsterdam: North-Holland Publishing Co., PP. 627-653.