1981

Lindley and Smith Type Improved Estimators of Regression Coefficients

Aman Ullah

ISSN: 0318-725X
ISBN: 0-7714-0285-6

RESEARCH REPORT 8113

LINDLEY AND SMITH TYPE IMPROVED ESTIMATORS OF REGRESSION COEFFICIENTS

by

Aman Ullah*

Abstract

In this paper we propose a class of Lindley and Smith-type operational estimators for the coefficients of linear regression. This class of estimators contains Shiller and Bayes Almon estimators, and also various Stein-type and operational Ridge-type estimators. The sampling properties of Lindley and Smith-type estimators are studied under the weighted mean squared error and matrix mean squared error criteria. The iterative version of these estimators is also analyzed.

*This paper was written at C.O.R.E., Belgium during my sabbatical leave from the University of Western Ontario. In preparing this paper I was benefitted by the comments from D. V. Lindley, J. Drèze, T. Kloek, J. Kmenta, V. K. Srivastava and M. Mouchart. Research support from Canada Council and NSERC is gratefully acknowledged. I am also thankful for the excellent research facilities provided at C.O.R.E.

August, 1981

Lindley and Smith Type Improved Estimators of Regression Coefficients

1. INTRODUCTION

In this paper we consider the general Bayesian linear regression model using a hierarchical form of prior structure due to Lindley (1971) and Lindley and Smith (1972). A two-stage hierarchy of priors is considered in which the first stage prior describes the relationship between parameters in the linear regression and the second stage prior describes the knowledge about this form of relationship.[1] In applied econometrics, this type of model has recently been used in a multi-regression context by Trivedi (1980) and in a single equation case by Ullah and Raj (1980) [also see Maddala (1977, Chapter 16)].

The plan of this paper is as follows. In Section 2 we propose a class of Lindley and Smith (L&S)-type biased estimators for the coefficients of linear regression. This class of estimators does not shrink the ordinary least squares estimator towards the arbitrary point zero.[2] The Zellner and Vandaele (1975) estimator, and the much used Shiller (1973) and Bayes-Almon estimators, are shown to be its special cases. The class of L&S-type estimators also contains various Stein-type and adaptive Ridge-type estimators. In Section 3, we analyze the sampling properties of the L&S-type estimators under the weighted mean squared error and matrix mean squared error criteria.

The iterative version of the L&S-type estimators is analyzed in Section 4. Finally, Section 5 gives the proofs of the results in Sections 3 and 4 .

2. THE MODEL AND ESTIMATORS

Consider a linear regression model :

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

where y is a $T \times 1$ vector of observations on the variable to be explained, X is a $T \times p$ matrix of observations on the explanatory variables, β is a $p \times 1$ vector of unknown parameters and u is a $T \times 1$ vector of disturbances. Let the disturbance vector u be distributed as multivariate normal with mean vector zero and covariance matrix $\sigma^2 I$, *i.e.* $u \sim N(0, \sigma^2 I)$.

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

$$(2.2) \qquad b = (X'X)^{-1}X'y$$

which is unbiased. Its matrix mean squared error Mtx MSE and the weighted MSE (WMSE) are given, respectively, as[3]

$$(2.3) \qquad \text{Mtx MSE}(b) = E(b-\beta)(b-\beta)' = \sigma^2(X'X)^{-1}$$

and

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

where Q is a known positive definite matrix, and "tr" represents the trace of the matrix.

We note from (2.1) that, given β ,

(2.5) $$y \sim N(X\beta, \sigma^2 I).$$

Now we assume the prior about β , given β_o , as

(2.6) $$\beta \sim N(R\beta_o, \sigma_o^2 A^{-1})$$

where A is a known $p \times p$ non-singular matrix, R is a $p \times r$ known matrix and β_o is an $r \times 1$ hyperparameters vector. Further, let us suppose a diffuse prior distribution for β_o , or alternatively,

(2.7) $$\beta_o \sim N(R_1\beta_1, A_1^{-1}), \qquad A_1 \to 0$$

where β_1 is an $r_1 \times 1$ hyperparameters vector with R_1 , $r \times r_1$, known. Then the equations (2.5) to (2.7) represent the linear model with a hierarchical form of prior structure [for details see Lindley and Smith (1972)].

It has been shown by Lindley and Smith (1972, p. 7) that the posterior distribution of the regression parameter β , given y , X , R , R_1 , A , β_1 and $A_1 \to 0$, is multivariate normal with the posterior mean (Bayes estimator),

(2.8) $$\bar{\beta} = [X'X + kA(I - R(R'AR)^{-1}R'A)]^{-1}X'y; \quad k = \frac{\sigma^2}{\sigma_o^2} .$$

Further the posterior mean (Bayes estimator) of the hyperparameters β_o , given y , X , R , R_1 , A , β_1 and $A_1 \to 0$, is [4]

(2.9) $$\bar{\beta}_o = [R'(A^{-1} + k(X'X)^{-1})^{-1}R]^{-1}R'(A^{-1} + k(X'X)^{-1})^{-1}b,$$

[see Smith (1973, p. 69)]. Notice that both $\bar{\beta}$ and $\bar{\beta}_o$ are free from the parameters in the prior of β_o given in (2.7).

Using (2.9), an alternative form of (2.8) can be written as :

$$(2.10) \qquad \bar{\beta} = (X'X + kA^{-1})(X'Xb + kAR\bar{\beta}_o)$$
$$= R\bar{\beta}_o + [I + k(X'X)^{-1}A]^{-1}(b - R\bar{\beta}_o)$$

where the first equality on the right is from Smith (1973, p. 69). The second line of (2.10) shows that $\bar{\beta}$ can be viewed as a shrinkage estimator which shrinks the OLS estimator b towards $R\bar{\beta}_o$, the Bayes estimator of $R\beta_o$. Similarly, an alternative form of $\bar{\beta}_o$ can be written as $\bar{\beta}_o = (R'AR)^{-1}R'A\bar{\beta}$.

Notice that when $R = 0$, $\bar{\beta}$ in (2.10) reduces to the ridge estimator (for $A = I$) which shrinks the OLS towards zero. Thus, even though there is an apparent similarity between the Bayes estimator $\bar{\beta}$ in (2.8) and the ridge estimator, it is clear from (2.10) that they differ with respect to the point of shrinkage.[5]

2.1. A Class of Lindley and Smith (L & S)-type Biased Estimators.

Regarding k in (2.8) L & S suggested an iterative procedure which essentially consists of starting with the OLS estimate of β $(k = 0)$ and then taking the estimates of σ^2 and σ_o^2 , respectively, as $(y - Xb)'(y - Xb)/n_1$ and $(b - Rb_o)'A(b - Rb_o)/n_2$ where $b_o = (R'AR)^{-1}R'Ab = \bar{\beta}_o$ for $k = 0$, and n_1 and n_2 are scalar numbers. Using these estimates of σ^2 and σ_o^2 we can formulate k as[6]

$$(2.11) \qquad \hat{k} = \frac{hs^2}{b'Bb} \quad ; \qquad s^2 = (y - Xb)'(y - Xb)$$

where h is an arbitrary scalar and

(2.12) $B = AJ = J'AJ = J'A; \; J = J^2 = I - R(R'AR)^{-1}R'A.$

Substituting $k = \hat{k}$ in (2.8) we can write the following class of L&S-type estimators

(2.13) $$\hat{\beta} = [I + \hat{k}DJ]^{-1}b = \left[I + \frac{hs^2}{b'Bb} DJ\right]^{-1} b$$

where $\hat{k}$, B and J are as given in (2.11) and (2.12), and [7]

(2.14) $D = (X'X)^{-1}A.$

An iterative version of this estimator is considered in Section 4.

An alternative form of $\hat{\beta}$, corresponding to (2.10), can be written as

(2.15) $$\hat{\beta} = (X'X + \hat{k}A)^{-1}(X'Xb + \hat{k}AR\hat{\beta}_o) = R\hat{\beta}_o + [I + \hat{k}D]^{-1}(b - R\hat{\beta}_o)$$

where from (2.9)

(2.16) $\hat{\beta}_o = [R'(A^{-1} + \hat{k}(X'X)^{-1})R]^{-1}R'(A^{-1} + \hat{k}(X'X)^{-1})^{-1}b.$

We can also write (2.16) as $\hat{\beta}_o = (R'AR)^{-1}R'A\hat{\beta}$. The L&S-type estimators $\hat{\beta}$, in the form (2.15), are weighted matrix combinations of the estimators b and $R\hat{\beta}_o$ and they shrink the OLS estimator towards $R\hat{\beta}_o$.

It is interesting to note the following members of the L&S estimators $\hat{\beta}$ in (2.13) or (2.15).

Lindley-Smith Estimator Under Exchangeable Prior : Considering $R = \iota$ and $A = I$ in (2.6) we get $B = J = I - \iota(\iota'\iota)^{-1}\iota'$ and $D = (X'X)^{-1}$ from

(2.12) and (2.14), where ι is a $p \times 1$ vector of unit elements. Substituting these values of B, J and D in (2.13) we can write :

$$(2.17) \qquad b_1 = \left[I + \frac{hs^2}{b'Jb}(X'X)^{-1}J\right]^{-1} b.$$

This is the first stage in the Lindley-Smith iterative estimator (1972, p.17) under an exchangeable prior.

Zellner-Vandaele Estimator : Choosing $R = \iota$ and $A = X'X = Q$ in (2.6) we get $B = QJ$; $J = I - \iota(\iota' Q\iota)^{-1}\iota' Q$ and $D = I$ from (2.12) and (2.14). Substituting these choices of B, J and D in (2.13) and (2.15) we get :

$$(2.18) \qquad b_2 = [I + \hat{k}J]^{-1}b = \iota\bar{b} + (1+\hat{k})^{-1}(b - \iota\bar{b})$$

where $\bar{b} = (\iota' X'X\iota)^{-1}\iota' X'Xb$ and $\hat{k} = hs^2/b'QJb$. This is the Zellner and Vandaele (1975, pp. 329-30) type estimator, which they obtained by a non-Bayesian method.

Bayes Almon Estimator : Taking R to be a $p \times r$ Almon transformation matrix of rank r and $A = I$ we get $B = J = I - R(R'R)^{-1}R'$ and $D = (X'X)^{-1}$. Substituting these in (2.13) we can write the Bayes-Almon-type estimator for the distributed lag model [see Maddala (1977) and Ullah and Raj (1980)].[8]

Stein and Ridge-type Estimators : When $R = 0$ and $A = I$ in (2.6) we have $J = B = I$ from (2.12) and $D = (X'X)^{-1}$ from (2.14). In this case $\hat{\beta}$ in (2.13) become the adaptive Ridge-type estimators [see Alam and Hawkes (1978) and the references therein] which shrink the OLS estimator towards zero. Further, for $J = I$ and $D = I$, $\hat{\beta}$ is Stein-type estimators, see

James and Stein (1961), Judge and Bock (1978) and Ullah and Ullah (1978) among others.

3. SAMPLING PROPERTIES OF THE ESTIMATOR $\hat{\beta}$

Assuming disturbances to be small, we present the small-disturbances approximations for the bias, MtxMSE and WMSE of the estimator $\hat{\beta}$. These are derived in Section 5. The properties of $\hat{\beta}$ compared to the OLS estimator b are then analyzed under the criteria of WMSE and MtxMSE.

The bias and MtxMSE , respectively, are

$$(3.1) \qquad E(\hat{\beta}-\beta) = -\frac{h\sigma^2 n}{\beta' B\beta} DJ\beta$$

and

$$(3.2) \qquad \text{MtxMSE}(\hat{\beta}) = \sigma^2 (X'X)^{-1} + \frac{\sigma^4 hn(n+2)}{\beta' B\beta}\left[\frac{h}{\beta' B\beta} DJ\beta\beta' J'D' - \frac{2}{n+2}\left\{K - \frac{L}{\beta' B\beta}\right\}\right]$$

where B and D are as given in (2.12) and (2.14), respectively, and

$$(3.3) \qquad n = T - p, \quad K = \frac{(X'X)^{-1}J'D' + DJ(X'X)^{-1}}{2},$$

$$L = DJ\beta\beta' B'(X'X)^{-1} + (X'X)^{-1}B\beta\beta' J'D'.$$

Further, since $E(\hat{\beta}-\beta)'Q(\hat{\beta}-\beta) = \operatorname{tr} E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'Q$ we get [9]

$$(3.4) \qquad \text{WMSE}(\hat{\beta}) = \sigma^2 \operatorname{tr}(X'X)^{-1}Q + \sigma^4 hn(n+2)\frac{\beta' J'D'QDJ\beta}{(\beta' B\beta)^2}\,.$$

$$\left[h - \frac{2}{n+2}\,\frac{\beta' B\beta}{\beta' J'D'QDJ\beta}\left\{\operatorname{tr} DJ(X'X)^{-1}Q - 2\,\frac{\beta' B'(X'X)^{-1}QDJ\beta}{\beta' B\beta}\right\}\right].$$

When J = I = D , the result in (3.4) compares with Ullah and Ullah (1978, p. 710).

We now compare $\hat{\beta}$ with b under (i) the WMSE and (ii) the MtxMSE criteria.

(i) WMSE Criterion

Before obtaining the main result we note that $J = J^2$ in (2.12) is an idempotent matrix of rank $\ell = p - r$. Thus it can be written as $J = GG'$ where G is a $p \times \ell$ matrix of ℓ orthonormal vectors such that $G'G = I$. Next, from Rao (1973, p. 74) we observe that if E is any $p \times p$ symmetric matrix and F is any $p \times p$ positive difinite matrix, then

$\lambda_p^* = \min_\beta \left(\frac{\beta'E\beta}{\beta'F\beta} \right)$ and $\lambda_1^* = \max_\beta \left(\frac{\beta'E\beta}{\beta'F\beta} \right)$ where λ_p^* and λ_1^* are the minimum and maximum eigenvalues, respectively, of EF^{-1}. From this it follows that $\min_\beta \left(\frac{\beta'J'EJ\beta}{\beta'J'FJ\beta} \right) = \lambda_\ell$ and $\max_\beta \left(\frac{\beta'J'EJ\beta}{\beta'J'FJ\beta} \right) = \lambda_1$ where $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_\ell$ are the eigenvalues of $(G'EG)(G'FG)^{-1}$ and we use $J = GG'$. Notice that λ_ℓ is the minimum and λ_1 is the maximum eigenvalue. Using this result and recalling (2.4) it can easily be verified that $\hat{\beta}$ has smaller WMSE than b , *i.e.*

$$(3.5) \qquad WMSE(\hat{\beta}) - WMSE(b) \leq 0$$

when

$$(3.6) \qquad 0 < h \leq \frac{2\lambda_\ell \mu_1}{n+2} (d-2); \; d = \frac{1}{\mu_1} \operatorname{tr} DJ(X'X)^{-1}Q = \frac{1}{\mu_1} \operatorname{tr} G'(X'X)^{-1}QDG > 2$$

where λ_ℓ and λ_1 are the minimum and maximum eigenvalues of $(G'AG)(G'D'QDG)^{-1}$ and $(G'AG)^{-1}(G'A(X'X)^{-1}QDG)$, respectively.

The result in (3.6) has been obtained for any $p \times p$ matrix D . For the choice of $D = (X'X)^{-1}A$ in (2.14), the condition (3.6) reduces to

$$(3.6)\text{a} \qquad 0 < h \leqslant \frac{2}{n+2}(d-2); \quad d = \frac{1}{\mu_1} \operatorname{tr} G'(X'X)^{-1}Q(X'X)^{-1}AG > 2.$$

The following observations about the dominance condition of $\hat{\beta}$ over b , given in (3.6) and (3.6)a, can now be made :

(i) It was noted earlier that for $J = I$, $\hat{\beta}$ is a subclass of estimators which shrink the OLS estimator towards zero. Since $J = I$ implies $G = I$, a $p \times p$ identity matrix, the dominance condition of these estimators, for any D , is

$$0 < h \leqslant \frac{2\lambda_p \mu_1}{n+2} (d_* - 2); \quad d_* = \frac{1}{\mu_1} \operatorname{tr} D(X'X)^{-1}Q > 2,$$

where λ_p and μ_1 are the minimum and maximum eigenvalues of $A(D'QD)^{-1}$ and $(X'X)^{-1}QD$, respectively.[10] When $D = (X'X)^{-1}A$ we get $0 < h \leqslant \frac{2}{n+2}(d_* - 2)$. Since $\operatorname{tr} JD(X'X)^{-1}Q \leqslant \operatorname{tr} D(X'X)^{-1}Q$, the range of h will be smaller for $J \neq I$ compared to $J = I$.

(ii) The dominance condition for the special cases of $\hat{\beta}$, *viz.* Zellner-Vandaele estimator, Bayes-Almon estimator and others given in Section 2.1, can be obtained by direct substitutions of the respective choices of A and R in (3.6) or (3.6)a. Note that the matrix G will be different for different choices of A and R .

(iii) It is clear from (3.6) that the range of h will, in general, depend upon the eigenvalues. However, if we choose D , A and Q such that

(3.7) $\quad D'QD = A \quad$ and $\quad (X'X)^{-1}QD = I,$

then we get estimators $\hat{\beta}$ from (2.13) whose dominance condition from (3.6) will be

(3.8) $\quad 0 < h \leq \frac{2}{n+2}(p-r-2); \; p > r+2,$

when $D = (X'X)^{-1}A$ as in (2.14), then the above condition follows from (3.6)a for the choice of Q and A for which $(X'X)^{-1}Q(X'X)^{-1}A = I$. Notice that the condition (3.8) is free from the eigenvalues and this condition can easily be verified in practice. The condition (3.8) implies that the estimators with D, A and Q or with A and Q (for $D = (X'X)^{-1}A$) satisfying (3.7) dominate the OLS estimator if the number of regressors are more than $r+2$, where r is the number of columns in the matrix R.

To see the estimators which satisfy (3.7) and (3.8), we consider the case of $D = (X'X)^{-1}A$. For this D and any arbitrary choice of Q in the WMSE, (3.7) gives $A = X'XQ^{-1}X'X$. Substituting these in (2.13), we get the following estimators :

(3.9) $$b_3 = \left[I + \frac{hs^2}{b'X'XQ^{-1}X'XJb} \; Q^{-1}X'XJ\right]^{-1} b$$

whose dominance condition is given by (3.8). Similarly, for a given choice of A in (2.6), (3.7) gives $Q = X'XD^{-1} = X'XA^{-1}X'X$. Thus the estimators $\hat{\beta}$ in (2.13), for a given A, dominate the OLS estimator b under the range (3.8) when we consider the WMSE $= E\,[(\hat{\beta}-\beta)'X'XA^{-1}X'X(\hat{\beta}-\beta)]$.

(iv) In the case when we consider $Q = (X'X)^{q}$ in (3.9) we can write it as :

$$(3.10) \qquad b_4 = \left[I + \frac{hs^2}{b'(X'X)^{2-q}Jb} (X'X)^{1-q}J \right]^{-1} b$$

where $J = I - R[R'(X'X)^{2-q}R]^{-1}R'(X'X)^{2-q}$ and q is any arbitrary number. This set of estimators dominates over b under the condition (3.8).

(ii) MtxMSE Criterion

According to the MtxMSE criterion, the estimator $\hat{\beta}$ is better than or superior to b if

$$(3.11) \qquad \text{MtxMSE}(b) - \text{MtxMSE}(\hat{\beta}) = \Delta \geqslant 0$$

where $\Delta \geqslant 0$ implies that Δ is a non-negative definite matrix.[11]

From (3.2) we note that

$$(3.12) \qquad \Delta = \frac{\sigma^4 hn(n+2)}{\beta'B\beta} \left[\frac{h}{\beta'B\beta} DJ\beta\beta'J'D' - \frac{2}{n+2} \left\{ K - \frac{L}{\beta'B\beta} \right\} \right]$$

where B and J are given in (2.12), D in (2.14), and n , K and L are in (3.3). Thus to see whether Δ is non-negative definite or not we need to show that the scalar

$$(3.13) \qquad \eta'\Delta\eta = -\sigma^4 hn(n+2) \frac{\eta'DJ\beta\beta'J'D\eta}{(\beta'B\beta)^2} \left[h - \frac{2}{n+2} (\varphi - 2) \right] \geqslant 0$$

for all $p \times 1$ vector $\eta \neq 0$, where using (2.12) and (2.14)

$$(3.14) \qquad \varphi = \frac{\eta'(X'X)^{-1}J'D'\eta}{\eta'DJ\beta\beta'J'D\eta} \beta'B\beta = \frac{(\delta'B\delta)(\beta'B\beta)}{(\delta'B\beta)^2} \; ; \quad \delta = (X'X)^{-1}\eta.$$

It is clear from (3.13) that $\eta'\Delta\eta = \delta'X'X\Delta X'X\delta \geqslant 0$ for all $\delta \neq 0$ if

$$(3.15) \qquad 0 < h \leqslant \frac{2}{n+2} (\varphi - 2) \quad \text{and} \quad \varphi > 2.$$

Now using Cauchy-Schwarz inequality [Rao (1973), p. 54)] we note that $\varphi \geq 1$. This implies that for some parameter space of β the condition $\varphi > 2$ can be satisfied. Thus the estimator $\hat{\beta}$ is a better estimator that b, under the MtxMSE criterion, for the range of h in (3.15) and in the parameter space of β which satisfies $\varphi > 2$.

4. ITERATIVE ESTIMATOR AND ITS PROPERTIES

Let us call the estimator $\hat{\beta}$ in (2.13) as the first step estimator and rewrite it as

$$\hat{\beta} = \bar{\beta}^{(1)} = [I + k^{(o)}DJ]^{-1}b \tag{4.1}$$

where from (2.11)

$$k^{(o)} = \frac{h(y - X\bar{\beta}^{(o)})'(y - X\bar{\beta}^{(o)})}{\bar{\beta}^{(o)\prime}B\bar{\beta}^{(o)}} = \frac{hs^{2(o)}}{\bar{\beta}^{(o)\prime}B\bar{\beta}^{(o)}} = \hat{k} \tag{4.2}$$

and $\bar{\beta}^{(o)} = b$ is the initial estimator $\bar{\beta}$ in (2.8) for $k = 0$. The estimator $\bar{\beta}^{(1)}$ can then be employed to provide another estimator of k and the process can be continued.

Suppose $\bar{\beta}^{(m+1)}$ denotes the estimator of β at the $(m+1)$-th iteration. Then we have :

$$\bar{\beta}^{(m+1)} = [I + k^{(m)}DJ]^{-1}b, \qquad k^{(m)} = \frac{hs^{2(m)}}{\bar{\beta}^{(m)\prime}B\bar{\beta}^{(m)}} \tag{4.3}$$

where for $m = 0,1,\ldots$

$$\begin{aligned} s^{2(m)} &= s^{2(o)} + (b - \bar{\beta}^{(m)})'X'X(b - \bar{\beta}^{(m)}) \\ \bar{\beta}^{(m)\prime}B\bar{\beta}^{(m)} &= b'Bb + (b - \bar{\beta}^{(m)})'B(b - \bar{\beta}^{(m)}) - 2b'B(b - \bar{\beta}^{(m)}). \end{aligned} \tag{4.4}$$

When $A = I$ so that $D = (X'X)^{-1}$, and $R = \iota$ such that $B = J = I - \iota(\iota'\iota)^{-1}\iota'$ then (4.3) is essentially the Lindley and Smith (1972, p.17) iterative modal estimation under the diffuse prior about σ^2 and independent, inverse $-\chi^2$ prior about σ_o^2 (also see footnote 6).

It has been hsown in the following section that upto the order of approximation considered

$$(4.5)\qquad \begin{aligned} &\text{WMSE}(\bar{\beta}^{(m+1)}) - \text{WMSE}(\hat{\beta}) = 0 \\ &\text{MtxMSE}(\bar{\beta}^{(m+1)}) - \text{MtxMSE}(\hat{\beta}) = 0. \end{aligned}$$

This result shows that the efficiency of the iterative estimator is the same as that of the first stage estimator $\hat{\beta} = \bar{\beta}^{(1)}$ in (4.1) and analysed in Section 2.1. Further the estimator $\bar{\beta}^{(m+1)}$ dominates the OLS estimator b under the range of h as given in (3.6).

An alternative iterative estimator can be considered as

$$(4.6)\qquad \tilde{\tilde{\beta}} = \bar{\beta}^{(m+1)} = [I + k^{(m)}DJ]^{-1}\bar{\beta}^{(m)}$$

where $k^{(m)}$ is as given in (4.3). This amounts to substituting $b = \bar{\beta}^{(o)}$ on the right of (4.1) and then changing it in each iteration. In a special case where $D = I$ and $J = I$, $\tilde{\tilde{\beta}}$ becomes the Vinod(1976)-type iterative estimator.

It has been shown in Section 5 that, upto the order of approximation considered, the bias, MtxMSE and WMSE of the iterative estimator $\tilde{\tilde{\beta}}$ is the same as in (3.1), (3.2) and (3.4), respectively, with h replaced by $(m+1)h$. Replacing h by $(m+1)h$ in (3.4), it can then easily be verified that, for any choice of D , (4.7) $\quad \mathrm{WMSE}(\bar{\beta}^{(m+1)}) - \mathrm{WMSE}(b) \leqslant 0 \qquad$ for $0 < h \leqslant \frac{2\lambda_{\ell}\mu_1}{(m+1)(n+2)}(d-2)$

and

(4.8) $\quad \mathrm{WMSE}(\bar{\beta}^{(m)}) - \mathrm{WMSE}(\bar{\beta}^{(m+1)}) \geqslant 0 \qquad$ for $0 < h \leqslant \frac{2\lambda_{\ell}\mu_1}{(2m+1)(n+2)}(d-2)$

where $\bar{\beta}^{(m+1)} = \tilde{\tilde{\beta}}$ as in (4.6), and λ_{ℓ} , μ_1 and d are as appearing in (3.6). When $D = (X'X)^{-1}A$, $\lambda_{\ell}\mu_1$ in (4.7) and (4.8) will become unity.

The result in (4.7) gives the condition under which the iterative estimator $\tilde{\tilde{\beta}} = \bar{\beta}^{(m+1)}$ dominates the OLS estimator, whereas (4.8) provides the condition under which the (m+1)th iteration will reduce the WMSE. These conditions suggest that the estimator $\tilde{\tilde{\beta}} = \bar{\beta}^{(m+1)}$ with $0 < h < \frac{2\lambda_{\ell}\mu_1}{(2m+1)(n+2)}(d-2)$ dominates the OLS estimator as well as $\bar{\beta}^{(m)}$. Thus, for this range of h it is worth going for iteration. This result is different than one obtained above for the iterative estimator in (4.3).

5. PROOF OF RESULTS IN SECTIONS 3 AND 4.

Let us write the model (2.1) as

(5.1) $\quad y = X\beta + \theta v$

where θ is a scalar assumed to tend to zero [see Kadane (1971)] and v is a $T \times 1$ random vector follwoing a multivariate normal distribution $N(0, \Psi I)$ with Ψ unknown. Notice that $\sigma^2 = \theta^2\Psi$.

It is easy to see that :

$$b - \beta = \theta(X'X)^{-1}X'v = \theta e \text{ (say)}$$

(5.2) $$Jb = J\beta + \theta Je$$

$$y - Xb = \theta Mv$$

where J is as given in (2.12) and $M = I - X(X'X)^{-1}X'$.

Now using (5.2), we get from (2.13)

(5.3) $$\begin{aligned}\hat{\beta} - \beta &= \left(I + \frac{h\theta^2 v'Mv}{b'Bb} DJ\right)^{-1} b - \beta \\ &= b - \beta - \frac{h\theta^2 v'Mv}{b'Bb} DJb + \\ &= b - \beta - \frac{h\theta^2 v'Mv}{\beta'B\beta}\left[1 + \frac{2\theta e'B\beta + \theta^2 e'Be}{\beta'B\beta}\right]^{-1}(DJ\beta + \theta DJe) + \cdots \\ &= b - \beta - \frac{h\theta^2 v'Mv}{\beta'B\beta}\left[1 - \frac{2\theta e'B\beta}{\beta'B\beta} + O(\theta^2)\right](DJ\beta + \theta DJe) + \cdots \\ &= \theta\xi_1 + \theta^2\xi_2 + \theta^3\xi_3 + O(\theta^4)\end{aligned}$$

where

(5.4) $$\begin{aligned}\xi_1 &= e\ , \qquad \xi_2 = -\frac{hv'Mv}{\beta'B\beta} DJ\beta \\ \xi_3 &= \frac{hv'Mv}{\beta'B\beta}\left[\frac{2e'B\beta}{\beta'B\beta} DJ\beta - DJe\right].\end{aligned}$$

Thus, the bias upto order $O(\theta^2)$ and the MtxMSE upto order $O(\theta^4)$ of the estimator $\hat{\beta}$ are

(5.5) $$E(\hat{\beta} - \beta) = \theta E(\xi_1) + \theta^2 E(\xi_2)$$

and

(5.6) $$E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \theta^2 E(\xi_1\xi_1') + \theta^3 E(\xi_2\xi_1' + \xi_1\xi_2') + \theta^4 E(\xi_2\xi_2' + \xi_3\xi_1' + \xi_1\xi_3').$$

Now using normality of v it is easy to verify that :

$$(5.7) \qquad E(\xi_1) = 0 \qquad\qquad E(\xi_2) = -\frac{h\Psi(T-p)}{\beta'B\beta} DJ\beta$$

$$(5.8) \qquad E(\xi_1\xi_1') = \Psi(X'X)^{-1}, \; E\xi_2\xi_1' = E\xi_1\xi_2' = 0$$

$$(5.9) \qquad E(\xi_2\xi_2') = \frac{h^2\Psi^2(T-p)(T-p+2)}{(\beta'B\beta)^2} DJ\beta\beta'J'D'$$

$$(5.10) \qquad E(\xi_3\xi_1') = \frac{h}{\beta'B\beta}\left[\frac{2DJ\beta\beta'B'(X'X)^{-1}X'E(v'Mv\cdot vv')X(X'X)^{-1}}{\beta'B\beta} - \right.$$

$$\left. - DJ(X'X)^{-1}X'E(v'Mv\cdot vv')X(X'X)^{-1}\right]$$

$$= \frac{h\Psi^2(T-p)}{\beta'B\beta}\left[\frac{2DJ\beta\beta'B'}{\beta'B\beta} - DJ\right](X'X)^{-1}$$

$$= \frac{h\Psi^2(T-p)}{\beta'B\beta} DJ\left[2\frac{\beta\beta'B'}{\beta'B\beta} - I\right](X'X)^{-1}$$

where we have used the results $E(v'Mv\cdot vv') = \Psi^2\,[\,\mathrm{tr}\,M + 2M]$ and $X'M = 0$. Further note that $E(\xi_1\xi_3')$ is the transpose of $E(\xi_3\xi_1')$.

The results for the iterative estimator $\tilde{\beta} = \bar{\beta}^{(m+1)}$ in (4.6) can be derived in the same way as done in the case of $\hat{\beta}$. In this context we first note from (5.3) that $b - \hat{\beta} = b - \bar{\beta}^{(1)}$ is at least of order θ^2. Thus from (4.4) :

$$(5.11) \qquad \begin{aligned} s^{2(m)} &= s^{2(o)} + 0(\theta^4) \\ \bar{\beta}^{(m)'}B\bar{\beta} &= b'Bb + 0(\theta^2) \end{aligned}$$

where $0(\theta^2)$ and $0(\theta^4)$ represent terms of at least orders θ^2 and θ^4 respectively.

Using (5.11), it can then be verified that

$$(5.12) \qquad \tilde{\beta} - \beta = \theta\xi_1^{(m+1)} + \theta^2\xi_2^{(m+1)} + \theta^3\xi_3^{(m+1)}$$

(5.13) $\xi_1^{(m+1)} = \xi_1$; $\xi_2^{(m+1)} = -h(m+1)\frac{v'Mv}{\beta'B\beta} DJ\beta$

and

(5.14) $\xi_3^{(m+1)} = h(m+1)\frac{v'Mv}{\beta'B\beta}\left[\frac{2e'B\beta}{\beta'B\beta} DJ\beta - DJe\right].$

Notice that $\xi_1^{(m+1)} = \xi_1$, and $\xi_2^{(m+1)}$ and $\xi_3^{(m+1)}$ are the same as ξ_2 and ξ_3 in (3.4) with h replaced by h(m+1). Thus using the expectations in (5.7) to (5.10), the results for the bias, MtxMSE and WMSE of $\tilde{\tilde{\beta}}$ can be found to be the same as those for $\hat{\beta}$ in Section 3, with h replaced by (m+1)h.

For the iterative estimator in (4.3), it can be similarly shown by using (5.11) that $\bar{\beta}^{(m+1)} - \beta = \theta\xi_1 + \theta^2\xi_2 + \theta^3\xi_3 = \hat{\beta} - \beta$, upto the order of approximation considered. Therefore, the result in (4.5) is obvious.

FOOTNOTES

[1]When only the first stage prior is used, one gets the Bayesian regression model as considered by Raiffa and Schlaifer (1961), Zellner (1971) and Box and Tiao (1973), among others.

[2]See Lindley (1962) for the problem of arbitrary origin in Stein's estimator.

[3]The MtxMSE of an estimator $\hat{\beta}$ of β is the second order moment matrix around β, *i.e.*, $\text{MtxMSE}(\hat{\beta}) = E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$. The sum of its diagonal elements is the total mean squared error which is $\text{MSE}(\hat{\beta}) = E(\hat{\beta}-\beta)(\hat{\beta}-\beta) = \text{tr}\,\text{MtxMSE}(\hat{\beta})$. The $\text{WMSE}(\hat{\beta}) = E(\hat{\beta}-\beta)'Q(\hat{\beta}-\beta) = \text{tr}\,\text{MtxMSE}(\hat{\beta})Q$.

[4]From (2.5) and (2.6) $y \sim N(XR\beta_o, \sigma^2 + \sigma_o^2 XA^{-1}X')$. Thus $\bar{\beta}_o$ is the generalised least squared estimator of β_o in this model. The equivalence with (2.9) follows by using the matrix inversion of $\sigma^2 + \sigma_o^2 XA^{-1}X'$.

[5]If we consider only the first stage prior in (2.6) then, for given β_o, the posterior mean of β is $\beta_* = R\beta_o + [I + k(X'X)^{-1}A]^{-1}(b - R\beta_o)$, see *e.g.* Zellner (1971) and Giles and Rayner (1979). Thus, in general $\bar{\beta}$ and β_* are different.

[6]One can consider $\hat{k} = hs^2/b'Bb + h_1$ where h_1 is a scalar constant. This would require replacing $\beta'B\beta$ by $\beta'B\beta + h_1$ in the WMSE expression of $\hat{\beta}$ in Section 3. However, the dominance condition on h remains the same for any $h_1 \geq 0$.

[7]The estimator $\hat{\beta}$ represents a general class of biased estimators for the arbitrary choices of D, J and B which are not constrained by (2.12) and (2.14). The results of this paper can be used for such arbitrary choices.

[8]The Shiller estimator can also be considered as a special case of $\hat{\beta}$ for $A = I$ and $J = H'H$ where H is a $(p-r) \times p$ matrix of rank $p-r$ defined in Shiller (1973, p. 777) which is orthogonal to the Almon transformation matrix R. However, Shiller's estimator may not be written in the form of (2.15) because $J = H'H \neq J^2$.

[9] The results (3.1), (3.2) and (3.4) can be used for various choices of Q, D, J and B.

[10] In the special cases of Stein-type $(A = X'X,\ Q = I)$ and Ridge-type $(A = I,\ Q = I)$ estimators the conditions on h reduce to those given in Judge and Bock (1978) and Ullah and Ullah (1978) for the Stein case, and in Alam and Hawkes (1978) and Ullah *et al.* (1981), among others, for the Ridge case.

[11] If we take $Q = \eta\eta'$, where η is a $p \times 1$ vector, then the criterion (3.11) is equivalent to $WMSE(\hat{\beta}) - WMSE(b) \leq 0$ for all $\eta \neq 0$. Also see Wallace (1972) and Giles and Rayner (1979).

REFERENCES

Alam, K and J.S. Hawkes, 1978, "Estimation of Regression Coefficients", *Scandinavian Journal of Statistics*, 169-172.

Box, G.E.P. and G.C. Tiao, 1973, *Bayesian Inference in Statistical Analysis*, Reading, Mass., Addison-Wesley Publishing Co.

Giles, D.E.A. and A.C. Rayner, 1979, "The Mean Squared Errors of the Maximum Likelihood and Natural-Conjugate Bayes Regression Estimator", *Journal of Econometrics 11*, 319-334.

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.

Kadane, J.B., 1971, "Comparison of k-Class Estimators when the Disturbances are Small", *Econometrica* , 723-738.

Lindley, D.V., 1962, "Discussion on Professor Stein's Paper", *Journal of the Royal Statistical Society*, Series B, (24), 285-287.

Lindley, D.V., 1971, *Bayesian Statistics : A Review*, Philadelphia, SIAM.

Lindley, D.V. and A.F.M. Smith, 1972, "Bayes Estimates for the Linear Model (with Discussion", *Journal of the Royal Statistical Society*, Series B, 1-41.

Maddala, G.S., 1977, *Econometrics*, McGraw Hill, New York.

Raiffa, H. and R. Schlaifer, 1961, *Applied Statistical Decision Theory*, Boston : Graduate School of Business Administration, Harvard University.

Rao, C.R., 1973, *Linear Statistical Inference and its Application*, John Wiley and Sons, New York.

Shiller, R.J., 1973, "A Distributed Lag Estimator Derived From Smoothness Priors", *Econometrica*, 11-19.

Smith, A.F.M., 1973, "A General Bayesian Linear Model", *Journal of the Royal Statistical Society*, Series B (35), 67-75.

Trivedi, P.K., 1980, "Small Samples and Collateral Information : An Application of the Hyperparameter Model", *Journal of Econometrics 12*, 301-318.

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. Kadiyala, 1981, "A Family of Improved Shrinkage Factors for the Ordinary Ridge Estimator", in *Selected Papers on Contemporary Econometric Problems* (editor : E.G. Charatsis), North-Holland, Netherlands, Forthcoming.

Ullah, A. and B. Raj, 1980, "A Polynomial Distributed Lag Model with Stochastic Coefficients and Priors", *Empirical Economics*, forthcoming.

Vinod, H.D., 1976, "Simulation and Extension of a Minimum MSE Estimator in Comparison with Stein's", *Technometrics 18*, 491-496.

Wallace, T.D., 1972, "Weaker Criteria and Tests for Linear Restrictions in Regression", *Econometrica 40*, 689-698.

Zellner, A., 1971, *An Introduction to Bayesian Inferences in Econometrics*, John Wiley and Sons, New York.

Zellner, A. and W. Vandaele, 1975, "Bayes-Stein Estimators for k-Means Regression and Simultaneous Equation Models", in : S.E. Fienberg and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics*, North-Holland Publishing Co., Amsterdam.