THE EGYPTIAN STATISTICAL JOURNAL.
ISSR, CAIRO UNIV., VOL.40, NO.1,1996

A MODIFICATION RULE ON SELECTION OF THE RIDGE PARAMETER

Sayed M. El-Sayed *Badr El-Dine E. Sofian*

Abstract

Many formulae have been sugessted to estimate the ridge parameters. Most of these formulae ignore the features of the model such as number of the explanatory variables, degree of illconditioning and the significancy of the model. In this article two formulae are proposed for selecting the ridge parameter K ; for ridge regression . A Monte Carlo study is conducted to compare the mean square errors of ridge regression under the new formulae and some other formulae. The numerical results of the simulation indicate that the performance of the new formulas does produce the smaller mean square errors.

Key Words : Illconditioning; Monte Carlo; MSE; Multicollinearity Ridge Parameter; Ridge regression.

1. Introduction

The regression model is

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{u} \qquad \ldots (1)$$

$\underset{\sim}{Y}$ is an (n×1) vector of observetion on a respons varaible, $\underset{\sim}{X}$ is an (n×p) matrix of observations on p explanatory varaibles, $\underset{\sim}{\beta}$ is the (p×1) vector of regression coefficients and $\underset{\sim}{u}$ is an (nx1) vector of unobservable errors satisfying $E(u)=0$, $E(uu')=\sigma^2 I$. It is assumed that $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ have been scaled so that $(\underset{\sim}{X}'\underset{\sim}{X})$ is the matrix and $(\underset{\sim}{X}'\underset{\sim}{Y})$ is a vector of correlation coefficients. Using the following canonical form

Let $\underset{\sim}{Q} = [q_1, q_2, q_3, \ldots, q_p]$ and $\Lambda = dig[\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_p]$ be such that $(\underset{\sim}{X}'\underset{\sim}{X}) = Q\Lambda Q'$ where q_j is the eigenvector associated with the eigenvalue λ_j. Without loss of generality, we can assume that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_p$. Further let $\underset{\sim}{Z} = \underset{\sim}{X}\,\underset{\sim}{Q}$ and $\underset{\sim}{\alpha} = Q'\beta$ so that $\underset{\sim}{X}\underset{\sim}{\beta} = \underset{\sim}{Z}\underset{\sim}{\alpha}$. Model (1) can be rewriten as

$$\underset{\sim}{Y} = \underset{\sim}{Z}\,\underset{\sim}{\alpha} + \underset{\sim}{u} \qquad \ldots (2)$$

Then the least squares estimate for $\underset{\sim}{\alpha}$ is given by

$$\hat{\underset{\sim}{\alpha}}(0) = \Lambda^{-1}\,\underset{\sim}{Q}'\,\underset{\sim}{X}'\,\underset{\sim}{Y} \qquad \ldots (3)$$

Further ,

$$\text{Cov}\ (\hat{\underset{\sim}{\alpha}}(0)) = \sigma^2\,\Lambda^{-1} \qquad \ldots (4)$$

and the mean square errors (MSE(0)) becomes

$$\text{MSE}(0) = \sigma^2 \sum_{j=1}^{p} \lambda_j^{-1} \qquad \ldots (5)$$

when λ_j is small, for some j $(1 \leq j \leq p)$, both the estimates and their variances are inflated ; see [(3) and (4)]. This is known as the multicollinearity problem. Section 1 is an introduction to the regression model and least squares estimate. Section 2 is an introduction to the ridge regression, while section 3 presents some formulae for selecting the ridge parameter . Section 4 contains the proposed formulae. The simulation description is given in section 5 . The results and the conclusion are presented in section 6 and 7.

2. Ridge Regression Model

To overcome the problem of multicollinearity in regression, Hoerl and Kennard (1970) suggested a class of estimators indexed by a parameter k ,k >0. Under this class, the estimator of model (2) is given by

$$\hat{\underset{\sim}{\alpha}}(K) = (\Lambda + KI)^{-1} \underset{\sim}{Q}' \underset{\sim}{X}' \underset{\sim}{Y} \qquad \ldots (6)$$

with

$$E(\hat{\underset{\sim}{\alpha}}(K)) = \Lambda(\Lambda + KI)^{-1}\alpha \qquad \ldots (7)$$

and

$$\mathrm{Cov}(\hat{\underset{\sim}{\alpha}}(k)) = \Lambda(\Lambda + KI)^{-2}\sigma^2 \qquad \ldots (8)$$

Let

$$\tau_1(k) = \sigma^2 \sum_{j=1}^{p} \lambda_j(\lambda_j + K)^{-2}$$

and

$$\tau_2(k) = K^2 \sum_{j=1}^{p} \alpha_j^2(\lambda_j + k)^{-2}$$

Horel and kennard (1970) showed that $\tau_1(k)$ is a continuous monotonically decreasing function of k, while $\tau_2(k)$ is continuous monotonically increasing function ofk ($\tau_2(k)$ is shown to approach $\beta'\beta$ as an upper limit). These properties show that is may be possible to allow a little bias and subtantially improve the MSE where the mean square errors is definde as

$$\mathrm{MSE}(\hat{\underset{\sim}{\alpha}}(K)) = \tau_1(k) + \tau_2(k)$$

$$\mathrm{MSE}(\hat{\underset{\sim}{\alpha}}(K)) = \sigma^2 \sum_{j=1}^{p} \lambda_j(\lambda_j + K)^{-2} + K^2 \sum_{j=1}^{p} \alpha_j^2(\lambda_j + k)^{-2} \qquad \ldots (9)$$

For $K > 0$ $\hat{\alpha}(K)$ is biased and the bias increase with K, [see equation (7)]. The idea of the ridge regression is to select a value for K such that the reduction in mean square errors is increasing by increasing the bias. Hoert and Kennard (1970) proved that there exists a value of K such that

$$\sigma^2 \sum_{j=1}^{p} \lambda_j(\lambda_j+K)^{-2} + K^2 \sum_{j=1}^{p} \alpha_j^2(\lambda_j+K)^{-2} \leq \sigma^2 \sum_{j=1}^{p} \lambda_j^{-1} \quad \ldots (10)$$

3. Some Formulae For Selecting The Ridge Parameter k

In this section we discuss some formulae for computing the value k which lead to improvement methods. McDonald and Galarneau (1975) ; noted that:

$$\hat{L} = \hat{\alpha}'(0)\hat{\alpha}(0) - \hat{\sigma}^2(0)\sum_{j=1}^{p} \lambda_j^{-1} \quad \ldots (11)$$

is an unbiased estimator of $\alpha'\alpha$. They suggested and studied the performance of the mean square errors to the following selection rule:

$$\begin{aligned} &\text{choose} \quad K \text{ such that } \hat{\alpha}'(k)\hat{\alpha}(k) = \hat{L} \quad \text{if } \hat{L} > 0; \\ &\text{choose} \quad k = K^* \quad ; \text{otherwise} \end{aligned} \quad \ldots (12)$$

they use $K^* = 0$ and $K^* = \infty$ which lead to $\hat{\alpha}(0)$ and 0, respectively.

Hoerl, Kennard and Baldwin (1975);(HKB); suggested that a reasenable choise of K is :

$$\hat{K}_{HKB} = p\,\hat{\sigma}^2(0) \,/\hat{\alpha}'(0)\hat{\alpha}(0) \quad \ldots (13)$$

Hoerl and Kennard (1970) developed an iteration scheme for calculating $\hat{K}_{HK}$ using $\hat{\underset{\sim}{\alpha}}'(K_t)\hat{\underset{\sim}{\alpha}}(k_t)$ instead of $\hat{\underset{\sim}{\alpha}}'(0)\hat{\underset{\sim}{\alpha}}(0)$ where;

$$\hat{K}_{HK}(t) = p\,\hat{\sigma}^2(0) \,/\, \hat{\underset{\sim}{\alpha}}'(K_t)\hat{\underset{\sim}{\alpha}}(K_t)$$

Lowless and Wang (1976) used a Bayesion argument to motivate the estimator, they suggested :

$$\hat{K}_{LW} = p\,\hat{\sigma}^2(0) \,/\, \hat{\underset{\sim}{\alpha}}'(0)\underset{\sim}{Z}'\underset{\sim}{Z}\,\hat{\underset{\sim}{\alpha}}(0)$$

an alternative method to overcome the difficulty that $\hat{\underset{\sim}{\alpha}}'(0)\hat{\underset{\sim}{\alpha}}(0)$ is an overestime to $\hat{\underset{\sim}{\alpha}}\underset{\sim}{\alpha}$, El-Bassiouni and El-Sayed (1986) introduced $\hat{K}_{BS1}$ and $\hat{K}_{BS2}$, which are defined as follow:

$$\text{choose} \begin{cases} \hat{K}_{BS1} = p\,\hat{\sigma}^2(0) \,/\, \hat{L} & \text{if } \hat{L} > 0 \\ \hat{K}_{BS1} = 0 & \text{if } \hat{L} \le 0 \end{cases} \qquad \ldots (14)$$

formula (14) means, we use the least square estimator if $\hat{L} \le 0$

$$\text{choose} \begin{cases} \hat{K}_{BS2} = p\,\hat{\sigma}^2(0) \,/\, \hat{L} & \text{if } \hat{L} > 0 \\ \hat{K}_{BS2} = \hat{K}_{HKB} & \text{if } \hat{L} \le 0 \end{cases} \qquad \ldots (15)$$

formula (15) means,we use the HKB estimator if $\hat{L} \le 0$.

4. The Improvement Selection Rule of K

Since formula (13) does not depend on the value of $\hat{L}$ defin in (11) ; we propose the new estimator denoted by $\hat{k}_1$ as combination between the selection rule defined in (12) and the estimator $\hat{K}_{HKB}$ defined in (13) where :

$$\text{choose} \begin{cases} \hat{K}_1 = p\,\hat{\sigma}^2(0) / \hat{\alpha}'(0)\alpha(0) & \text{if } \hat{L} > 0 \\ \hat{K}_1 = 0 & \text{if } \hat{L} \leq 0 \end{cases} \quad \ldots (16)$$

Let $\hat{K}_2 \equiv \hat{k}_{BS1}$ defined in (14). Both of $\hat{K}_1$ and $\hat{k}_2$ depend on the value of $\hat{L}$ defined in (11). The two estimators $\hat{K}_1$ and $\hat{K}_2$, di not consider the properties of the model under investigation as:

(1) The number of explanatory variables in the model p ;

(2) The degree of illconditioning in $(\underset{\sim}{X}'\underset{\sim}{X})$ matrix ;

(3) The value of σ^2.

The new selection rule for k developed in such a way that previous points are taken in consideration to determine the critical value of $\hat{L}$. If $\hat{L} > 0$, this implies that the range of $\hat{k}_1$ and $\hat{k}_2$ is ;

$$0 < \hat{K}_1 \text{ and } \hat{k}_2 < \frac{p}{\sum_{j=1}^{p} \lambda_j^{-1}}$$

so that the ranges of $\hat{k}_1$ and $\hat{k}_2$ become short or narrow when one or more of the eigen values are small. In such case the upper pound value $\frac{p}{\sum_{j=1}^{p} \lambda_j^{-1}} \to 0$ as $\lambda_j \to 0$ $(1 \le j \le p)$. It is possible to this range not contain the appropriate value of K which can reduce the mean square errors in (9). The new selection rule allows the values of $\hat{K}$ to have a wide range ; say $0< \hat{K} <c$ where c is positive constant value greater than $\frac{p}{\sum_{j=1}^{p} \lambda_j^{-1}}$. In this case the critical value of $\hat{L}$,say $\hat{L}_c$, becomes as follows:

$$\hat{L}_c = - \hat{\sigma}^2(0)[\sum_{j=1}^{p} \lambda_j^{-1} - p\, c^{-1}] \quad .... \ (17)$$

Put c = 1 in (17) then $0< \hat{K} <1$ and the critical point value of $\hat{L}_c$ will be reduced to

$$\hat{L}_1 = - \hat{\sigma}^2(0) \ [\sum_{j=1}^{p} \lambda_j^{-1} - p] \quad .... \ (18)$$

Notation (18) is a function at $\hat{\sigma}^2(0)$, λ_j and p . The value of $[\sum_{j=1}^{p} \lambda_j^{-1} - p]$ may also be used as an index of illconditioning factor (ICF). After the previous discsussion , we can introduce the following estimators denoted by $\hat{K}_3$ and $\hat{K}_4$ which depend on the new selection rule as follow :

choose
$$\begin{cases} \hat{K}_3 = p\,\hat{\sigma}^2(0) \,/\, \hat{\underset{\sim}{\alpha}}'(0)\hat{\underset{\sim}{\alpha}}(0) & \text{if} \quad \hat{L} > \hat{L}_1 \\ \hat{K}_3 = 0 & \text{if} \quad \hat{L} \le \hat{L}_1 \end{cases} \quad \ldots (1$$

while

choose
$$\begin{cases} \hat{K}_4 = p\,\hat{\sigma}^2(0) \,/\, \hat{L} & \text{if} \quad \hat{L} > 0 \\ \hat{K}_4 = p\,\hat{\sigma}^2(0) \,/\, \hat{\underset{\sim}{\alpha}}'(0)\hat{\underset{\sim}{\alpha}}(0) & \text{if} \quad \hat{L}_1 < \hat{L} \le 0 \\ \hat{K}_4 = 0 & \text{if} \quad \hat{L} < \hat{L}_1 \end{cases} \quad \ldots (20$$

The advantage behind the improvement selection rule is on makin the range of $\hat{k}$ larger than it was, and bounded by unity. Also we can compute $\hat{k}$ even if $\hat{L}$ less than zero up to $\hat{L}_1$.

5. The simulation Study

In this section we describe the simulation technique which was used to examine the performance of the modified ridge estima tores using, ($\hat{k}_3$ and $\hat{k}_4$) in contrast to the least square and the unmodified estomatores using, ($\hat{k}_1$ and $\hat{k}_2$) through MSE criteria.

5.1. Models Considered

We consider three basic models , these models are used by Lee and Compbell (1985);

(i) Four factor model (Hald , (1952)) : p = 4, n = 13 with eigenvalues

$$\Lambda = dig[\ 2.2357,\ 1.5761,\ 0.1866,\ 0.0016\ ]$$

and ICF = 618.3 .

(ii) Ten factor model (Gorman and Toman (1966)):p= 10, n= 36 with eigenvalues

$$\Lambda = dig[\ 3.6923,\ 1.5418,\ 1.2927,\ 1.0457,\ 0.9719,\ 0.6587,\ 0.3574,\ 0.2197,\ 0.1513,\ 0.0681\ ]$$

and ICF = 23.8 .

(iii) Fifteen factor model (McDonald and Schwing (1973): p = 15, n = 60 with

$$\Lambda = dig[\ 4.5272,\ 2.7547,\ 2.0545,\ 1.3487,\ 1.2277,\ 0.9605,\ 0.6124,\ 0.4729,\ 0.3708,\ 0.2163,\ 0.1665,\ 0.1275,\ 0.1142,\ 0.0460,\ 0.0049\ ]$$

and ICF = 248.2 . The above three sets of eigenvalues are taken from Lawless (1978).

5.2. Regression Coefficients and Orientations

The simulation methodologies of Hoerl, Kennard and Baldwin (1975) and Lawless and Wang (1976) were subject to strong criticizms [see Pagel (1981)] who showed that such methodologies are strongly biased in favor of ridge regression. To avoid the pitfalls in such methodologies we decided to follow the same pattern as those reported by Newhouse and Oman (1971), McDonald and Galarneau (1975), Gunst and Mason (1977), Wichern and Churchill (1978) and El-Bassiouni and El-Sayed(1986) who used different orientations of β to the eigen vectores of $(\underset{\sim}{X}'\underset{\sim}{X})$ matrix. Five differents orientations associated with the eigen vectores of $(\underset{\sim}{X}'\underset{\sim}{X})$ have been used. Let vector q_p is the eigen vector corresponding to the smallest eigen value and the vecto q_1 is the eigen vector corresponding to the largest eigen value and the vector $\underset{\sim}{\alpha} = \underset{\sim}{Q}' \underset{\sim}{\beta}$. Put $\underset{\sim}{\beta} = q_1, q_2, \sum_{j=r+1}^{p} q_j, \sum_{j=1}^{r} q_j$ and $\sum_{j=1}^{p} q_j$ we obtaine the following five orientaions:

(1) $\underset{\sim}{\alpha} = [0, 0, 0, \ldots, 1]'$ (21.a)

(2) $\underset{\sim}{\alpha} = [1, 0, 0, \ldots, 0]'$ (21.b)

(3) $\underset{\sim}{\alpha} = [0, 0, \ldots, 0, 1, 1, \ldots, 1]'$ (21.c)

(4) $\underset{\sim}{\alpha} = [1, 1, \ldots, 1, 0, 0, \ldots, 0]'$ (21.d)

(5) $\underset{\sim}{\alpha} = [1, 1, 1, \ldots, 1, 1, 1]'$ (21.e)

It is known [see Gibbons (1981) that the choice $\underset{\sim}{\beta} = q_p$ in (21.a) is unfavorable to ridge while $\underset{\sim}{\beta} = q_1$ is favorable in (21.b). The other orientations in [(21.c) - (21.e)] described different orientation situations. Six different values from (0.0001) to (2.5) were used for σ^2.

5.3. Replications

For each of the three models, for each of the six values of σ^2 and for each of the five orientations considered; (3x6x5=90 combinations); 1000 samples were generated. In each sample $\hat{\alpha}_j(0)$ was generated via $\hat{\alpha}_j \sim N(\alpha_j, \sigma^2/\lambda_j)$ and $\hat{\sigma}^2(0)$ was generated via $(n-p)\hat{\sigma}^2(0)/\sigma^2 \sim \chi^2(n-p)$. Each of the estimators $\hat{K}_1, \hat{K}_2, \hat{K}_3$ and $\hat{K}_4$

were computed an comparmentwise by using the estimate of $MSE(\hat{\alpha}(k)$ formula (9) also compared with least squares by formula (5).

5.4. Summary Statistics

For each of the 90 combinations,we kept track,over all 1000 replications, of:

(1) The average and the expected values of F and R^2 statistics;

(2) The minimum ,maximum, mean and CV of $\hat{K}$;

(3) The mean of MSE(k) and MSE(0);

(4) % of runs for which $\hat{L} <0$;

(5) % of runs for which $\hat{L} < \hat{L}_1$;

(6) % of runs $MSE(\hat{K}_i) < MES(0)$, i=1,2,3,4.

6. The Results

The main numerical results of the simulation are summarized in Tables (1) to (3). These tables just in case of orientation five which considere the common case where all the explanatory variables to includ in the model. We begin by establishing the validity of our simulation, then we discuss the MSE's.

6.1. The Validity of Our Simulation

Table (1) containes the average values of $\hat{F}$ and $\hat{R}^2$. For each model, α and σ combination, the values of $\hat{F}$ and $\hat{R}^2$ were computed and the averages over 1000 runs of $\hat{F}$ and $\hat{R}^2$ were computed. Recall that

$$\hat{F} = \hat{\alpha}'(0) \Lambda \hat{\alpha}(0) / p \hat{\sigma}^2(0)$$

$$\hat{R}^2 = \hat{\alpha}'(0) \Lambda \hat{\alpha}(0) \; / \; (Y-\bar{Y}1)'(Y-\bar{Y}1).$$

An examination of this table will reveal that the F statistic ranges from highly significant to insignificant levels, where as $\hat{R}^2$ ranges from 0.999 to below 0.40.

To check an simulation values, the expected values of F (EF) were computed using

$$E(\hat{F}) = (n-p-1)/(n-p-3) \; [(\hat{\alpha}'(0) \; \Lambda \; \hat{\alpha}(0))/p \, \sigma^2) + 1]$$

and are presented in the same table. They are closs to the simulated values. Further evidence of the validity of the simulations is given in Table (2). It contains the values of the simulated MES of least squares and the other estimates. One can notice the closes agreament between the simulated and theoretical values. These calculations indicate the accuracy of the MSE resultes to be presented in the sequal and testify to the validity of our simulation.

6.2. Mean Squared Error

Ridge estimators are constructed with the aim of having smaller MSE than least squares. We use the percentages of the number of times that $MSE(\hat{k}_i) < MSE(0)$ as a measure of the improvement obtained by using $\hat{\alpha}(\hat{K}_i)$; i =1,2,3,4. Table (3) shows these percentages denoted by $P(k_i,0)$. Also the table presents the percentages of number of times $\hat{L} < 0$ and $\hat{L} < \hat{L}_1$ denotes in the table by $P(L_o)$ and $P(L_1)$ respectivelly . The following observations are made upon examination the table.

(1) From Table (2) we can notice that the average values of the $MSE(k_3)$ and $MSE(k_4)$are less than that of the average of

MSE(0). This means that by applaying ridge regression using ($\hat{k}_3$ or $\hat{k}_4$) we have reduced the MSE values. Comparing the average values of MSE(k_1) and MSE(k_2) with the corresponding MSE(k_3) and MSE(k_4), we notice clearly the advantages of $\hat{k}_3$ and $\hat{k}_4$.

(2) From Table (3) the value of $P(L_1)$ is less than the value of $P(L_o)$ for all σ^2, this means that The percentages number of times we can compute $\hat{k}_3$ or $\hat{k}_4$ and then applay the ridge method is greater than we compute $\hat{k}_1$ or $\hat{k}_2$ under the new selction rule.

(3) For very small σ^2 (= 0.0001); the values $p(k_1,0)$, $p(k_2,0)$, $p(k_3,0)$ and $p(k_4,0)$ have the same values. That means all the estimators are presented the same results where all of them made about 100% improvement [see table (3)]. In fact we can expact this result from relation (18) that $\hat{L}_1 \cdot 0$ as $\hat{\sigma}^2 \cdot 0$, in this manner $\hat{L}_1 = \hat{L}_o$.

(4) With small $\hat{\sigma}^2(0)$ (= 0.01) and the high multicolinearity (the λ_j's are widely spread range) the new estimators $\hat{K}_3$ and $\hat{k}_4$ made good improvemets than the old $\hat{k}_1$ and $\hat{k}_2$. For not less than 84% and 74% the MSE($\hat{k}_3$) and MSE($\hat{K}_4$) are found to be less than MSE(0) respectivelly. Referring to table(2) we notice that the average values of simulated MSE($\hat{k}_3$) and MSE($\hat{k}_4$) are less tha MSE(0) where the values are 2.7623, and 3.2878 and 6.2849

respectivelly .

(5) The comparison between the two new estimates $\hat{K}_3$ and $\hat{k}_4$ it appears that $\hat{k}_3$ is better than $\hat{k}_4$ where $P(k_3,0)$ is always greater or equal to $P(k_4,0)$. It means that the percentage of improvements made using $\hat{k}_3$ is greater than the improvements under $\hat{k}_4$. Also, the average values of $MSE(k_3)$ are always less than the average values of $MSE(k_4)$.

7. Conclusion

Two ridge parameters estimates, denoted by $\hat{k}_3$ and $\hat{k}_4$, were suggested depend on new selection rule. This new rule depends on the number of observations (n), number of variables (p) and the variance of the error term (σ^2) in the model to determine certain critical value. The advantage behind this improvement selection rule is on making the range of $\hat{k}$ larger than it was, and bounded by unity. The simulation technique and MSE criteria were used to compare between the different ridge estimates. The results show that the MSE values using both $\hat{K}_3$ and $\hat{k}_4$ are smaller than MSE not only when using the least squares but also when using $\hat{K}_1$ and $\hat{k}_2$. By large σ^2, the values of MSE using $\hat{k}_3$ is smaller than MSE using $\hat{k}_4$. In cases where the multicollinearity is severe, we recommend to use $\hat{k}_3$.

References

(1) El-Bassiouni, M.Y. and El-Sayed, M,S (1986)" On the Selection of the Ridge Parameter",The 21st Annual Confierence in Statistics, Computer Science and Operation Research, ISSR, Cairo University.

(2) Gibbons , D.G.(1981)"A simulation Study of Some Ridge Estimatiors",JASA,76, 131-139.

(3) Gorman , J.W. and Toman, R.J.(1966) " Selection of Variables for Fitting Equations to Data",Tech.,B.,27-51.

(4) Gunst, R.T. and Mason, R.L. (1977)"Biased Estimation in Regression : An Evaluation Using Mean squaressd errors", JASA,72, 616 -628.

(5) Hald, A.(1952) " *Statistical Theory With Engineering Applications"*,John Willy and Sons, New York.

(6) Hoerl, A.E. and Kennard, R.W. (1970)" Ridge Regression: Biased Estimation for Non-Orthognal Problem", Tech.,12,55-67.

(7) Hoerl,A.E. and Kennard, R.W. (1970) "Ridge Regression : Iterative Estimation of the Biasing Parameter",Commun. Statist., A5,77-88.

(8) Hoerl, A.E. , Kennard, R.W. and Baldwin , K.F. (1975) " Ridge Regression : Some Simulation " , Commun. Statist., A4, 105-123.

(9) Lawless, J.F. (1978) " Ridge and Related Estimation Procedures Theory and Practice" , Commun. Statist. 7, 139-163.

(10) Lawless, j.F. and Wang, p.(1976)"A simulation Study of Ridge and Other Regression Estimators " , Commun. Statist., A5, 307-323.

(11) Lee, T.S. and Campbell, D.B.(1985) " Selection the Optemum K in Ridge Regression " , Commun. Statist.14,1589-1604.

(12) McDonald, G.C. and Galarneau, D.I.(1975)"A Monte Carlo Evaluation of Some Ridge Type Estimators " , JASA ,70, 407-416.

(13) McDonald, G.C. and Schwing, R.W. (1973) " Instability of Regression Estimates relating Air Pollution to Mortality" Technometrics, 15, pp. 463-481.

(14) Newhouse, J.P. and Oman, S.D.(1971)"An Evaluation of Ridge Estimators " , Rand Report no.R-716-PR,1-28.

(15) Pagel, M.D. (1981)"Comment on Hoerl and Kennard's Ridge Regression Simulation Methodology"Commun. Statist.A10, 2361-2367.

(16) Wichern, D.W. and Churchill, G.A.(1978)" A comparison of Ridge Estimators", Tech.20,301-311.

Table(1) The Average and Expected Values of F and R^2 Statistics.

Model	σ^2	EF	F	ER^2	R^2
i	0.0001	13334.78	12924.58	0.99985	0.99980
	0.0100	134.67	132.03	0.98537	0.98029
	1.0000	2.67	2.70	0.57143	0.48809
	1.5000	2.22	2.37	0.52632	0.44902
	2.0000	2.00	2.06	0.50000	0.42260
	2.5000	1.87	1.89	0.48276	0.40256
ii	0.0001	10870.22	11018.04	0.99977	0.99975
	0.0100	109.78	110.78	0.97773	0.97604
	1.0000	2.17	2.20	0.46511	0.44197
	1.5000	1.81	1.82	0.42016	0.39486
	2.0000	1.63	1.66	0.39473	0.37400
	2.5000	1.52	1.51	0.37838	0.35117
iii	0.0001	10480.59	10423.30	0.99972	0.99971
	0.0100	105.84	107.15	0.97303	0.97207
	1.0000	2.10	2.08	0.41671	0.39980
	1.5000	1.75	1.76	0.37316	0.36207
	2.0000	1.57	1.53	0.34886	0.33105
	2.5000	1.47	1.48	0.33335	0.32471

Table (2) Theoretical and Averages of the Simulated MSE Values of the Estimators.

Model	σ^2	Theoretical	Simulated				
		MSE(0)	MSE(0)	MSE(K_1)	MSE(K_2)	MSE(K_3)	MSE(K_4)
i	0.0001	0.0622	0.0637	0.0591	0.0591	0.0591	0.0591
	0.0100	6.2228	6.2849	[illegible].0142	6.1397	2.7623	3.2878
	1.0000	622.2799	641.7007	642.5825	697.7755	305.5692	360.7623
	1.5000	933.4199	907.3860	910.4467	986.3149	437.9143	513.7825
	2.0000	1244.5598	1256.1572	1257.0043	1364.1293	626.6008	732.8257
	2.5000	1555.6998	1574.3633	1570.9306	1705.4192	751.1047	885.5932
ii	0.0001	0.0034	0.0033	0.0033	0.0033	0.0033	0.0033
	0.0100	0.3384	0.3344	0.3087	0.3088	0.3087	0.3088
	1.0000	33.8397	33.2858	28.7452	31.0111	19.1582	21.4241
	1.5000	50.7595	49.9612	44.2613	47.2889	29.3165	32.3441
	2.0000	67.6793	67.2670	61.1504	65.1163	40.0083	43.9741
	2.5000	84.5992	85.0027	77.6511	82.7211	51.9322	57.0023
iii	0.0001	0.0263	0.0264	0.0260	0.0260	0.0260	0.0260
	0.0100	2.6316	2.6047	1.8013	1.8629	1.8013	1.8629
	1.0000	263.1618	262.5225	299.0338	314.8259	178.8157	194.6078
	1.5000	394.7428	391.3706	441.7546	464.4032	259.8177	282.4663
	2.0000	526.3237	531.5898	593.6357	622.3979	342.3381	371.1003
	2.5000	657.9046	657.2157	754.5502	793.9130	460.8518	500.2146

Table (3) % of Times Estimators Are Better than Least Squares.

σ^2	Model	$P(L_0)$	$P(K_1,0)$	$P(k_2,0)$	$P(L_1)$	$P(K_3,0)$	$P(K_4,0)$
0.0001	i	0.00	99.90	99.80	0.00	99.90	99.80
	ii	0.00	100.00	100.00	0.00	100.00	100.00
	iii	0.00	100.00	100.00	0.00	100.00	100.00
0.0100	i	43.50	41.40	30.80	0.00	84.90	74.30
	ii	0.00	97.40	96.80	0.00	97.40	96.80
	iii	0.00	77.50	75.00	0.00	77.50	75.00
1.0000	i	63.90	16.60	0.60	1.00	79.50	63.50
	ii	42.30	47.30	38.40	1.20	88.40	79.50
	iii	66.10	9.10	4.50	0.00	75.20	70.60
1.5000	i	64.60	14.60	0.90	1.70	77.50	63.80
	ii	47.60	43.30	35.00	3.10	87.80	79.50
	iii	65.30	9.70	4.80	0.00	75.00	70.10
2.0000	i	65.10	15.00	0.40	2.10	78.00	63.40
	ii	50.70	39.80	29.70	3.90	86.60	76.50
	iii	68.40	8.30	3.60	0.10	76.60	71.90
2.5000	i	66.50	15.50	0.40	1.60	80.40	65.30
	ii	52.30	38.70	29.00	5.50	85.50	75.80
	iii	66.10	8.00	4.20	0.10	74.00	70.20