Journal of Computer and Mathematical Sciences, *Vol.8(10), 504-511 October 2017* *ISSN 0976-5727 (Print)*
(An International Research Journal), *www.compmath-journal.org* *ISSN 2319-8133 (Online)*

Predictive Efficiency of Restricted and Mixed Regression Estimator

Amit Sharma[1] and Manoj Kumar[2*]

[1]Department of Statistics,
University of Jammu, Jammu – 180 006, INDIA.
[2]Corresponding Author, Department of Statistics,
Panjab University, Chandigarh – 160 014, INDIA.
email: amitstat99@gmail.com, mantiwa@gmail.com.

(Received on: September 23, Accepted: October 17, 2017)

ABSTRACT

In this article we have considered the composite target function to expose the ridge regression and also the estimators obtained by utilizing the prior non-sample information for prediction purpose. Finally, the performance properties of the estimators are analyzed.

Keywords: Linear regression model, ridge regression estimator, restricted ridge estimator mixed regression estimator, target function, prior non-sample information.

1. INTRODUCTION

Generally predictions from a linear regression are made either for the actual values of the study variable or for the average values at a time, see. Rao and Toutenberg (1995). However situation may arise in which one may be required to consider the predictions of the both actual and the average values simultaneously. For example, manufacturer of a certain product would like to know the life of his product on the average. On the other hand, a customer who purchases this product would be more interested in knowing the actual life of the product, which he has purchased. Thus, the manufacturer is interested in the prediction of the average values but he may not completely ignore the interest of the customer in the predictions of actual value and so he may assign higher weightage to the prediction of actual values. Similarly, the customer may give more weightage to the prediction of actual value in comparison to that of the average values. Appreciating the need of simultaneous prediction of actual and average values of the study variable, Shalabh (1995) proposed a composite target function, which considers the prediction of both the actual and average values together.

But if the usual assumption of the full column rank of the design matrix X is violated in the linear regression model $Y = X\beta + \sigma U$ is violated, the problem of multicollinearity arises. This leads to unstable estimates of the regression coefficients, which yields unstable predictors, so they (predictors) are practically useless. To overcome this, different remedial actions have been proposed. A popular numerical technique to deal with multicollinearity is ridge regression proposed by Hoerl and Kennard (1970a,b). Consequently, several authors have suggested different estimators, see. E.g., Sarkar (1992, 1998), Liu(1993) and Kaciranlar *et al*. (1999), Akdeniz,F, Erol,H (2003).

In this paper the author consider the composite target function to expose ridge regression estimator and the estimator obtained by considering the prior non-sample information, for prediction purpose and analyses their performance properties

The organization of the paper is as follows; section 2 deals with model specification and presents target function for the prediction of actual and average values of the study variables. In section 3 the author assumes the availability of prior non-sample information and formulates the estimators, viz., ridge regression restricted ridge and mixed ridge. In section 4 two predictors are presented and their properties are analyses. In section 5 the author compares the predictors on the basis of their properties, which were analyzed in section 4 and lastly the results of the theorem presented in section 4 are derived in section 6.

2. MODEL SPECIFICATIONS AND TARGET FUNCTION

Let us consider a linear regression model

$$Y = X\beta + \sigma U \tag{2.1}$$

Where Y is a $(n\,x1)$ vector of n-observations on study variables; X is a $(n\,x\,p)$ matrix with full column rank of observations on p-explanatory variables; β is a $(p\,x1)$ vector of regression coefficients; σ being an unknown positive scalar and U is a $(n\,x1)$ vector of n–observable disturbances with mean zero and variance unity.

If *b* denotes the estimate of β, then the predictor for values of the study variable with in the sample is generally formulated as $\hat{T} = Xb$ which is used for predicting either the actual value of Y or the average value $E[Y] = X\beta$ at a time.

When the situation demands prediction for both the average and actual together, we may define the following target function

$$T(Y) = \alpha Y + (1-\alpha)E(Y) = T \tag{2.2}$$

and use $\hat{T} = Xb$ for predicting it, where $0 \le \alpha \le 1$ is a non stochastic scalar specifying the weightage to be assigned to the prediction of actual and average values of the study variable,see.,Shalabh (1995)

3. PRIOR NON-SAMPLE INFORMATION

Suppose in addition to the sample information given in (2.1) there exists non-sample information which is in the form of stochastic restrictions i.e.

$$r = R\beta \qquad (3.1)$$

Where r is a $(J \times 1)$ vector of known elements and R is a $(J \times p)$ full column rank matrix of known elements

If we ignore the restrictions (3.1) then the least squares estimator of β is

$$b_o = X^{+}Y \qquad (3.2)$$

Where $X^{+} = (X'X)^{-1}X'$

And following the ridge regression method of estimation proposed by Hoerl & Kennard (1970a, b) the author shall write

$$b(k) = W^{-1}X'Y \qquad (3.3)$$

Where *k* is a non-negative scalar characterizing the estimator and $W = (X'X + kI)$

The estimator defined in (3.2) and (3.3) doesn't obey the restriction (3.1), it is not so with the restricted ridge estimator given by

$$b_R(k) = b(k) - (X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}\left[Rb(k) - r\right] \qquad (3.4)$$

This obeys the restrictions (3.1)

Suppose the restrictions are stochastic in nature, i.e

$$r = R\beta + V \qquad (3.5)$$

Where V is a $(J \times 1)$ random vector with mean zero and variance-covariance matrix Ω which is positive definite and known.

Further it is assumed that U and V are stochastically independent

The mixed ridge type estimator of β is

$$b_M(k) = \left[I_p - kW^{-1}\right]b_M \qquad (3.6)$$

Where

$$b_M = \left[X'X + S^2R'\Omega^{-1}R\right]^{-1}\left[XY + S^2R\Omega^{-1}r\right] \qquad (3.7)$$

$$S^2 = \frac{1}{n-p}\left((Y - X\beta)(Y - X\beta)'\right) \qquad (3.8)$$

Here S^2 is an unbiased estimator of σ^2 ,see., Theil and Goldberger (1961)

4. PREDICTIONS

Employing (3.2), (3.3) and (3.6) yields the following predictors for the values of study variables

$$\hat{T}(k) = Xb(k) \qquad (4.1)$$

$$\hat{T}_R(k) = Xb_R(k) \qquad (4.2)$$

$$\hat{T}_M(k) = Xb_M(k) \qquad (4.3)$$

It is easy to see from (2.2) that both the predictors defined in (4.1) and (4.2) are biased with bias vectors

$$B\left[\hat{T}(k)\right] = -kXW^{-1}\beta \qquad (4.4)$$

$$B\left[\hat{T}_R(k)\right] = -kXAW^{-1}\beta \qquad (4.5)$$

Further their mean squared error matrices and predictive risks are

$$M\left[\hat{T}(k)\right] = k^2 XW^{-1}\beta\beta' W^{-1}X' + \sigma^2\left\{\alpha^2 I_n - 2\alpha XW^{-1}X' + XW^{-1}X'XW^{-1}X'\right\} \qquad (4.6)$$

$$M\left[\hat{T}_R(k)\right] = M\left[\hat{T}(k)\right] - [k^2\{XA^*W^{-1}\beta\beta'W^{-1}X' + XAW^{-1}\beta\beta'W^{-1}A^{*\prime}X'\} + \sigma^2\{XA^*W^{-1}X'XW^{-1}X' + XA\,W^{-1}X'XW^{-1}A^*X' - \alpha X(A^*W^{-1} + W^{-1}A^{*\prime})X'\}] \qquad (4.7)$$

Where

$$A = I - A^* \qquad (4.8)$$

$$A^* = (XX)^{-1}R'\left[R(X'X)R'\right]^{-1}R \qquad (4.9)$$

$$\rho\left[\hat{T}(k)\right] = k^2\beta'W^{-1}X'XW^{-1}\beta + \sigma^2\left[(n-p)\alpha^2 + \sum_{j=1}^{p}\left\{\frac{\alpha k}{\lambda_j + k} - \frac{(1-\alpha)\lambda_j}{\lambda_j + k}\right\}^2\right] \qquad (4.10)$$

$$\rho\left[\hat{T}_R(k)\right] = \rho\left[\hat{T}(k)\right] - \left[k^2\beta W^{-1}X'XA^*W^{-1}\beta + \sigma^2\left\{(1-2\alpha)\sum_{j=1}^{p}\left(\frac{\lambda_j}{\lambda_j + k}\right)^2 - 2\alpha k\sum_{j=1}^{p}\frac{\lambda_j}{(\lambda_j + k)^2}\right\}\right]$$

Where λ_j 's are the Eigen values of $X'X$

In order to derive the expression for bias, mean squared error matrix and predictive risk of $T_M(k)$, we employ the small disturbances asymptotic theory introduced by Kadane (1971).

Theorem: When the disturbances in the model (2.1) are normally distributed, the predictive bias, mean squared error matrix and risk of $T_M(k)$ up to order $o(\sigma^4)$ of approximations is given by

$$B\left[\hat{T}_M(k)\right] = -kXW^{-1}\beta \qquad (4.12)$$

$$M\left[\hat{T}_M(k)\right] = M\left[T(k)\right] - \sigma^4 \left[\left(1 - \frac{2}{n-p}\right)XW^{-1}R\Omega^{-1}RW^{-1}X' - \alpha(X^+R'\Omega^{-1}RW^{-1}X' + XW^{-1}R\Omega^{-1}R^+)\right] \qquad (4.13)$$

$$\rho\left[\hat{T}_M(k)\right] = \rho\left[T(k)\right] - \sigma^4 \left[\left(1 - \frac{2}{n-p}\right)\sum_{j=1}^{p}\left(\frac{tr\dot{R}\Omega^{-1}R\dot{X}X}{(\lambda_j + k)^2}\right) - 2\alpha\sum_{j=1}^{p}\left(\frac{trR'\Omega^{-1}R}{\lambda_j + k}\right)\right]$$

5. A COMPARISON

On comparing the expression (4.10) and (4.11), the author finds that up to the unit power of k and $0 \le \alpha < 0.5$, the predictive risk associated with $\hat{T}_R(k)$ is less than the predictive risk associated with $T(k)$, if the characterizing scalar k satisfies the constraint

$$0 < k \le \left[\left(\frac{1-2}{2\alpha}\right)\sum_{j=1}^{p}\left(\frac{\lambda_j}{(\lambda_j + k)}\right)^2\right]\left[\sum_{j=1}^{p}\frac{\lambda_j}{(\lambda_j + k)^2}\right]^{-1} \qquad (5.1)$$

and just reverse holds, if $0.5 < \alpha \le 1$.

For $\alpha = 0.5$,the predictive risk associated with $\hat{T}_R(k)$ is smaller than the predictive risk associated with $\hat{T}(k)$, if the lower bound of k is

$$k \ge \sigma^2\left[\beta' W^{-1}X'XA^*W^{-1}\beta\right]\left[\sum_{j=1}^{p}\frac{\lambda_j}{(\lambda_j + k)^2}\right] \qquad (5.2)$$

On comparing the expression (4.10) and (4.14), the author finds that for a specified value of *k* suggested by Hoerl, Kennard and Baldwin (1975), the predictive risk associated with $\hat{T}_M(k)$ is smaller than the predictive risk associated with $\hat{T}(k)$ so long as

$$\alpha < \left(\frac{1}{2} - \frac{1}{n-p}\right) \qquad (5.3)$$

Case I: When the aim is to predict the average value of the study variable $(\alpha = 0)$

$$\rho_{AV}\left[\hat{T}(k)\right] = k^2\beta' W^{-1}X'XW^{-1}\beta + \sigma^2\sum_{j=1}^{p}\left(\frac{\lambda_j}{\lambda_j + k}\right)^2 \tag{5.4}$$

$$\rho_{AV}\left[\hat{T}_R(k)\right] = k^2\beta' W^{-1}X'XAW^{-1}\beta \tag{5.5}$$

$$\rho\left[\hat{T}_M(k)\right] = \rho\left[\hat{T}\ (k)\right] - \sigma^4\left(1 - \frac{2}{n-p}\right)\sum_{j=1}^{p}\left(\frac{trR'\Omega^{-1}RX'X}{(\lambda_j + k)^2}\right) \tag{5.6}$$

On comparing the expression (5.4) and (5.5), the author finds that for the average value prediction the predictor $\hat{T}_R(k)$ is better than that of $\hat{T}(k)$ for a fixed value of *k.*

On comparing the expression (5.4) and (5.6), the author finds that for the average value prediction the predictor $\hat{T}_M(k)$ has smaller risk than that of $\hat{T}(k)$ if

$n > (p+2)$

Case II: When the aim is to predict the actual values of the study variable $(\alpha = 1)$

$$\rho_{AC}\left[\hat{T}(k)\right] = k^2\beta' W^{-1}X'XW^{-1}\beta + \sigma^2\left\{n - 2\sum_{j=1}^{p}\left(\frac{\lambda_j}{\lambda_j + k}\right) + \sum_{j=1}^{p}\left(\frac{\lambda_j}{\lambda_j + k}\right)^2\right\} \tag{5.7}$$

$$\rho_{AC}\left[\hat{T}_R(k)\right] = k^2\beta' W^{-1}X'XAW^{-1}\beta + n\sigma^2 \tag{5.8}$$

$$\rho_{AC}\left[\hat{T}_M(k)\right] = \rho_{AC}\left[\hat{T}(k)\right] - \sigma^4\left[\left(1 - \frac{2}{n-p}\right)\sum_{j=1}^{p}\left(\frac{trR'\Omega^{-1}RX'X}{(\lambda_j + k)^2}\right) - 2\sum_{j=1}^{p}\left(\frac{trR'\Omega^{-1}R}{\lambda_j + k}\right)\right] \tag{5.9}$$

On comparing the expressions (5.7) and (5.8) the author finds that for actual value prediction, $\hat{T}_R(k)$ have smaller risk than $\hat{T}(k)$, if the lower bound of k is

$$k \geq \sigma\left[\beta' W^{-1}X'XA^{*}W^{-1}\beta\right]^{-1/2}\left[\left\{2\sum_{j=1}^{p}\left(\frac{\lambda_j}{\lambda_j + k}\right) - \left(\frac{\lambda_j}{\lambda_j + k}\right)^2\right\}\right]^{1/2} \tag{5.10}$$

On comparing the expression (5.7) and (5.9) we see that

$$\rho_{AC}\left[\hat{T}(k)\right]-\rho_{AC}\left[\hat{T}_M(k)\right] = \sigma^4\left[\left(1-\frac{2}{n-p}\right)\sum_{j=1}^{p}\left(\frac{trR'\Omega^{-1}RX'X}{(\lambda_j+k)^2}\right) - 2\sum_{j=1}^{p}\left(\frac{trR'\Omega^{-1}R}{\lambda_j+k}\right)\right] \quad (5.11)$$

Thus for actual value prediction, $T_M(k)$ performs better than $\hat{T}(k)$,if the terms on the right hand side of the expression (5.7) are positive.

APPENDIX

Employing (2.1) and (2.2) in (3.6) provides

$$\hat{T}_M(k)\text{-}T = \phi_o + \sigma\phi_1 + \sigma^2\phi_2 + \sigma^3\phi_3 + o(\sigma^4) \quad (6.1)$$

Where

$$\phi_0 = -kXW^{-1}\beta \quad (6.2)$$

$$\phi_1 = (XW^{-1}X'U) \quad (6.3)$$

$$\phi_2 = \left(\frac{U'MU}{n-p}\right)X'W^{-1}R'\Omega^{-1}V \quad (6.4)$$

$$\phi_3 = -\left(\frac{U'MU}{n-p}\right)X'W^{-1}R'\Omega^{-1}RX^{+}U \quad (6.5)$$

Here it is easy to see that for normally distributed disturbances

$$E[\phi_0] = \phi_0 \quad (6.6)$$

$$E[\phi_1] = E[\phi_2] = E[\phi_3] = 0 \quad (6.7)$$

Utilizing the expression (6.6) and (6.7) in (6.1) yield the result (4.12) of the theorem

Now

$$\left[\hat{T}_M(k)-T\right]\left[\hat{T}_M(k)-T\right]' = \phi_0\phi_0' + \sigma\left[\phi_o\phi_1' + \phi_1\phi_0'\right] + \sigma^3\left[\phi_3\hat{\phi}_0 + \phi_2\hat{\phi}_1 + \phi_2\phi_1' + \phi_0\phi_3'\right] + \sigma^4\left[\phi_1\phi_3' + \phi_2\phi_2' + \phi_3\phi_1'\right] \quad (6.8)$$

It is easy to verify that

$$E[\phi_0\phi_1'] = k^2XW^{-1}\beta\beta'W^{-1}X' \quad (6.9)$$

$$E[\phi_0\phi_1'] = E[\phi_0\phi_2'] = E[\phi_0\phi_3'] = E[\phi_1\phi_2'] = 0 \quad (6.10)$$

$$E[\phi_1\phi_1'] = \left\{\alpha^2I_n - 2\alpha XW^{-1}X' + XW^{-1}X'XW^{-1}X'\right\} \quad (6.11)$$

$$E[\phi_1\phi_3'] = -\{XW^{-1}R'\Omega^{-1}RW^{-1}X' - \alpha X^{+}R'\Omega^{-1}RW^{-1}X'\} \quad (6.12)$$

$$E[\phi_2\phi_2'] = \left(\frac{n-p+2}{n-p}\right)XW^{-1}R'\Omega^{-1}RW^{-1}X' \quad (6.13)$$

Utilizing expressions (6.9), (6.10), (6.11), (6.12) and (6.13) in (6.8) yields the result (4.13) of the theorem

Again

$$\rho\left|\hat{T}_M(k)\right| = [T_M(k) - T]' \; [T_M(k) - T] = trM\left|\hat{T}_M(k)\right| \quad (6.14)$$

and thus result (4.14) of the theorem follows from the expression (6.14)

REFERENCES

1. Hoerl, A.E and Kennard, R.W., ”Ridge regression: biased estimation for non-orthogonal problems” *Technometrics*, 12, 55-67 (1970 a).
2. Hoerl, A.E and Kennard, R.W., “Ridge regression: biased estimation for non-orthogonal problems” *Technometrics*, 12, 69-82 (1970 b).
3. Hoerl, A.E and Kennard, R.W., and Baldwin, K.F.,“Ridge regression: some simulation”, *Commn. Stat*., 4, 105-123 (1975).
4. Kaciranlar,S., Sakallioglu, S., Akdeniz, F., Styan , G.P.H., and H.J., Werner. “A new biased estimator in linear regression and detailed analysis of widely analyses data on Portland Cement”, *Sankhya, Series B*, 61, 443–459 (1999)
5. Kadane, J.B., “Comparisons of k-class estimators when the disturbances are small”, *Econometrica*, 39, 723-737 (1971).
6. Rao, C.R., and H., Toutenburg. Linear Models: Least Squares and Alternatives, Springer, New York (1995).
7. Shalabh. “Performance of Stein-Rule procedure for simultaneous prediction of actual and average values of the study variable in linear regression model”, *Bulletin of the International Statistical Institute* 56, 1375-90 (1995).
8. Sarkar, N., “A new estimator combining the ridge regression and restricted least squares methods of estimation” *Comm. in Statistics, Series A*, 21, 1987-2000 (1992).
9. Sarkar, N., Erratum: “A new estimator combining the ridge regression and restricted least squares methods of estimation” *Comm. in Statistics, Series A*, 27, 1019-1020 (1998).
10. Theil, H and Goldberger, A.S., “On pure and mixed estimation in econometrics”, *International Economic Review* 2, 65-78 (1961).
11. Akdeniz, F and Erol, H. “Mean Squared Error Matrix Comparison of Some Biased Estimators in linear Regression”. *Communication in Statistics-theory and Methods* 32 (7), 2389-2413 (2003).