Yugoslav Journal of Operations Research
27 (2017), Number 2, 243–247
DOI: https://doi.org/10.2298/YJOR170114014T

PREDICTIVE EFFICIENCY OF RIDGE REGRESSION ESTIMATOR

Manoj TIWARI
Department of Statistics, Panjab University, Chandigarh - 160 014, INDIA
mantiwa@gmail.com

Amit SHARMA
Department of Statistics, University of Jammu, Jammu - 180 001, INDIA
amitstat99@gmail.com

Received: January 2017 / Accepted: May 2017

Abstract: In this article we have considered the problem of prediction within and outside the sample for actual and average values of the study variables in case of ordinary least squares and ridge regression estimators. Finally, the performance properties of the estimators are analyzed.

Keywords: Linear Regression Model, Ridge Regression Estimator, Least Squares Predictor, Rdge Regression Predictor, Prediction Within and Outside Sample, Prior non-sample Information.

MSC: 62J05, 62J07.

1. INTRODUCTION

The main aim of a linear regression model is to make prediction, either for the actual values or average values of the study variable; see., Rao and Toutenburg [4] for an interesting account. In both the cases, the estimated equation derived from least squares estimation of parameters provides the best linear unbiased predictions. If unbiasedness is not crucial and can be dropped, several shrinkage estimators are available which may bring substantial gain in precision at the cost of little bias. For example, the method of ridge regression see., Hoerl and Kennard [1],[2] provides an estimator, though biased, which has smaller mean square error than that obtained by the method of least squares. In addition to predicting average and actual values of a study variable within the sample, one may be interested in knowing performance when the aim is to predict the values outside the sample, for example for the purpose of forecasting and policy prescriptions. In the present paper, an attempt is made to compare the predictive efficiency of the least squares predictor and the predictor obtained by using the method of ridge regression. The organisation of the paper is as follows; Section 2 deals with model specification, Sections 3 and 4 deal with the problem of prediction within and outside the sample, respectively. Finally, the predictors are compared in Section 5.

2. MODEL SPECIFICATIONS

Let us postulate the following linear regression model

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

where, Y is a $(n \times 1)$ vector of n - observations on the study variable, X is a $(n \times p)$ full column rank matrix of n observations on p explanatory variables, β is a $(p \times 1)$ vector of the regression coefficients, U is a $(n \times 1)$ vector of n - observable disturbances assumed to follow multivariate normal distribution with mean zero and variance-covariance matrix $\sigma^2 I$, σ is unknown scalar, I is identity matrix of order $(n \times n)$.
In addition to the given n-observation on the study variables and the explanatory variables, let us suppose that further m-value of the same set of explanatory variables are known but the corresponding observations on the study variables are not available. Assuming that the model remains unchanged, we can express it as

$$Y_f = X_f\beta + U_f \tag{2.2}$$

where Y_f is a $(m \times 1)$ unobserved vector of study variables; X_f is a $(m \times p)$ matrix formed by m observations on p-explanatory variables; U_f is a $(n \times 1)$ vector of n observable disturbances, and it is also assumed to follow multivariate normal distribution with mean zero and variance-covariance matrix $(\sigma^2 I)$.

Further it is assumed that U and U_f are stochastically independent, see ,e.g.,Trenkler and Toutenburg [6].

3. PREDICTION WITHIN SAMPLE

If β is estimated by the method of least squares, the predictor for actual and average values of a study variable is given by

$$P_0 = HY \tag{3.1}$$

where $H = XX^+$ is the hat matrix and $X^+ = (X'X)^{-1}X'$.
Similarly, when the ridge regression method of estimation is used to estimate β, the predictor for actual and average values of a study variable is given by

$$P(k) = HY - kXW^{-1}X^{+}Y \tag{3.2}$$

where k is a positive scalar characterizing the predictor and $W = (X'X + kI)$. Let us first consider the use of P_0 and P_k for predicting the average values of the study variable. It is well known that P_0 is unbiased predictor with prediction variance

$$V_{AV}[P_0] = \sigma^2 p \tag{3.3}$$

Similarly, it is found that $P(k)$ is a biased predictor for average values of the study variable. The prediction mean square error defined as trace of mean squared error matrix $P(k)$is given by

$$M_{AV}[P(k)] = \sigma^2 p - k\sigma^2\Big[\sum_{j=1}^{p}\Big(\frac{1}{(\lambda_j + k)}\Big) + \sum_{j=1}^{p}\Big(\frac{\lambda_j}{(\lambda_j + k)^2}\Big)\Big] + k^2\beta' W^{-1}X'XW^{-1}\beta \tag{3.4}$$

where $\lambda_j^{'s}$; $(j = 1, 2, \cdots, p)$ are the eigenvalues of matrix $X'X$.

Now if we use P_0 and $P(k)$ for predicting the actual values of the study variable, we see that P_0 is unbiased predictor with prediction variance

$$V_{AC}[P_0] = \sigma^2(n - p) \tag{3.5}$$

where $P(k)$ is a biased predictor with mean square error as

$$M_{AC}[P(k)] = V_{AC}[P_0] + k^2\sigma^2\sum_{j=1}^{p}\Big(\frac{\lambda_j}{(\lambda_j + k)}\Big)^2 + k^2\beta' W^{-1}X'XW^{-1}\beta \tag{3.6}$$

4. PREDICTIONS OUTSIDE SAMPLE

If we use the least squares and ridge regression methods of estimation to estimate β from model (2.1) and to formulate the prediction from (2.2), we obtain the following predictors

$$P_{f_0} = X_f X^{+}Y \tag{4.1}$$

$$P_f = X_f X^{+}Y - kX_f W^{-1}X^{+}Y \tag{4.2}$$

If we compare these two predictors with respect to criterion of mean squared error, we find that conclusion remains unaltered whether these predictors are used for average values of Y_f or the actual values.

Using (2.1) and (2.2), we observe from (4.1) that P_{f_0} is an unbiased predictor and its prediction variance is given by

$$V_{AV}[P_{f_0}] = \sigma^2 T \tag{4.3}$$

where $T = (X'X)^{-1}X'_fX_f$
Similarly, we can see that $P_f(k)$ is a biased predictor and its prediction mean square error is given by

$$M_{AV}[P_f(k)] = \sigma^2T - k\sigma^2\Big[\sum_{j=1}^{p}\Big\{\frac{\lambda_{fj}}{\lambda_j}\Big(\frac{1}{\lambda_j+k)}\Big)\Big\} + \sum_{j=1}^{p}\Big(\frac{\lambda_{fj}}{(\lambda_j+k)}\Big)\Big] + k^2\beta'W^{-1}X'_fX_fW^{-1}\beta \tag{4.4}$$

where$\lambda_{fj}{}^{'s}$, $(j = 1, 2, \cdots p)$ are the eigenvalues of X'_fX_f.

5. A COMPARISON

* On comparing the expression (3.3) and (3.4), we find that the prediction mean square of $P(k)$ is smaller than the prediction variance of P_0, if the characterizing scalar k satisfies the following constraint,

$$0 < k \le \frac{\sigma^2}{\beta'W^{-1}X'XW^{-1}\beta}\Big[\sum_{j=1}^{p}\Big(\frac{1}{(\lambda_j+k)}\Big) + \sum_{j=1}^{p}\Big(\frac{\lambda_j}{(\lambda_j+k)^2}\Big)\Big] \tag{5.1}$$

** On comparing the expression (3.5) and (3.6) , we observe that for any value of k, $(0 < k < \infty)$ the prediction variance of P_0 is smaller than the prediction mean square of P_k.

Thus, we see that the ridge regression predictor with characterizing scalar k satisfying constraint (5.1) is better than the least squares predictor while predicting the average values of the study variable, but if we predict the actual values, the least squares predictor is unbeaten.

*** On comparing the expression (4.3) and (4.4), we see that the prediction mean square of $P_f(k)$ is smaller than the prediction variance of P_{f_0}, if the characterizing scalar k satisfies the following constraint,

$$0 < k < \frac{\sigma^2}{\beta'W^{-1}X'_fX_fW^{-1}\beta}\Big[\sum_{j=1}^{p}\Big\{\frac{\lambda_{fj}}{\lambda_j}\Big(\frac{1}{\lambda_j+k)}\Big)\Big\} + \sum_{j=1}^{p}\Big(\frac{\lambda_{fj}}{(\lambda_j+k)}\Big)\Big] \tag{5.2}$$

Thus the ridge regression predictor is better than the least squares predictor under constraint (5.2), irrespective of their use for prediction of the average values or the actual values of the study variable. When (5.2) does not hold true, the least squares predictor remains unbeaten.

REFERENCES

[1] Hoerl, A.E., and Kennard, R.W., "Ridge regression: biased estimation for non-orthogonal problems", *Technometrics*, 12 (1970 a) 55-67.
[2] Hoerl, A.E., and Kennard, R.W., "Ridge regression: biased estimation for non-orthogonal problems", *Technometrics*, 12 (1970 b) 69-82.
[3] Hoerl, A.E., Kennard, R.W., and Baldwin, K.F., "Ridge regression: some simulation", *Communications in Statistics*, 4 (1975) 105-123.

[4] Rao, C.R., and Toutenburg, H., *Linear Models: Least Squares and Alternatives*, Springer, New York, 1995.

[5] Theil, H., and Goldberger, A.S., ”On pure and mixed estimation in econometrics”, *International Economic Review*, 2 (1961) 65-78.

[6] Trenkler, G., and Toutenburg, H., ”Pre-test procedures and forecasting in the regression model under restrictions”, *Journal of Statistical Planning and Inference*, 30 (1992) 245-256.