Asian Journal of Mathematics and Statistics 4 (2): 90-97, 2011
ISSN 1994-5418 / DOI: 10.3923/ajms.2011.90.97

An Efficient Estimation Procedure For Determining Ridge Regression Parameter

Moawad El-Fallah Abd El-Salam
Department of Statistics and Mathematics and Insurance, Faculty of Commerce, Zagazig University, Zagazig, Egypt

ABSTRACT

A common problem in multiple regression models is multicollinearity, which produces undesirable effects on the least squares estimator. One of the popular methods for handling this problem is ridge regression estimation. In ridge regression, ridge parameter or biasing constant plays an important role in parameter estimation. Many researchers are suggested various procedures for determining the ridge parameter. In this study, we introduce an alternative procedure for determining the ridge parameter. The efficiency of the proposed procedure is evaluated and compared with some existing procedures in the literature, through simulation study in terms of Mean Square Error (MSE). The procedure developed in this study seems to be very reasonable because of having smaller MSE.

Key words: Multicollinearity, ridge regression estimation, ridge parameter

INTRODUCTION

It is well known that the Ordinary Least Squares (OLS) estimator is unbiased estimator. However, in the presence of multicollinearity, OLS estimator could becomes unstable due to their large variance, which leads to poor prediction. The popular solution method to solve this problem is ridge regression. In ridge regression method, we modify the least squares method to allow biased estimators of the regression coefficients. Therefore, these biased estimators are preferred over the least squares estimator, because they will have a larger probability of being close to the true parameter values with smaller MSE of regression coefficients. In the presence of multicollinearity, selection of ridge parameter plays an important role, by adding a small constant to the diagonal elements of the matrix for decreasing its "condition number" (D'Ambra and Sarnacchiaro, 2010). Many researchers are suggested various methods for selecting ridge parameter in ridge regression. These methods have been suggested by Khalaf and Shukur (2005), Mardikyan and Cetin (2008) and others.

In this study, we suggest an efficient estimation procedure for selecting the ridge parameter and hence ridge estimator.

MODEL AND ESTIMATORS

Consider widely used linear regression model:

$$Y = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where, Y is a n×1 vector of observations on a response variable Y. $\boldsymbol{\beta}$ is a p×1 vector of unknown regression coefficients, X is a matrix of order (n×p) of observations on p predictor (regressor) variables X_1, X_2, ..., X_p and $\boldsymbol{\varepsilon}$ is an $n\times 1$ vector of random variables which are distributed as $N(0, \boldsymbol{\sigma}^2 I_n)$. The most common estimator for $\boldsymbol{\beta}$ is the least squares estimator $\hat{\beta}=(X'X)^{-1}X'Y$. For the sake of convenience, we assume that the matrix X is standardized in such a way that $X'X$ is a non-singular correlation matrix. This study is concerned with dealing the situation $X'X$ has at least one small eigen value leading to a high MSE for $\boldsymbol{\beta}$ meaning that $\hat{\beta}$ is an unreliable estimator of $\boldsymbol{\beta}$. Let L and T be the matrices of eigen values and eigen vectors of $X'X$, respectively, satisfying $T X'T = L = (\lambda_1, \lambda_1, ..., \lambda_p)$ where, λ_i being the ith eigen value of $X'X$ and $T'T = TT' = I_p$. We obtain the equivalent model:

$$Y = Z\alpha + \varepsilon \tag{2}$$

where, $Z = XT$, it implies that $Z'Z = L$ and $\alpha = T'\boldsymbol{\beta}$ (Montgomery *et al*., 2006). Then Ordinary Least Squares (OLS) estimator of is given by:

$$\hat{\alpha} = (Z'Z)^{-1} = Z'Y = L^{-1}Z'Y \tag{3}$$

Therefore, OLS estimator of $\boldsymbol{\beta}$ is given by:

$$\hat{\beta} = T\hat{\alpha}$$

The Ordinary Ridge Regression (ORR) estimator of suggested by Hoerl and Kennard (1970) is written as:

$$\hat{\alpha}_{RR} = [I - KA_k^{-1}]\hat{\alpha},\ k \geq 0 \tag{4}$$

where, $A_k = (L + kI_p)$ and:

$$k = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$$

Hence ridge regression estimator of $\boldsymbol{\beta}$ is $\hat{\beta}_{RR} = T\hat{\alpha}_{RR}$ and mean square error of $\hat{\alpha}_{RR}$ is:

$$\text{MSE}\ (\hat{\alpha}_{RR}) = \text{Variance}\ (\hat{\alpha}_{RR}) + [\text{Bias}\ (\hat{\alpha}_{RR})]^2$$

$$= \hat{\sigma}^2 \sum_{i=1}^{p} \lambda_i / (\lambda_i + k)^2 + k^2 \sum_{i=1}^{p} \alpha_i^2 / (\lambda_i + k)^2 \tag{5}$$

where, $\hat{\sigma}^2$ is the OLS estimator of $\boldsymbol{\sigma}^2$ i.e:

$$\hat{\sigma}^2 = \frac{Y'Y - \hat{\alpha}'Z'Y}{n-p-1},\ \alpha = T'\beta$$

We observe that, when $k = 0$ in Eq. 4, OLS estimator of is recovered. As k increases the ridge regression estimators are biased but more precise than OLS estimator. Hoerl *et al*. (1975) suggested that, the value of "k" is chosen small enough, for which the mean squared error of ridge estimator, is less than the mean squared error of OLS estimator.

Many researchers have been suggested different ways of estimating the ridge parameter. Some of the well known methods for choosing ridge parameter value are listed below:

$$K_{HKB} = \frac{p\hat{\alpha}^2}{\hat{\alpha}'\hat{\alpha}} \quad \text{(Hoerl and Kennard, 1970)} \tag{6}$$

$$K_{LW} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p}\lambda_i\hat{\alpha}_i^2} \quad \text{(Lawless and Wang, 1976)} \tag{7}$$

$$K_{HMO} = p\hat{\alpha}^2 / \sum_{i=1}^{p}[\hat{\alpha}_i^2 / \{1+\lambda_i(\hat{\alpha}_i^2/\hat{\sigma}_i^2/\hat{\sigma}^2)^{\frac{1}{2}})\}], i=1,2,\ldots,p. \quad \text{(Nomura, 1988)} \tag{8}$$

$$K_{KS} = (\lambda_{max}\hat{\sigma}^2)/((n-p-1)\hat{\sigma}^2+\lambda_{max}\hat{\alpha}^2{}_{max} \quad \text{(Khalaf and Shukur, 2005)} \tag{9}$$

PROPOSED RIDGE PARAMETER

Hoerl and Kennard (1970) showed that ridge estimator is biased estimator and its squared bias is continuous and monotonically increasing function of "k". Also they proved that the MSE of $\hat{\alpha}_{RR}$ is less than MSE of $\hat{\alpha}$ when:

$$0 \le k \le \frac{\hat{\sigma}^2}{\hat{\sigma}^2{}_{max}}$$

where, $\hat{\alpha}^2{}_{max}$ is the largest element of α^2 and σ^2 is replaced by its estimate:

$$\hat{\sigma}^2 = \frac{Y'Y - \hat{\alpha}'Z'Y}{n-p}$$

Many researchers are interested in ridge estimator, such that this estimator having smaller total MSE than OLS estimator. The MSE of ridge estimator is depends on the ridge parameter (k).

In this study, we have suggested an alternative method for determining ridge parameter "k" and it is defined as:

$$k_D = \max\left(0, \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} = \frac{1}{n(VIF_j)max}\right) \tag{10}$$

where, $VIF_j = \frac{1}{1-R_j^2}$, j = 1, 2, .., p is variance inflation factor of jth regressor.

Our suggested estimator is modification of "k_{HKB}". The small amount:

$$\frac{1}{n(VIF_j)max}$$

is subtracted from "k_{HKB}". This amount however, varies with the size of the sample (n) used and strength of the multicollinearity in the model. Now we discuss some results related to the proposed method.

Result 1: If $(VIF_j)_{max}$ is too large, then k_D is an approximately "k_{HKB}".

Proof: The proposed ridge parameter is:

$$k_D = \max\left(0, \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} = \frac{1}{n(VIF_j)\max}\right)$$

If $(VIF_j)_{max}$ is too large then:

$$\frac{1}{n\,(VIF_j)_{max}} \rightarrow 0$$

Therefore:

$$\frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} - \frac{1}{n(VIF_j)_{max}} \rightarrow \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$$

Hence; we rewrite the proposed estimator as:

$$k_D = \max\left(0, \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}\right)$$
$$\Rightarrow k_D \cong k_{HKB} \qquad \text{since } k_{HKB} \geq 0$$

Result 2: If $(VIF_j)_{max}$ is close to one, then k_D is either 0 or k_{HKB}-1/n.

Proof: If $(VIF_j)_{max}$ is close to 1, then the quantity 1/n (VIF_j) max is approximately 1/n.
Hence k_{HKB}-1/n may be positive or negative. So that, we have considered two cases:

Case I: If $k_{HKB} \leq 1/n$, then $k_{HKB}-1/n \leq 0$.
Hence by definition of k_D, $k_D = 0$.

Case II: If $k_{HKB} > 1/n$ that implies, $k_D = k_{HKB}-1/n > 0$.
Therefore, $k_D = k_{HKB}-1/n$.

Result 3: $0 \leq K_D \leq K_{HKB}$

Proof: The proposed ridge parameter k_D, is:

$$kD = \max\left(0, k_{HKB} - \frac{1}{n\,(VIF_j)\max}\right)$$

The possible values of k_D are:

$$kD = 0 \text{ if } k_{HKB} \leq \frac{1}{n\,(VIF_j)\max}$$
$$kD > 0 \text{ if } k_{HKB} > \frac{1}{n\,(VIF_j)\max}$$

from above relation:

$$k_D \geq 0 \tag{11}$$

Let k_{HKB}, n and $(VIF)_{max}$ be th nonnegative. Hence:

$$k_{HKB} - \frac{1}{n\,(VIF_j)max} \leq k_{HKB}$$

Therefore:

$$k_D \leq k_{HKB} \tag{12}$$

from inequality Eq. 11 and 12:

$$0 \leq k_D \leq k_{HKB} \tag{13}$$

Hoerl *et al*. (1975) have shown that:

$$k_{HKB} \leq \frac{\sigma^2}{\hat{\alpha}^2_{max}}$$

Using this, inequality Eq. 13 becomes:

$$k_D \leq k_{HKB} \leq \frac{\sigma^2}{\hat{\alpha}^2_{max}}$$

Hence proposed ridge parameter (k_D) satisfy the upper bound of ridge parameter stated by Hoerl and Kennard (1970).

PERFORMANCE OF THE PROPOSED RIDGE PARAMETER

Here, we examined the performance of the ridge estimator using the proposed ridge parameter over the different ridge parameters (k). We examined the MSE ratio of the ridge estimator using proposed ridge parameter and other ridge parameters over OLS estimator.

We have considered two examples. In example 1, we generate data for two predictor variables with different combinations of sample size, correlation between predictor variables and variance of the error terms. In example 2, same simulation study is carried out for 4 predictor variables.

Example 1: We have generated random sample of size n for two predictor variables. To exhibit multicollinearity in the simulated data, we use the different degree of correlation between the variables included in the model. Here we put correlation values ρ = 0.999 and 0.9999. We have used sample size n = 20, 50, 75 and 100. The variance of the error terms are taken as σ^2 = 5, 10, 25 and 100. Ridge estimates are computed using different ridge parameters given in Eq. 6 to 10. These estimates are : k_{HKB} (Hoerl and Kennard (1970) using Eq. 6), k_{LW}(Lawless and Wang (1976) using Eq. 7), k_{HMO} (Nomura, 1988) using Eq. 8, k_{KS} (Khalaf and Shukur (2005) using Eq. 9) and k_D (The proposed estimator using Eq. 10). The MSE of such ridge regression parameters are obtained using Eq. 5. This experiment is repeated 1500 times and obtains the average MSE (AMSE). Firstly, we

Table 1: Ratio of AMSE of OLS over various ridge estimators for different "k"

p	K	Different n with $\sigma^2 = 5$				Different n with $\sigma^2 = 10$			
		20	50	75	100	20	50	75	100
0.999	LS/HKB	2.821	2.739	2.714	2.826	2.480	2.847	2.836	2.734
	LS/LW	1.776	1.401	1.151	1.164	1.969	2.013	1.589	1.115
	LS/HMO	1.903	1.829	1.819	1.974	1.783	1.973	1.950	1.827
	LS/KS	2.212	1.762	1.742	1.652	2.102	2.230	2.299	1.766
	LS/kD	2.875	2.806	2.787	2.896	2.482	2.866	2.863	2.706
	HKB/LS	0.354	0.365	0.368	0.354	0.403	0.351	0.353	0.370
	HKB/LW	0.630	0.512	0.424	0.412	0.793	0.707	0.560	0.412
	HKB/HMO	0.674	0.668	0.670	0.699	0.718	0.693	0.688	0.676
	HKB/KS	0.784	0.643	0.642	0.585	0.847	0.783	0.811	0.653
	HKB/kD	1.019	1.025	1.027	1.025	0.999	1.006	1.009	1.012
0.9999	LS/HKB	3.078	2.806	2.781	3.011	3.304	2.712	3.406	2.797
	LS/LW	1.095	0.909	0.934	1.239	1.954	1.234	1.302	1.051
	LS/HMO	1.999	1.823	1.888	2.117	2.424	1.875	2.335	1.928
	LS/KS	1.701	1.588	1.447	1.941	2.784	1.734	2.223	1.745
	LS/kD	3.177	2.889	2.846	3.088	3.329	2.741	3.445	2.828
	HKB/LS	0.325	0.356	0.36	0.332	0.303	0.369	0.294	0.358
	HKB/LW	0.356	0.324	0.336	0.411	0.591	0.455	0.382	0.376
	HKB/HMO	0.649	0.649	0.679	0.703	0.734	0.691	0.686	0.690
	HKB/KS	0.553	0.566	0.520	0.644	0.842	0.640	0.653	0.624
	HKB/kD	3.078	2.806	2.781	3.011	3.304	2.712	3.406	2.797
		Different n with $\sigma^2 = 25$				Different n with $\sigma^2 = 100$			
p	K	20	50	75	100	20	50	75	100
0.999	LS/HKB	2.221	2.562	2.796	2.635	2.087	2.487	2.862	2.695
	LS/LW	1.804	1.988	1.625	1.198	1.719	1.990	1.957	1.208
	LS/HMO	1.730	1.801	1.856	1.799	1.874	1.843	1.961	1.802
	LS/KS	1.864	2.137	2.046	1.699	1.735	2.087	2.351	1.775
	LS/ kD	2.222	2.567	2.816	2.655	2.088	2.489	2.874	2.711
	HKB/LS	0.447	0.390	0.358	0.379	0.474	0.402	0.349	0.371
	HKB/LW	0.807	0.776	0.581	0.455	0.815	0.799	0.684	0.448
	HKB/HMO	0.774	0.703	0.664	0.683	0.889	0.740	0.685	0.669
	HKB/KS	0.834	0.834	0.732	0.645	0.823	0.839	0.821	0.659
	HKB/kD	0.994	1.002	1.007	1.007	0.990	0.999	1.004	1.006
0.9999	LS/HKB	2.759	2.409	3.107	2.864	2.543	2.977	2.530	2.396
	LS/LW	1.471	1.003	1.283	1.297	1.424	1.335	0.960	0.932
	LS/HMO	1.896	1.588	2.124	2.072	1.706	2.001	1.658	1.626
	LS/KS	1.966	1.445	1.758	2.048	1.726	1.904	1.425	1.508
	LS/ kD	2.777	2.430	3.131	2.882	2.556	2.995	2.547	2.410
	HKB/LS	0.362	0.415	0.322	0.349	0.393	0.336	0.395	0.417
	HKB/LW	0.533	0.416	0.413	0.453	0.560	0.448	0.379	0.389
	HKB/HMO	0.687	0.659	0.684	0.723	0.671	0.672	0.655	0.679
	HKB/KS	0.713	0.60	0.566	0.715	0.679	0.640	0.563	0.629
	HKB/kD	1.007	1.008	1.008	1.006	1.005	1.006	1.007	1.006

computed the AMSE ratios of OLS estimator over different estimators. Secondly, AMSE ratios of ridge estimator using ridge parameter "k_{HKB}", over OLS and different ridge estimators are computed and these ratios are reported in Table 1. For the results of Table 1, we consider the method that lead to the maximum ratio to the best from the MSE point of view.

The results of Table1 show that, the proposed estimator (k_D) performs better than other ridge parameters for all combinations of correlation between predictors (ρ), variance of the error term (σ^2) and sample size (n), especially, when $\sigma^2 = 10$, n = 75 and $\rho = 0.9999$, used in this simulation study.

Example 2: We have generated random sample of size n from N_4 (0, Σ_1) on X_1, X_2, X_3 and X_4 where:

$$\Sigma_1 = \begin{bmatrix} 1 & 0.2290 & -0.8240 & -0.2450 \\ 0.2290 & 1 & -0.139 & -0.973 \\ -0.8240 & -0.139 & 1 & -0.030 \\ 0.2450 & -0.973 & -0.030 & 1 \end{bmatrix}$$

We consider the model as:

$$Y = 10 + X_1 + X_2 + 2X_3 + X_4 + \varepsilon, \text{ where, } \varepsilon \sim N(0, \sigma^2)$$

We have generated the data with sample sizes n = 20, 50, 75 and 100. The variance of the error terms are taken as $\sigma^2 = 1$, 5, 10 and 25. Same simulation study carried out as in Example 1 and the MSE ratios of different estimators over OLS estimator are reported in Table 2.

From the results of Table 2, we conclude that, the proposed method for estimating the ridge parameter performs quite well than all other ridge parameters for all combinations of variance of the error term (σ^2) and sample sizes (n) in our study, especially when $\sigma^2 = 25$ and n = 20.

Table 2: Ratio of AMSE of OLS over various ridge estimators for different "k"

	Different n with $\sigma^2 = 1$				Different n with $\sigma^2 = 5$			
K	20	50	75	100	20	50	75	100
LS/HKB	1.82	2.08	2.25	2.13	2.35	2.21	2.31	2.15
LS/LW	1.61	1.53	1.53	1.24	1.58	1.11	1.12	1.01
LS/HMO	1.20	1.46	1.63	1.54	1.74	1.6	1.69	1.56
LS/KS	1.64	1.86	2.07	1.84	1.96	1.78	1.84	1.68
LS/LD	1.83	2.10	2.28	2.16	2.38	2.23	2.34	2.18
HKB/LS	0.55	0.48	0.44	0.47	0.43	0.45	0.43	0.47
HKB/LW	0.88	0.74	0.68	0.58	0.67	0.51	0.49	0.46
HKB/HMO	0.66	0.70	0.72	0.72	0.74	0.73	0.73	0.73
HKB/KS	0.90	0.89	0.92	0.86	0.83	0.81	0.79	0.78
HKB/kD	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01
	Different n with $\sigma^2 = 10$				Different n with $\sigma^2 = 25$			
K	20	50	75	100	20	50	75	100
LS/HKB	2.17	2.25	2.15	1.99	2.40	2.21	2.20	2.28
LS/LW	1.02	1.01	0.98	0.86	1.17	0.97	1.01	1.04
LS/HMO	1.57	1.64	1.57	1.41	1.82	1.59	1.62	1.67
LS/KS	1.70	1.76	1.56	1.41	1.90	1.61	1.72	1.85
LS/kD	2.20	2.28	2.18	2.02	2.43	2.24	2.22	2.31
HKB/LS	0.46	0.44	0.47	0.51	0.42	0.45	0.46	0.44
HKB/LW	0.47	0.45	0.46	0.43	0.49	0.44	0.47	0.46
HKB/HMO	0.72	0.73	0.73	0.71	0.76	0.72	0.74	0.73
HKB/KS	0.78	0.78	0.74	0.72	0.79	0.73	0.78	0.81
HKB/kD	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01

CONCLUSION

The proposed method for estimating the ridge parameter in ridge regression has been given. The proposed ridge estimator (k_D) for estimating the ridge parameter is based on number data points (n) and strength of multicollinearity in the data. The performance of the proposed ridge parameter is evaluated through the simulation study, by comparing the ratio of average MSE of the proposed ridge estimator (k_D) with other ridge parameter estimators in literature. The performance of the proposed ridge parameter is better than other ridge parameters used in ridge regression.

ACKNOWLEDGMENT

Author would like to thank Prof. Elhammoumi, M. (Edgecombe College, NC,USA) for the English correction and helpful comment.

REFERENCES

D'Ambra, A. and P. Sarnacchiaro, 2010. Some data reduction methods to analyze the dependence with highly collinear variables: A simulation study. Asian J. Math. Stat., 3: 69-81.

Hoerl, A.E. and R.W. Kennard, 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12: 55-67.

Hoerl, A.E., R.W. Kennard and K.F. Baldwin, 1975. Ridge regression: Some simulations. Commun. Stat., 4: 105-123.

Khalaf, G. and G. Shukur, 2005. Choosing ridge parameter for regression problem. Commun. Stat. Theory Methods, 34: 1177-1182.

Lawless, J.F. and P. Wang, 1976. A simulation study of ridge and other regression estimators. Commun. Stat. Theory Methods, 5: 307-323.

Mardikyan, S. and E. Cetin, 2008. Efficient choice of biasing constant for ridge regression. Int. J. Contemp. Math. Sci., 3: 527-536.

Montgomery, D.C., E.A. Peck and G.G. Vining, 2006. Introduction to Linear Regression Analysis. Wiley, New York.

Nomura, M., 1988. On the almost unbiased ridge regression estimation. Commun. Stat. Simulation Comput., 17: 729-743.