International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

An Alternative Approach for Selecting Ridge Parameter for Ordinary Ridge Regression Estimator

Hazim Mansoor Gorgees[1], Fatimah Assim Mahdi[2]

[1, 2]Department of Mathematics, College of Education for Pure Science, Ibn-Al-Haitham, University of Baghdad, Iraq

Abstract: ***In the presence of multicollinearity, the parameter estimation method based on the ordinary least squares procedure is unsatisfactory. In 1970, Hoerl and Kennard introduced alternative method distinguished as ridge regression estimator. In such estimator, ridge parameter or biasing constant plays an important role in estimation. Various methods were suggested by many researchers for choosing the ridge parameter. In this article we employed the concept of condition number to suggest a new method for selecting the ridge parameter. The performance of the proposed method is assessed and compared with other traditional methods through simulation study in terms of mean square error (MSE). The method developed in this paper seems to be reasonable since it has smaller MSE than the other stated methods.***

Keywords: multicollinearty, ridge regression, ridge parameter, singular value decomposition, condition number

1. Introduction

In this article we deal with classical linear regression model:

$$y = X\beta + \varepsilon \quad \dots \quad (1)$$

where

y is(n×1)vector of response variable,

X is (n×p)matrix of explanatory variables and n > p,

βis(p ×1)vector of unknown parameters,

εis(n ×1)vector of unobservable random errors and

$$E(\varepsilon) = 0 \ , \ var(\varepsilon) = \sigma^2 I$$

Currently, a considerable attention is focused on biased estimation of the regression model. This attention is due to the inability of ordinary least squares to provide reasonable point estimates when the matrix of explanatory variables is ill conditioned. Despite possessing the very desirable property of being minimum variance in the class of linear unbiased estimators under the usual conditions imposed on the model, the ordinary least squares estimators can nevertheless, have extremely large variances when the data are inter correlated which is one form of ill conditioning. Much researches; therefore, on obtaining biased estimators with better overall performance than the ordinary least squares estimators were conducted. This paper states the ridge regression estimators as an alternative to the ordinary least squares estimators with multicollineardata. In contrast to ordinary least squares, these estimators allow a small amount of bias in order to achieve a major reduction in the variance.

2. The Case of Multicollinearity

The problem of multicollinearity occurs when there exists an exact linear relationship or an approximate linear relationship among two or more explanatory variables, two types of multicollinearity may be faced in regression analysis, exact and near multicollnearity. During regression calculations, the exact linear relationship causes a division by zero which in turncauses the calculations to be aborted. When the relationship is not exact, the division by zero does not occur and the calculations will not abort. However, the division by a very small quantity still distorts the results. Hence, one of the first steps in regression analysis is to determine if multicollinearity is a problem.

Multicollinearity can be thought of as a situation where two or more explanatory variables in the data set move together, as a consequence it is impossible to use this data set to decide which of the explanatory variables is producing the observed change in the response variable.

Some multicollinearty is nearly always present, but the important point is whether it is serious enough to cause appreciable damage to the regression analysis. Indicators of multicollineaity include a low determinant of the information matrix X'X, a very high correlation among two or more explanatory variables, very high correlation among two or more estimated coefficients, a very small (near zero) eigen values of the correlation matrix of the explanatory variables and the too large condition number. Relationship is existing between two or more independent variables.

3. The Class of Shrinkage Estimators

Applying the singular value decomposition technique we can decompose the matrix X as follows [1]

$$X = H\,\Lambda^{\frac{1}{2}}\,G' \quad \dots \quad (2)$$

where H is$(n \times p)$matrix satisfying $H'H = I_p$, $\Lambda^{\frac{1}{2}}$is a $(p \times p)$diagonal matrix of ordered singular values of X. $\lambda_1^{\frac{1}{2}} \geq \lambda_2^{\frac{1}{2}} \geq \cdots \geq \lambda_p^{\frac{1}{2}} > 0$, G is a$(p \times p)$orthogonal matrix whose columns represent the normalized eigenvectors of X'X.

Consequently, the ordinary least squares estimator of the regression parameters vector β can be rewritten as:

$$b_{OLS} = (X'X)^{-1}\,X'\,Y$$
$$= (G\,\Lambda\,G')^{-1}\,G\,\Lambda^{\frac{1}{2}}H'\,Y$$
$$= G\,\Lambda^{\frac{-1}{2}}H'Y = G\,C$$

Where$C = \Lambda^{\frac{-1}{2}}H'Y = G'\,b_{OLS}$ is the vector of uncorrelated components ofb_{OLS}.

This can be noticed by considering the variance-covariance matrix of C that can be easily shown to equal the diagonal matrix$\sigma^2\Lambda^{-1}$.

The generalized shrinkage estimators denoted by b_{SH} can be defined as:

$$b_{SH} = G\,\Delta\,C = \sum_{j=1}^{p} \vec{g}_j\,\delta_j\,C_j \quad \ldots \quad (3)$$

Where:

$\vec{g}_j$ is the j^{th} column of the matrix G.

δ_j is the j^{th} diagonal element of the shrinkage factorsdiagonal matrixΔ,$0 \le \delta_j \le 1,\ j = 1, 2, \ldots, p$, and$C_j$ is the j^{th} element of the uncorrelated components vector C.

4. Ordinary Ridge Regression Estimators

The most popular method that has been proposed to deal with multicollinearity problem is the ordinary ridge regression.

This method is the modification of ordinary least squares method to allow biased estimators of regression coefficients. The ridge estimators depend crucially upon an exogenous parameter, say k, called the ridge parameter or the biasing parameter of the estimator. For any $k \geq 0$, the corresponding ordinary ridge estimator denoted by$b_{RR}$is defined as:

$$b_{RR} = (X'X + k\,I)^{-1}\,X'\,Y \quad \ldots \quad (4)$$

Where$k \geq 0$is a constant selected by the statistician according to some intuitively plausible criteria put forward by Hoerl and Kennard [2].

It can be shown that the ridge regression estimator given in equation (4) is a member of the class of shrinkage estimators as follows:

By using matrix algebra and singular value decomposition approach we get:

$$\begin{aligned} b_{RR} &= (X'X + k\,I)^{-1}\,X'\,Y \\ &= [G(\Lambda + k\,I)\,G']^{-1}\,G\,\Lambda^{\frac{1}{2}}H'\,Y \\ &= G\,(\Lambda + k\,I)^{-1}\,G'G\,\Lambda^{\frac{1}{2}}H'\,Y \\ &= G\,(\Lambda + k\,I)^{-1}\Lambda^{\frac{1}{2}}H'\,Y \\ &= G\,[(\Lambda + k\,I)^{-1}\Lambda]\Lambda^{\frac{-1}{2}}H'Y = G\,\Delta\,C \quad \ldots \quad (5) \end{aligned}$$

Where: $\Delta = (\Lambda + k\,I)^{-1}\Lambda$

Equivalently, the shrinkage factors $\delta_j, j = 1,2, \ldots,$ pof the ridge estimator has the form:

$$\delta_j = \frac{\lambda_j}{\lambda_j + K} \quad \ldots \quad (6)$$

Where $\lambda_j$is the j^{th} element (eigenvalue) of the diagonal matrixΛ, and K is the ridge parameter.

The mean square error of ordinary ridge regression estimator can easily demonstrated to be[2].

$$MSE(b_{RR}) = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + K)^2} + K^2\beta'(X'X + k\,I)^{-2}\beta \ldots (7)$$

The first term can be shown to be the sum of variances(total variance) of the parameter estimates and the second term can be considered to be the square of the bias introduced when b_{RR} is used instead ofb_{OLS}.

5. Choice of Ridge Parameter

The ordinary ridge regression estimator does not provide a unique solution to the multicollinearity problem, but provide a family of solutions. These solutions depend upon the value of k (the ridge parameter). No explicit optimum value can be found for k. Yet, several stochastic choices have been proposed for this ridge parameter. Some of these choices may be summarized as follows

Hoerl and Kennard (1970).Suggested graphical method called ridge traceto select the value of the ridge parameter k. When viewing the ridge trace, the analyst picks the value of k for which the regression coefficients have stabilized.

Often, the regression coefficients will vary widely for small values of k and then stabilize. We have to choose the smallest value of k(which introduces the smallest bias) after which the regression coefficients have seem to remain constant.

Hoerl, Kennard and Baldwin in (1975),proposed another method to select a single value of K given as[3]

$$\hat{K}_{HKB} = \frac{p\,S^2}{b_{OLS}'\,b_{OLS}} \quad \ldots \quad (8)$$

Where p is the number of explanatory variables, $S^2$is the OLS estimator of$\sigma^2$and $b_{OLS}$is the OLS estimator of the vector of regression coefficients β.

Lawless and Wang (1976)proposed selecting the value of K by using the formula [4]

$$\hat{K}_{LW} = \frac{p\,S^2}{b_{OLS}'\,X'X\,b_{OLS}} \quad \ldots \quad (9)$$

Assuming that the regression coefficients vector has certain prior distribution srivastava followed Bayesian approach to estimate the ridge parameter. He concluded that [5]

$$\hat{K}_{Bayes} = Max\left[0, \frac{tr\,(X'X)}{\left[\frac{n-p-3}{n-p-1}\left(\frac{b_{OLS}'\,X'X\,b_{OLS}}{S^2}\right) - p\right]}\right] \ldots (10)$$

Where $tr\,(X'X)$ denote the trace of the matrix $X'X$.

6. Proposed Method

Our contribution in this topic represented by utilizing the concept of condition number in order to select the ridge parameter. The condition number is defined to be the ratio of the largest to the smallest singular value of the matrix of the explanatory variables X.

The suggested estimator denoted as$\hat{K}_{CN}$ is defined as:

$$\hat{K}_{CN} = Max\left[0, \frac{p\,S^2}{b_{OLS}'\,b_{OLS}} - \frac{1}{CN}\right] \quad \ldots \quad (11)$$

Where CN referred to condition number.

Our proposed estimator is the modification of$\hat{K}_{HKB}$.

The small amount$\frac{1}{CN}$ is subtracted from$\hat{K}_{HKB}$.

This amount, however, varies with the strength of multicollinearity in the model.

If the condition number is too large, then$\hat{K}_{CN}$ would coincide with$\hat{K}_{HKB}$ since in such case, the fraction$\frac{1}{CN}$ would approach to zero.

On the other hand if the condition number is too small (approximately equal to 1) then the possibility that $\left(\frac{p\,S^2}{b_{OLS}{}'\,b_{OLS}} - \frac{1}{CN}\right)$ be negative is too large.

In this case we choose $\hat{K}_{CN}$ to be equal to zero which means that the ridge regression estimator would coincide with the ordinary least squares estimator and the data set is not influenced by the multicollinearity problem.

7. Generalized Ridge Regression

Again, using the singular value decomposition technique in order to derive the generalized ridge regression, we can rewrite the linear regression model as

$$y = X\,\beta + \epsilon = \left(H\Lambda^{\frac{1}{2}}\right)(G'\beta) + \epsilon$$

Or

$$y = Z\,\alpha + \epsilon \quad \ldots \quad (12)$$

Where:

$$Z = H\Lambda^{\frac{1}{2}} \quad , \quad \alpha = G'\beta$$

The model in equation (12) is called the canonical model or uncorrelated components model. The OLS estimator of α is given as

$$\alpha_{OLS} = (Z'Z)^{-1}Z'y = \Lambda^{-1}Z'y \quad \ldots \quad (13)$$

And $Var(\alpha_{OLS}) = \sigma^2(Z'Z)^{-1} = \sigma^2\Lambda^{-1}$ which is diagonal.

This shows the important property of this parameterization since the elements of α_{OLS}, namely, $(\alpha_1, \alpha_2, \ldots, \alpha_p)_{OLS}$ are uncorrelated.

The generalized ridge estimator for α is given by:

$$\alpha_{GRR} = (Z'Z + K)^{-1}Z'y = (\Lambda + \mathrm{K})^{-1}Z'y \quad \ldots \quad (14)$$
$$= (\Lambda + \mathrm{K})^{-1}Z'Z\,\alpha_{OLS} = (\mathrm{I} + \mathrm{K}\,\Lambda^{-1})^{-1}\,\alpha_{OLS}$$
$$W_K\,\alpha_{OLS} = dig\left(\frac{\lambda_i}{\lambda_i + K_i}\right)\alpha_{OLS} \quad , \quad i = 1, 2, \ldots, p$$

Where $K = dig\,(K_1, K_2, \ldots, K_p)$ and:

$$W_K = (\mathrm{I} + \mathrm{K}\,\Lambda^{-1})^{-1} = dig\left(\frac{\lambda_i}{\lambda_i + K_i}\right) \quad , \quad i = 1,2, \ldots, p$$

The mean square error of α_{GRR} is given by:

$$MSE(\alpha_{GRR}) = var(\alpha_{GRR}) + (bias\;\alpha_{GRR})(bias\;\alpha_{GRR})'$$
$$= \sigma^2\,tr(W_K\Lambda^{-1}W_K') + (W_K - I)\,\alpha_{OLS}\,\alpha_{OLS}(W_K - I)'$$
$$= \sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i + K_i)^2} + \sum_{i=1}^{p}\frac{\mathrm{K}_i{}^2 + \alpha^2{}_{i(OLS)}}{(\lambda_i + K_i)^2} \quad \ldots \quad (15)$$

To obtain the value of K_i that minimize $MSE(\alpha_{GRR})$

We differentiate equation (15) with respect to K_i and equating the resultant derivative to zero. Thus

$$\frac{\partial MSE(\alpha_{GRR})}{\partial K_i} = -\sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i + K_i)^3} + \sum_{i=1}^{p}\frac{\lambda_i \mathrm{K}_i\alpha^2{}_{i(OLS)}}{(\lambda_i + K_i)^3} = 0$$

Solving for K_i we obtain $K_i = \frac{\sigma^2}{\alpha^2{}_{i(OLS)}}$

Since the value of σ^2 is usually unknown, we use the estimated value $\hat{\sigma}^2$. Therefore, when the matrix K satisfies

$$\hat{K}_i = \frac{\hat{\sigma}^2}{\alpha^2{}_{i(OLS)}} = dig\left(\frac{\hat{\sigma}^2}{\alpha^2{}_{1(OLS)}}, \frac{\hat{\sigma}^2}{\alpha^2{}_{2(OLS)}}, \ldots, \frac{\hat{\sigma}^2}{\alpha^2{}_{p(OLS)}}\right)$$

Then the MSE of generalized ridge regression attains the minimum value.

The original form of generalized ridge regression estimator can be converted back from the canonical form by

$$b_{(GRR)} = G\,\alpha_{GRR} \quad \ldots \quad (16)$$

8. The Simulation Results

To exhibit multicollinearity in the simulated data, we use different degrees of correlation between the variables included in the model. Specifically, we assume correlation values to be $\rho = 0.70, 0.90\ and\ 0.95$ four predictor variables have been generated. Since the performance of different estimators is influenced by the sample size, we have used three types of samples, small of size 10, median of size 40 and large of size 100. The standard deviations of the error terms are taken as $\sigma = 5, 10\ and\ 20$. Ordinary ridge estimates are computed using different ridge parameters given in equations (8) to (11) and the generalized ridge estimate is obtained from equation (16).

The mean square error (MSE) is used as a criterion in order to assess the performance of the stated methods. This experiment is repeated 1000 times. And the results are presented in tables below

Table 1: The values of MSE at $\rho = 0.70$

n	Method	Standard deviation σ		
		5	10	20
10	G_{RR}	0.0197	0.0227	0.0233
	$\hat{K}_{HKB}$	0.0229	0.0238	0.0235
	$\hat{K}_{LW}$	0.0024	0.0024	0.0024
	$\hat{K}_{Bayes}$	0.0039	0.0039	0.0039
	$\hat{K}_{CN}$	7.3728e-017	7.3728e-017	7.3728e-017
40	G_{RR}	0.0126	0.0125	0.0125
	$\hat{K}_{HKB}$	0.0764	0.0747	0.0742
	$\hat{K}_{LW}$	0.0031	0.0031	0.0031
	$\hat{K}_{Bayes}$	0.1128	0.1134	0.1136
	$\hat{K}_{CN}$	0.0037	5.7051e-018	7.5922e-017
100	G_{RR}	0.0078	0.0078	0.0078
	$\hat{K}_{HKB}$	0.0355	0.0303	0.0273
	$\hat{K}_{LW}$	4.3631e-004	4.3625e-004	4.3623e-004
	$\hat{K}_{Bayes}$	0.0436	0.0438	0.0439
	$\hat{K}_{CN}$	0.0012	3.8272e-004	3.9968e-017

Table 2: The values of MSE at $\rho = 0.90$

n	Method	Standard deviation σ		
		5	10	20
10	G_{RR}	0.0192	0.0226	0.0232
	$\hat{K}_{HKB}$	0.0227	0.0238	0.0235
	$\hat{K}_{LW}$	0.0024	0.0024	0.0024
	$\hat{K}_{Bayes}$	0.0039	0.0039	0.0039
	$\hat{K}_{CN}$	1.0533e-017	6.6706e-017	6.6706e-017
40	G_{RR}	0.0126	0.0125	0.0125
	$\hat{K}_{HKB}$	0.0766	0.0748	0.0742
	$\hat{K}_{LW}$	0.0031	0.0031	0.0031
	$\hat{K}_{Bayes}$	0.1127	0.1134	0.1136
	$\hat{K}_{CN}$	0.0042	2.7648e-017	4.9591e-017
100	G_{RR}	0.0078	0.0078	0.0078
	$\hat{K}_{HKB}$	0.0359	0.0307	0.0274
	$\hat{K}_{LW}$	4.3632e-004	4.3625e-004	4.3623e-004
	$\hat{K}_{Bayes}$	0.0435	0.0438	0.0439
	$\hat{K}_{CN}$	0.0013	4.7300e-004	2.8866e-017

Table 3: The values of MSE at $\rho = 0.95$

n	Method	Standard deviation σ		
		5	10	20
10	G_{RR}	0.0185	0.0225	0.0232
	$\widehat{K}_{HKB}$	0.0225	0.0237	0.0235
	$\widehat{K}_{LW}$	0.0024	0.0024	0.0024
	$\widehat{K}_{Bayes}$	0.0039	0.0039	0.0039
	$\widehat{K}_{CN}$	1.0533e-016	0	0
40	G_{RR}	0.0126	0.0125	0.0125
	$\widehat{K}_{HKB}$	0.0768	0.0749	0.0743
	$\widehat{K}_{LW}$	0.0031	0.0031	0.0031
	$\widehat{K}_{Bayes}$	0.1126	0.1134	0.1136
	$\widehat{K}_{CN}$	0.0047	5.3979e-017	4.8713e-017
100	G_{RR}	0.0078	0.0078	0.0078
	$\widehat{K}_{HKB}$	0.0363	0.0311	0.0276
	$\widehat{K}_{LW}$	4.3633e-004	4.3625e-004	4.3623e-004
	$\widehat{K}_{Bayes}$	0.0435	0.0438	0.0439
	$\widehat{K}_{CN}$	0.0013	5.5274e-004	1.7764e-017

9. Conclusions

Our proposed method for estimating the ridge parameter depends upon the level of multicollinearity between the explanatory variables. This method shows the importance of the condition number as an indicator of the presence of multicollinearity problem.

Moreover, the simulation results imply that the proposed method performs well in the sense of MSE. It seems to be better than other studied methods in all conditions of multicollinearity levels, sample sizes and the standard deviations of the error terms.

References

[1] Gorgees, H.M., (2009) "Using Singular Value Decomposition Method for Estimating the Ridge Parameter", Journal of Economic and Administrative Science, Vol.5,No-15, PP. 1-10.
[2] Hoerl, A. E. and kennard, R.W., (1970), "Ridge Regression: Biased Estimation of Nonorthogonal Problems", Techno metrics, Vol. 12, PP. 55-67.
[3] Hoerl, A.E., Kennard, R. W. and Baldwin, K.F., (1975), "Ridge Regression: some simulation" Communications in Statistics, Vol. 4, PP. 105-123.
[4] Lawless, J.F and Wang, p., (2005), "A simulation study of Ridge and other Regression Estimators", Communications in Statistics - theory and Methods Vol.34, PP. 1177-1182.
[5] srivastava, M.S, (2002), "Methods of Multivariate Statistics", Wiley, New York.
[6] Gogess, H. M and Ali, B. A., (2013), "Employing Ridge Regression Procedure to Remedy the Multicollinearity problem", Ibn Al-Hatham Journal for Pure and Applied science, Vol. 26, No.1, PP. 320-327.