Journal of Advances in Mathematics and Computer Science

34(5): 1-9, 2019; Article no.JAMCS.53467

ISSN: 2456-9968

(Past name: British Journal of Mathematics & Computer Science, Past ISSN: 2231-0851)

Performance of the New Ridge Regression Parameters

Mowafaq Muhammed Al-Kassab[1*] and Mohammed Qasim Al-Awjar[2]

[1]Department of Mathematics Education, Faculty of Education, Tishk International University, Iraq.
[2]Department of Statistics and Informatics, College of Computers and Mathematics, Mosul University, Iraq.

Authors' contributions

This work was done in collaboration with two authors. Author MMAK wrote the sections 1, 2 and 3 whereas author MQAA wrote the sections 4 and 5. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JAMCS/2019/v34i530225
Editor(s):
(1) Dr. Metin Basarir, Professor, Department of Mathematics, Sakarya University, Turkey.
Reviewers:
(1) Oyindamola B. Yusuf, University of Ibadan, Nigeria.
(2) Donald E. Ramirez, University of Virginia, USA.
Complete Peer review History: http://www.sdiarticle4.com/review-history/53467

Original Research Article

Received: 20 October 2019
Accepted: 26 December 2019
Published: 01 January 2020

Abstract

A new approach is presented to find the ridge parameter k when the multiple regression model suffers from multicollinearity. This approach studied two cases, for the value k, scalar, and matrix. A comparison between this proposed ridge parameter and other well-known ridge parameters evaluated elsewhere, in terms of the mean squares error criterion, is given. Examples from several research papers are conducted to illustrate the optimality of this proposed ridge parameter k.

Keywords: Least squares; multicollinearity; ridge parameters; scalar; vector; matrix; mean squared error.

1 Introduction

Multiple linear regression is a well-known method for studying the relationship between dependent variable and explanatory variables. The mathematical from for this regression is:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i,$$

$i = 1, 2, \ldots, n$

**Corresponding author: E-mail: mowafaq.muhammed@ishik.edu.iq, mowafaq.muhammed@tiu.edu.iq, Mowafaq2002@yahoo.co.uk;*

where n is the number of observations. The matrix form of this regression is:

$$y = X\beta \ + e \tag{1}$$

where y is a nx1 vector of response variables, X is a known n x p matrix of the explanatory variables with full rank, e is a nx1 vector of errors with zero mean and var-cov. matrix $\sigma^2 I_n$. β is a px1 vector of unknown regression coefficients. For simplicity we assume that the matrix X and the vector y are standardized such that the matrix $X^T X$ is the correlation matrix of the explanatory variables and $X^T Y$ is the correlation between the matrix X and the vector Y. The following ordinary least squares (OLS) estimators are obtained by minimizing the sum of the squares of error $e^T e$.

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y \tag{2}$$

where $\hat{\beta}_{OLS}$ is an unbiased estimate for β. This ordinary least squares estimator works with some assumptions such as linearity independent identically distributed errors with zero mean and constant variance, homoscedasticity.

In regression analysis, when two or more regressed variables are linearity related it is called multicollinearity. Multicollinearity problem tends to produce OLS estimates that are unstable, having wrong sign coefficients, raises the value of the variance of the coefficient estimates and it makes it more difficult to specify the correct model. Therefore, alternative methods have been proposed to overcome the problem of multicollinearity. Ridge regression method is introduced [1] to overcome the problem of multicollinearity. The paper is organized such that the ridge regression and the ridge parameter are presented in section two. The proposed ridge parameters are given on section three. Application of this proposed ridge parameters on three examples from researched papers is given in section four and section five end by conclusion.

2 Ridge Regression

Ridge regression method is one of the most popular methods that has been proposed by [1]. It is obtained by adding a small positive number k to the diagonal elements of the matrix $X^T X$, so the ridge regression estimator will be:

$$\hat{\beta}_R = (X^T X + k\, I_P)^{-1} \tag{3}$$

where $k \geq 0$ is known as the ridge (or the biased) parameter and estimated from the studied data. [1] showed that, for $k \geq 0$, the ridge regression estimator provides a smaller mean squares error (MSE) than the least squares estimator. The most important problem is to find the value of the ridge parameter k, much discussion, concerning the problem of finding a good value for k has been held. Many different techniques have been proposed or suggested by various researchers. Some of them are, [1-29], and very recently, [30], among others.

All of the above mentioned technique treat the ridge parameter k as a scalar and calculate its value depending on the estimated mean squares error (MSE) and (or) the eigenvalues of the $X^T X$ matrix, maximum value, variance inflation factor (VIF) and the estimated values of β_{OLS} ($\hat{\beta}_i$, $\hat{\beta}_{max}$, $\hat{\beta}^T \beta$).

3 Proposed Estimator for Ridge

In this article, we present a new method to estimate the ridge parameter k. As we mentioned earlier that most of the researcher depending on the ridge parameter k as a constant value. In this section, we will consider the ridge parameter k as a scalar, and as a matrix.

3.1 Case one k is constant

In this case, we assume that the ridge parameter k is constant. Equation (3) which represent the ridge regression estimator is:

$$\widehat{\beta}_R = (X^T X + k\, I)^{-1}\, X^T Y$$

which can be written as:

$$k\, \widehat{\beta}_R \;=\; X^T Y - \;(X^T X\,)\; \widehat{\beta}_R$$

or

$$k \;= \frac{1}{\sum_{i=1}^{p} \widehat{\beta}_{iR}^{\;2}}\,[\,\widehat{\beta}_R^T\, \widehat{\beta}_R\, X^T Y - \,\widehat{\beta}_R^T\, (X^T X\,)\; \widehat{\beta}_R\,]$$

To find the value of the ridge parameter k, we suggest to use the eigenvalues $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_p$ of the $X^T X$ matrix instead of the ridge regression estimator $\widehat{\beta}_R$, so that:

$$\widehat{k}_R \;= \frac{1}{\sum_{i=1}^{p} \lambda_i^2}\,[\,\lambda^T\, X^T Y - \lambda^T\, (X^T X\,)\lambda\,] \tag{4}$$

Substituting $\widehat{k}_R$ from equation (4) instead of k in equation (3), we get

$$\widehat{\beta}_R^{*} = (X^T X + \widehat{k}_R\, I_P)^{-1}\, X^T Y \tag{5}$$

Note that $\widehat{\beta}_R^{*}$ depends on the correlation between X and Y, and the eigenvalues of $X^T X$.

3.2 Case 2 K is a matrix

Assuming that the ridge parameter K is a vector, from equation (3), we get

$$(X^T X + K\,)\; \widehat{\beta}_R = \; X^T Y \tag{6}$$

$$K\, \widehat{\beta}_R = \; X^T Y - \;(X^T X)\, \widehat{\beta}_R$$

We suggest to use the vector λ of the eigenvalues of $X^T X$ matrix instead of the ridge regression estimator $\widehat{\beta}_R$.

i.e.

$$K\, \lambda \;=\; X^T Y - \;(X^T X)\, \lambda$$

We use Moore - Penros inverse [31] to find the inverse of λ which states that:

$$\lambda^{-1} = (\,\lambda^T \lambda\,)^{-1}\, \lambda^T = \frac{1}{\lambda^T \lambda}\, \lambda^T \quad \lambda \neq\; 0 \tag{7}$$

which implies that

$$K^* = \left(X^T Y - X^T X \lambda \right) \lambda^{-1} \quad (8)$$

Substituting equation (8) in equation (6), we get:

$$\hat{\beta}^*_R = (X^T X + K^*)^{-1} X^T Y \quad (9)$$

Notice that K^* in eqution (9) can also calculated from $X^T X$:
a: as a diagonal matrix depending on the eigenvalues of $X^T X$.
b: or a matrix depending on the values of the eigenvector of $X^T X$.

The two methods for calculating K^* in equation (9) in addition to equation (4) will be used in the next section.

4 Examining the Performance of the Proposed Ridge Parameter

In this section, we give some numerical examples to illustrate the performance of our proposed estimate for the ridge parameter k.

4.1 Example 1

We consider the data set on Portland Cement originally due to [32] which has since then been widely analyzed by [33,34,35,36,37]. These data come from an experimental investigation of the heat evolved during the setting and hardening of Portland cements for various compositions. The dependence of this heat is on the percentages of four compounds, these compounds are tricalcium aluminate: 3CaO. Al_2O_3, tricalcium silicate: 3CaO.SiO_2, tetracalecium aluminoferrite: 4CaO. Al_2O_3 .$Fe_2O_3$and β -dicalcium silicate: 2CaO.SiO_2, denoted by x_1, x_2, x_3 and x_4 respectively. The heat evolved after 180 days of curing, denoted by y, is measured in calories per gram of cement. To assess the effects of the four compounds, the following multiple linear regression model is assumed to specify the relationship between the normalized response variable y and the normalized explanatory variables X_1, X_2, X_3 and X_4:

$$\hat{y}_i = \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \hat{\beta}_4 x_{i4} \quad (10)$$

The following table gives the estimated values of the regression coefficient using equations (2), (4) and (6) for ordinary least squares and the proposed ridge parameters with their MSE and the values of K.

Table 1. Estimated values of the regression coefficient with the MSE for OLS and the proposed ridge parameter

Methods	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	MSE
O.L.S.	0.6065	0.5277	0.0434	-0.1603	0.0264
Proposed ridge parameter constant	0.3885	0.5904	-0.4423	-0.5241	0.03512
Proposed ridge parameter vector	0.5014	0.2870	-0.0669	-0.4165	0.00224
Proposed ridge parameter matrix	-0.8799	2.8113	0.2969	-0.0021	0.73174

From Table 1 we see that the estimated MSE for the proposed ridge regression K vector is smalls than the proposed ridge parameter (k scalar and K matrix), and it is also smaller than the O.L.S.

4.2 Example 2

We considered the data obtained from Tagi gas filling company for the time period 2008 – 2016 [38]. To assess the effects of the four explanatory variable X_1, X_2, X_3 and X_4 on the response variable Y, where Y represents the annual output of liquid gas cylinders and the explanatory variables refer to craftsman, administrators, technicians and engineers respectively. [38] used the Tagi gas filling company, data to compare the performance of the following ridge parameters:

$$b_{ORR} = (X^T X + KI \ \ )^{-1} X^T Y$$

$$\widehat{K}_{HKB} = \frac{P \ \ S^2}{b_{OLS}^T \ \ \ b_{OLS}}$$

$$\widehat{K}_{LW} = \frac{P \ \ S^2}{b_{OLS}^T \ X^T X \ \ b_{OLS}}$$

$$\widehat{K}_{Bays} = Max[0 \, , \quad \frac{tr(X^T X \,)}{\left[\frac{n-p-3}{n-p-1} \left(\ \frac{b_{OLS}^T \, X^T X \ \ b_{OLS}}{S^2} \ \right) - p \right]}]$$

and

$$\widehat{K}_{CM} = Max \, [\, 0 \, , \quad \frac{P \ \ S^2}{b_{OLS}^T \ \ \ b} - \frac{1}{CN} \ \ ]$$

To illustrate the performance of our proposed ridge regression parameter, Table 2 gives the MSE of this data using: ordinary least squares, the compared procedures given by [38], and our proposed ridge regression parameter.

Table 2. Estimated mean squares error MSE for the O.L.S. given by [38] and our proposed ridge regression K

MSE	Methods							
	O.L.S.	$\widehat{K}_{HKB}$	$\widehat{K}_{LW}$	$\widehat{K}_{Bays}$	$\widehat{K}_{CN}$	**Proposed ridge K**		
						Constant	**Vector**	**Matrix**
	0.051646	0.071295	0.157356	0.173969	0.099103	0.03358	0.015783	0.24745

From Table 2, we see that the estimated mean squares error MSE for the proposed ridge regression K (vector) is the smallest value.

4.3 Example 3

The following regression model is fitted to the data in which the number of persons employed y are regressed on five predictor variables

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \, x_{i1} + \hat{\beta}_2 \, x_{i2} + \hat{\beta}_3 \, x_{i3} + \hat{\beta}_4 \, x_{i4} + \hat{\beta}_5 \, x_{i5}$$

Where:

X_1: is the land cultivated (million hectares)
X_2: is the inflation rate %, X_3 is the number of establishment, X_4 is the population (million) and X_5 is the literacy % [39].

The following table gives the estimated values of the regression coefficients, using equations (2), (4) and (6) for ordinary least squares and the proposed ridge parameter with their mean squares error and the values of K.

Table 3. Estimated values of the regression coefficients with the mean squares error

Methods	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	**MSE**
O.L.S.	0.0333	-0.03837	-0.1817	1.1328	0.0322	0.0130
Pasha & Shah						
K = 0.055	0.1870	0.0501	0.1203	0.4702	0.18812	2.5006
K_{Lw}= 0.4233	0.211	0.088	0.182	0.257	0.216	1.366
Proposed ridge parameter k						
Constant	0.255194	0.342032	0.257692	0.261923	0.250528	0.0060518
Vector	0139150	0.018756	0.110335	0.498856	0.241100	0.0007416
Matrix	1.88264	0.10350	0.01497	-0.04181	-0.01050	0.0424754

Where

$$K_{LW} = \frac{P \ \sigma^2}{\Sigma \lambda_i \ \ g^2{}_i} \qquad [5]$$

g = $\beta^T \beta$, λ_i are the eginvalues

We see, from Table 3, that the estimated mean squares error for the proposed ridge parameter K when it is a vector is the smallest value.

5 Conclusion

In this paper, we found a new formula to obtain the ridge regression parameter k, we studied two cases for k, scalar and matrix from. We have compared the proposed ridge parameter to well known ridge parameters through three data sets and demonstrated, using two case studies, the improvements in MSE using our the proposed ridge parameter for the case k is a vector.

Acknowledgements

The authors are thankful to the referees for their valuable comments which certainly improved the quality of the paper. The first author gratefully acknowledges the support provided by Tishk International University.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Horel AE, Kennard RW. Ridge regression: Biased estimation for non orthogonal problems. Technometrics. 1970;12:55-68.
DOI: 10.1080/00401706.1970.10488634

[2] Horel R, Kennard W, Baldwin KF. Ridge regression: Some simulation. Commun. Statist. 1975;4:105-123.
DOI: 10.1080/ 03610927508827 232

[3] McDonald GC, Galarneau DI. A Monte Carlo evaluation of some ridge type estimators. J. Amer. Statist. Assoc. 1975;20:407-416.
DOI: 10.1080/01621459.1975.10479882

[4] Hocking F, Speed M, Lynn MG. A class of biased estimators in linear regression. Technometerics. 1976;18(4):425-437.
DOI: 1080/00401706.1976.10489474

[5] Lawless JF, Wang P. A simulation study of ridge and other regression estimators. Commun. Statist – Theory and Methods. 1976;14:1589-1604.

[6] Dempster AP, Schatzoff M, Wermuth N. A simulation study of alternatives to ordinary least squares. Journal of The American Statistical Association. 1977;72:77–91.
DOI: 10.1080/ 01621459.1977.10479910

[7] Gunst RF, Mason RL. Biased estimation in regression: An evaluation using mean squared error. Journal of the American Statistical Association. 1977;72:616-628.
DOI: 1080/01621459.1977.10480625

[8] Wichern DW, Churchill GA. A comparison of ridge estimators. Technometrics. 1978;20:301-311.
DOI: 10.1080/00401706.1978.10489675

[9] Hemmerle WJ, Brantel TF. Explicit and constraint generalized ridge estimator. Techno Metrics. 1978;2:109-120.
DOI: 10.1080/00401706.1978.10489634

[10] Lawless JF. Ridge and related estimation procedure. Communications in Statistics. 1978;A7:139-164.

[11] Gloub M. Helth, Wahaba G. Generalized cross validation as method for choosing a good ridge parameter. Techno Metrics. 1979;21:215-223.
DOI: 10.1080/00401706. 10489751

[12] Gibbons DG. A simulation study of some ridge estimators. Journal of the American Statistical Association. 1981;76: 131–139.
DOI: 10. 1080/ 0162149.1981 .10477619

[13] Nordberg L. A procedure for determination of a good ridge parameter in linear regression. Communications in Statistics. 1982;A11:285-409.

[14] Saleh A. K. Md., Kibria BMG. Performance of some new preliminary test ridge regression estimators and their properties. Communication in Statistics – Theory and Methods. 1993;22:2747–2764.
DOI: 10.1080/03610929308831183

[15] Kibria BMG. On preliminary test ridge regression estimators for linear restriction in a regression model with non – normal disturbances. Communications in Statistics. 1996;a25:2349-2369.
DOI: 10.1080/03610929608831843

[16] Haq MS, Kibria BMG. A shrinkage estimator for the restricted liner regression model: Ridge regression approach. Journal of Applied Statistical Science. 1996;3(4):301-316.

[17] Singh S, Tracy DS. Ridge regression using scrambled responses. Metrika. 1999;147-157.

[18] Wencheko E. Estimation of the signal-to – Noise in the linear regression model. Ststistical Papers. 2000;41:327-343.
DOI: 10.1007/BF02925926

[19] Kibria BMG. Performance of some new ridge regression estimators. Commun. Statist – Simulation and Computation. 2003;32(2):419-435.

[20] Khalaf G, Shukur G. Choosing ridge parameter for regression problem. Commun. Statist – Theory and Methods, Methods. 2005;34:1177-1182.

[21] Zhang J, Ibrahim M. A simulation study on SPSS ridge regression and ordinary least squares regression procedures for multicollinearity data. Journal of Applied Statistics. 2005;32:571-588.
DOI: 10.1080/02664760500078946

[22] Alkhamisi MA, Khalaf G, Shakur G. Some modifications for choosing ridge parameter. Communications in Statistics. 2006;35:1-16.

[23] Alkhamisi MA, Shukur G. A Monte Carlo study of ridge parameters commun. Statist. Simulation and Computation. 2007;36 (3):535-547.
DOI: 10.1080/03610910701208619

[24] Mardikyan S, Cetin E. Efficient choice of biasing constant for ridge regression. Int. J. Contemp. Math. Science. 2008;3(11):527-547.

[25] Batah FS, Ramnathan T, Gore SD. The efficiency of modified Jackknife and its applications. 2008;24(2):157-174.

[26] Muniz G, Kibria BMG. On some ridge regression estimators: An empirical comparison. Communications in Statistics-Simulation and Computation. 2009;38:621-630.
DOI: 10.1080/03610910802592838

[27] Mansson K, Shukur G, Kibria BMG. A simulation study of some ridge regression estimators under different distributional assumptions. Communications in Statistics-Simulations and Computations. 2010;39(8):1936-1670.
DOI: 10.1080/03610918.2010.508862

[28] Dorugade AV, Kashid DN. Alternative method for choosing ridge parameter for regression. International Journal of Applied Mathematics Sciences. 2010;4(9):447-456.

[29] Al-Hassan Y. Performance of new ridge regression estimators. Journal of the Association of Arab Universities of Basic Applied Science. 2010;9:23-26.

[30] Muniz G, Kibria BMG, Kristofer M, Shukur G. On developing ridge regression parameters: A graphical investigation. Sort-Statistics and Operations Research Transactions. 2012;36(2):115-138.

[31] Campbell SL, Meyer Jr. CD. Generalized inverses of linear transformations. Dver. ISBH 978 - 0 – 486 – 66693 – 8; 1991.

[32] Woods Hubert: Steinour, Harold H, Starke Howard R. Effect of composition of portland cement on heat evolved during hardening. Industrial and Engineering Chemistry. 1932;24:1207-1214.

[33] Hald Anders. Statistical theory with engineering applications. Wiley, New York; 1952.

[34] Hamaker HC. On multiple regression analysis. Statistica Neerlandica. 1962;16:31-56.

[35] Gorman JW, Toman RJ. Selection of variables for fitting equation to data. Technometrics. 1966;8:27-51.

[36] Daniel Cuthbert, Wood Fred S. Fitting equations to data: Computer analysis of multifactor data. Second Edition with the Assistance of John W. Gorman. Wiley. New York; 1980.

[37] Nomura Masuo. On the almost unbiased ridge regression estimator. Communications in Statistics – Simulation and Computation. 1988;17:729-743.

[38] Gorgees HM, Mahdi FA. Using the ridge regression procedures to estimate the multiple linear regression coefficient. Journal of Physics, Conference Series 1003; 2018. DOI: 10.1088/1742–6596/1003/1/0/2049

[39] Pasha GR, Shah MAA. Application of ridge regression to multicollinear data. Journal of Research (Science). Bahauddin Zakariya University, Multan, Pakistan. 2004;15(1)97–106.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/53467