Asian Journal of Advanced Research and Reports

7(2): 1-8, 2019; Article no.AJARR.52893
ISSN: 2582-3248

Modified Ridge Regression Estimator with the Application of Peanut Production in Pakistan

Asifa Mubeen[1], Nasir Jamal[1], Muhammad Hanif[1] and Usman Shahzad[1*]

[1]Department of Mathematics and Statistics, Pir Mehr Ali Shah Arid Agricultural University, Rawalpindi, Pakistan.

Authors' contributions

This work was carried out in collaboration among all authors. Author AM designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Authors NJ and MH managed the analyses of the study. Author US managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJARR/2019/v7i230172
Editor(s):
(1) Dr. Longan Shang, Professor, School of Biotechnology and Chemical Engineering, Ningbo Institute of Technology, Zhejiang University, Zhejiang, PR China.
(2) Dr. Naeem Khan, Department of Plant Sciences, Quaid-I-Azam University, Islamabad, Pakistan.
(3) Dr. Him Lal Shrestha, Associate Professor, Coordinator - UNIGIS Programme, Kathmandu Forestry College, Koteshwor, Kathmandu, Nepal.
Reviewers:
(1) Philbert Modest Luhunga, Tanzania.
(2) Essam Fathy Mohamed El-Hashash, Al-Azhar University, Egypt.
(3) Huichen Jiang, Beihang University, China.
Complete Peer review History: http://www.sdiarticle4.com/review-history/52893

Original Research Article

Received 19 September 2019
Accepted 21 November 2019
Published 30 November 2019

ABSTRACT

The main objective of the present study was to develop a new ridge regression estimator and fit the ridge regression model to the peanut production data of Pakistan. Peanut production data has been used to analyze the results. The data has been taken peanut production and growth rate of Pakistan. The mean square error of the proposed estimator is compared with some existing ridge regression estimators. In this study, we proposed a ridge regression estimator. The properties of proposed estimators are also discussed. The real data set of peanut production is used for assuming the performance of proposed and existing estimators. Numerical results of real data set show that proposed ridge regression estimator provides best results as compare to reviewed ones.

Keywords: Regression; ridge regression; peanut production.

**Corresponding author: Email: usman.stat@yahoo.com;*

1. INTRODUCTION

Francis and Dickson [1] was firstly suggested the term 'regression' in "family likeness in stature" study. Which is degree of relationship among the average value of one variable with the other variable. The regression line with one independent variable is known as simple Linear regression (LR) model. When the independent variable is more than one then it is known as multivariate LR model. In some complicated cases there is a nonlinear regression models. The regression models and regression analysis are widely use for prediction and forecasting.

Linear regression analysis is one of the most popular statistical procedure. Linear model is frequently use for regression analysis. However, in such conditions, uncertainty the correlation among independent and dependent variable is not linear then some transform models should be used such as quadratic, exponential and log [2].

The ridge regression is the modified form of regression. The term ridge regression was firstly introducing by [3,4]. In case of more than one independent variable shows that the parameter estimators are depends upon MMSE with having big amount of probability of unacceptable. They suggested an evaluation method which is based on add up of some positive quantity in diagonals of $X'X$. Presented by the ridge trace, the procedure for figure out by two sizes the properties of non-orthogonality. And described that how $X'X$ is calculated biased estimator with minimum mean square errors.

2. REVIEW OF LITERATURE

Prais and Winsten [5] suggested the ridge regression estimator by the using the roh ($\hat{\rho}$) in ridge parameter. As the $\hat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_t^2}$, whereas the e_t are the residuals. The ridge parameter of Prais and Winsten is $\hat{\beta}_{(p\&w)} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$, where Ω is define as roh (ρ). The e_t and e_{t-1} are the auto correlated residuals.

Hoerl et al. [6] introduced the new ridge regression estimator by utilizing the [5] estimator and the variance term. It can be written as $\hat{\beta}_{(hkb)} = (X'\hat{\Omega}^{-1}X^{-1} + k_1 I_{p\times p})^{-1}X'\hat{\Omega}^{-1}Y$, whereas Ω is define as roh (ρ), same as in Prais and Winston. The K is define as $k_1 = \rho s^2 / \hat{\beta}'_{(p\&w)}\hat{\beta}_{(p\&w)}$, as ρ is calculated from auto correlated errors, and s^2 is the variance, and $\hat{\beta}_{(p\&w)}$ is the ridge estimator of Prais and Winston. As the ridge estimator of HKB gives the less mean square error as compared to the Prais and Winston 's ridge parameter's MSE. As the variance term is calculated in the form of

$$s^2 = \frac{\left\{y - X\hat{\beta}_{(p\&w)}\right\}' \Omega^{-1} \left\{y - X\hat{\beta}_{(p\&w)}\right\}}{n-p}$$

Where $\hat{\beta}_{(p\&w)}$ is the ridge parameter of [5] and Ω is define as $\hat{\rho}$, n is the number of observation and p is the number of variables.

Lawless and Wang [7] proposed the new form of ridge regression estimator which can be define as the generalized Lawless wang ridge and it can be written as

$$\hat{\beta}_{(lw)} = (X'\hat{\Omega}^{-1}X^{-1} + k_2 I_{p\times p})^{-1}X'\hat{\Omega}^{-1}Y$$

$$= (X'\hat{\Omega}^{-1}X^{-1} + k_2 I_{p\times p})^{-1}X'\hat{\Omega}^{-1}X\hat{\beta}_{(p\&w)}$$

With having the value of k_2 is

$$k_2 = \rho s^2 / \left\{\hat{\beta}_{(p\&w)}\right\}' X'\Omega^{-1}X\left\{\hat{\beta}_{(p\&w)}\right\}$$

Here k_2 is the ridge parameter of lawless and wang (1976), and in k_2 the ρ is the roh hat, $\hat{\beta}_{(p\&w)}$ is the ridge parameter of [5] with dependent variable and transpose of dependent variable matrix, as the ridge parameter of lawless and wang gives the mininmum MSE as compared to the [5] and [6] ridge regression estimator.

Hoerl and Kennard [3] also suggested the new ridge regression estimator which is define as generalized general ridge regression estimator and it can be written as

$$\hat{\beta}_{(hk)} = (X'\hat{\Omega}^{-1}X^{-1} + \hat{Q}k_1\hat{Q}')^{-1}X'\hat{\Omega}^{-1}Y$$

$$= (X'\hat{\Omega}^{-1}X^{-1} + \hat{Q}k_1\hat{Q}')^{-1}X'\hat{\Omega}^{-1}\hat{\beta}(p\&w)$$

As the $\widehat{Q}$ is the square matrix of eigen vectors of $X'\widehat{\Omega}^{-1}X$ matrix. And the ridge parameter k_1 is define as $k_1 = s^2 / \widehat{d}^2$, $\widehat{d} = \widehat{q}' \ \widehat{\beta}_{(p\&w)}$, here the $\widehat{q}'$ is the $\widehat{Q}'\widehat{Q} = I$ matrix with having order of matrix according to the dependent variable matrix order. As a result of the [3] are similer to the result of [5] ridge regression estimator.

Acer and Ozkale [8] proposed inspiration trials in ridge regression through auto-correlation residuals. The proposed outcomes demonstrate by mathematically illustration also the monte-Carlo simulation explained that outcomes of ACC.

Adwale and Ayinde [9] reviewed the ridge estimator prediction methods and divide in various kinds and numerous categories. In this paper they studied the different measures such as constant Max, continuous Max, average, geometric mean, harmonic mean and median and many various kinds are root square, reciprocal and reciprocal of root mean square. These measures are used to predict the new predicting methods for ridge estimator. They have proposed new estimator by examine the simulation. the relative efficiency of the predicted procedure which is the ratio between MSE and OLS is compared with the procedures. Outcomes provide better methods than the existing results, generally the ridge estimator is the best procedure in formula of harmonic mean, fix Max and fluctuating more in real and root square kinds.

Nasir and Rind [10] has been worked on ridge estimator by proposing modified form of ridge regression parameter to get the best results about wheat production. The Urea fertilizer, DAP fertilizer and manures performs a substantial part to improve the production rate of wheat yield. The number of cultivation of wheat area is substantial factor to rise the wheat production. They also analyzed that the good rain level in rainy months is meaningfully helpful to raise the wheat production whereas the higher temperature in summer months is reason to reducing the production of wheat crop. The descriptive variables which are used in the model are performed their part for predicting the production of wheat with respect to their previous prospects in agriculture field.

The main determination of present study is to develop a new ridge regression estimator and fit the ridge regression model to the peanut production data of Pakistan. Some others measure are also used to investigation, ridge parameters, coefficients, MS errors and biased.

3. PROPOSED ESTIMATOR

Let us suppose, a general linear model is

$$Y = X\beta + \varepsilon \tag{4.1}$$

Where Y is $n\times 1$ vector of data and dependent variable; X is a complete rank $n\times p$ ($p>n$) non-stochastic matrix of data in p independent variables; β is $p\times 1$ unknown coefficient and ε is $n\times 1$ vector of unknown observation. Such that

$$\varepsilon = \rho\varepsilon_{t-1} + \omega_t \ |\rho| < 1, -1 < \rho < 1 \ , \ t = 1,2,3,\dots,n \tag{4.2}$$

And

$$\omega_t \sim N(0,\sigma^2), \ E(\omega_t\omega_{t'}) = 0, \quad \forall \ \ t,t' \neq t, \tag{4.3}$$

As it is known that in the model of OLS estimator β ,

$$\widehat{\beta} = (X'X)^{-1}X'Y$$

Is unbiased but inefficient. So we have proposed a modified ridge estimator

$$\widehat{\mathrm{B}}_J = (X'X + JI)^{-1}X'Y \tag{4.4}$$

Where as

$$J = \sigma^2 / n\mathrm{B}'\mathrm{B} \ , \tag{4.5}$$

I. Bias

The bias of modified ridge regression estimator is

$$\widehat{B}_J = \left\{X'X + \left(\sigma^2 / n\mathrm{B}'\mathrm{B}\right)I_{p\times p}\right\}^{-1}X'Y \tag{4.6}$$

$$= J.\sigma^2(X'X)^{-1}J'$$

$$= E(\widehat{\beta}'J'J\widehat{\beta}) - E(\beta'J\beta) - E(\beta'J'\beta) + E(\beta'\beta)$$

$$E(\widehat{B}_J)$$

$$= \left[E\left\{I_{p\times p} + X'X\left(\sigma^2 / n\mathrm{B}'\mathrm{B}\right)I_{p\times p}\right\}^{-1}(X'X)^{-1}X'Y\right]$$

So,

Bias ($\widehat{B}_J$) $= \left[\left\{ X'X + \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right\}^{-1} X'Y \right]$ (4.7)

II. Variance of $\widehat{B}_J$

Var ($\widehat{B}_J$) $= Var\left[\left\{ X'X + \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right\}^{-1} X'Y \right]$ (4.8)

Put

J $= \left\{ I_{p\times p} + X'X \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right\}$ (4.9)

$= Var(J\widehat{\beta})$

As by the property

$Var(AY) = A.Var(Y)A'$

So,

Var ($\widehat{B}_J$) $= \left\{ I_{p\times p} + X'X \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right\}$

$= \sigma^2 (X'X)^{-1} \left\{ I_{p\times p} + X'X \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right\}'$ (4.10)

Hence proved.

III. Mean Square Error Of $\widehat{B}_J$

MSE ($\widehat{B}_J$) $= E(\widehat{\beta}_J - \beta)'(\widehat{\beta}_J - \beta)$ (4.11)

$$= E(\widehat{\beta}'J'J\widehat{\beta} - \beta'J\beta - \beta'J'\beta + \beta'\beta)$$

$$= E(\widehat{B}'J'J\widehat{B}) - B'J'JB - B'J'JB + E(B'J'JB)$$
$$- E(B'J'JB) + B'J'JB + B'J'JB$$
$$- B'JB - B'J'B + E(B'B)$$

$$= E(\widehat{B}'J'J\widehat{B}) - E(B'J'JB) + B'J'JB - B'JB - BJB' + B'B$$

$$= E(\widehat{B} - B)'J'J(\widehat{B} - B) + B'J'B(J - I) - B'B(J - I)$$

$$= E(\widehat{B} - B)'J'J(\widehat{B} - B) + B'(J - I)'B(J - I) \quad (4.12)$$

As

$\widehat{B} = (X'X)^{-1} X'Y$

From LR model

$$Y = X\beta + \mu$$
$$= (X'X)^{-1} X'(X\beta + \mu)$$
$$\widehat{B} = \beta + X'\mu(X'X)^{-1}$$
$$\widehat{B} - \beta = X'\mu(X'X)^{-1} \quad (4.13)$$

So,

$E(\widehat{B} - \beta)'(\widehat{B} - \beta)$ is

$$E(\widehat{B} - \beta)'(\widehat{B} - \beta) = E\left\{ X'\mu(X'X)^{-1} \right\}' \left\{ X'\mu(X'X)^{-1} \right\}$$
$$= E\left\{ X'(X'X)^{-1} \right\}' \mu'\mu \left\{ X'(X'X)^{-1} \right\}$$
$$= \left\{ (X'X)^{-1} \right\}' E(\mu'\mu) \quad (4.14)$$

So,

$$= \left\{ (X'X)^{-1} \right\}' \delta_\mu^2$$
$$= E(\widehat{B} - B)'J'J(\widehat{B} - B) + B'(J - I)'B(J - I)$$
$$= \left\{ (X'X)^{-1} \right\}' \delta_\mu^2 J'J + B'(J - I)'B(J - I)$$

$$= \left\{ (X'X)^{-1} \right\}' \delta_\mu^2 \left\{ I_{p\times p} + X'X \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right\}'$$
$$\left\{ I_{p\times p} + X'X \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right\}$$
$$+ B'\left[X'X \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right]' B\left[X'X \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right]$$

So,

MSE ($\widehat{B}_J$) $= \left\{ (X'X)^{-1} \right\}' \delta_\mu^2 \left\{ I_{p\times p} + X'X \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right\}'$

$$\left\{ I_{p\times p} + X'X \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right\}$$
$$+ B'\left[X'X \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right]'$$
$$B\left[X'X \left(\sigma^2 / n\mathrm{B}'\mathrm{B} \right) I_{p\times p} \right] \quad (4.15)$$

4. PEANUT PRODUCTION; A REAL LIFE EXAMPLE

The peanuts are also kwon as groundnut, they are mostly cultivated in rain fed areas of Pakistan, mostly in Punjab, also in some areas of Sindh and North West Frontier Province (NWFP). As the peanuts contain 50% of good oil quality.

The data of peanut production is collected from the website of Pakistan named as index mundi. We take the data of peanuts production in Pakistan. Of the year 1972-2017. As this data having some outlier's values or sequence trend values out of variation. It has 45 observations.

As years represented as y, peanut production is represented as x1 and growth rate of peanut production is represented as x2.

The histogram of the original data is demonstrate as.

The time series model for the present study is

$$Y = X\beta + \varepsilon_t$$

where

$$\hat{\beta}_J = (X'X + JI_{P\times P})^{-1} X'Y$$

And

$$J = \partial^2 / n\left[\hat{\beta}(p\&w)\right]'\left[\hat{\beta}(p\&w)\right]$$

where

$$\partial^2 = \left\{y - X\hat{\beta}(p\&w)\right\}' \Omega^{-1}\left\{y - X\hat{\beta}(p\&w)\right\}$$

$$= \begin{bmatrix} 5.043640e+09 & -1574661750 & 1.152026e+02 \\ -1.574662e+09 & 5535261009 & -1.574662e+09 \\ 1.152026e+02 & -1574661750 & 5.043640e+09 \end{bmatrix}$$

$$J = \partial^2 / n\left[\hat{\beta}(p\&w)\right]'\left[\hat{\beta}(p\&w)\right]$$

$$\hat{\beta}_J = (X'X + JI_{P\times P})^{-1} X'\hat{\beta}(p\&w)$$

$$= \begin{bmatrix} -8.123668 \\ 23.175352 \\ -2.591989 \end{bmatrix}$$

$$\hat{Y}_J = X_o(-8.123668) + X_1(23.175352) + X_2(-2.591989)$$

As shown in the Fig. 2 shows a representation of dot and line plot of proposed ridge regression estimator $\hat{Y}_J$. As this figure shows the structural breaks but it has some variation. It is stationary with some structural breaks. This figure also has some peak points like lowest and highest. The lowest point appears between zero to 10 and the highest peak point appears between the points 30 to 40.

$$\hat{\varepsilon}_{t(J)} = Y - \hat{Y}_J$$

As shown in the Fig. 3 shows a representation of dot and line plot of proposed ridge regression estimator $\hat{Y}_J$. As this figure shows the structural breaks but it has some variation. It is stationary with some structural breaks. This figure also has some peak points like lowest and highest. The lowest point appears between 30 to 40 and the highest peak point appears between the points zero to 10.

$$MSE(\hat{B}_J) = \left\{(X'X)^{-1}\right\}' \delta_\mu^2 \left\{I_{p\times p} + X'X\left(\sigma^2 / n\mathrm{B}'\mathrm{B}\right) I_{p\times p}\right\}'$$
$$\left\{I_{p\times p} + X'X\left(\sigma^2 / n\mathrm{B}'\mathrm{B}\right) I_{p\times p}\right\}$$
$$+B'\left[X'X\left(\sigma^2 / n\mathrm{B}'\mathrm{B}\right) I_{p\times p}\right]'$$
$$B\left[X'X\left(\sigma^2 / n\mathrm{B}'\mathrm{B}\right) I_{p\times p}\right]$$

=241943.5

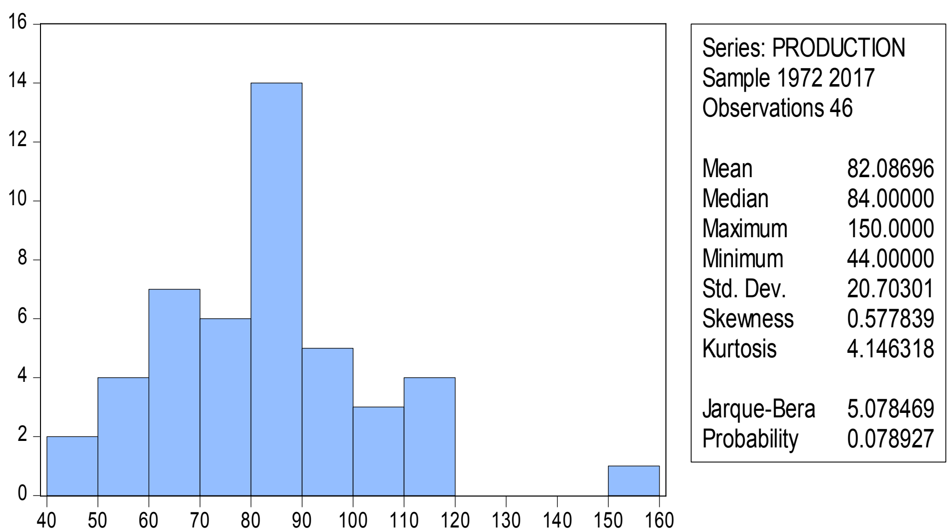

Fig. 1. Shows the data which has been used in this study is time series data, (1972-2017) and has obtained from the website of Pakistan index mandi

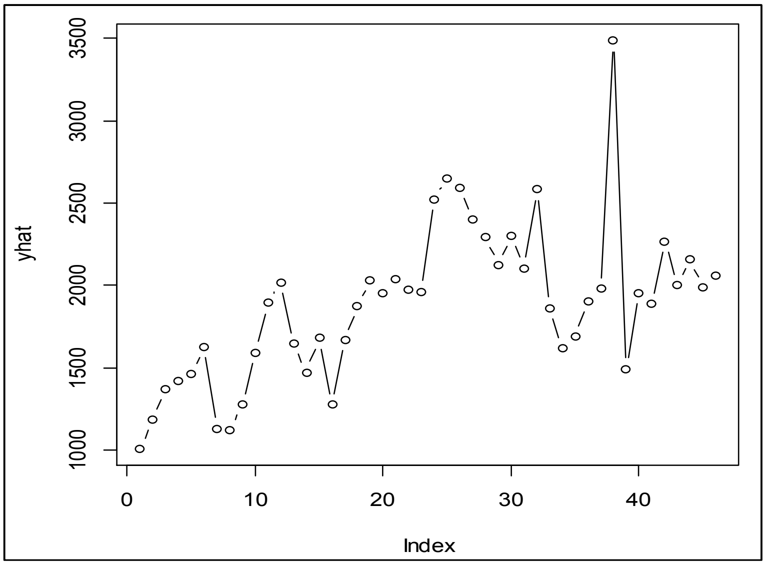

Fig. 2. Plot of proposed ridge regression estimator $\widehat{Y}_J$

As shown in Table 1 the results of $\widehat{B}_J$, The MSE of $\widehat{B}_J$ is 241943.5 which is minimum than all the above MS errors of [5] which is 5043639858, [6] which is 4914310857, [7] which is 4255756 and [3] which is 5043639858. As the proposed estimator provides a best results so it is best estimator for current situation.

Table 1. Results of reviewed and proposed ridge regression estimators

	Ridge estimator	MSE
[5]	$\widehat{\beta}_{(p\&w)} = \begin{bmatrix} -53108.3417 \\ -153.6729 \\ -192.0004 \end{bmatrix}$	5043639854
[6]	$\widehat{\beta}_{(hkb)} = \begin{bmatrix} -52413.8104 \\ -151.3938 \\ -191.1470 \end{bmatrix}$	4914310857
[7]	$\widehat{\beta}_{(lw)} = \begin{bmatrix} -7.024215e-15 \\ -2.524781e-14 \\ -6.167698e-15 \end{bmatrix}$	4255756
[3]	$\widehat{\beta}_{(hk)} = \begin{bmatrix} -53108.3417 \\ -153.6729 \\ -192.0004 \end{bmatrix}$	5043639858
Proposed	$\widehat{\beta}_J = \begin{bmatrix} -8.123668 \\ 23.175352 \\ -2.591989 \end{bmatrix}$	241943.5

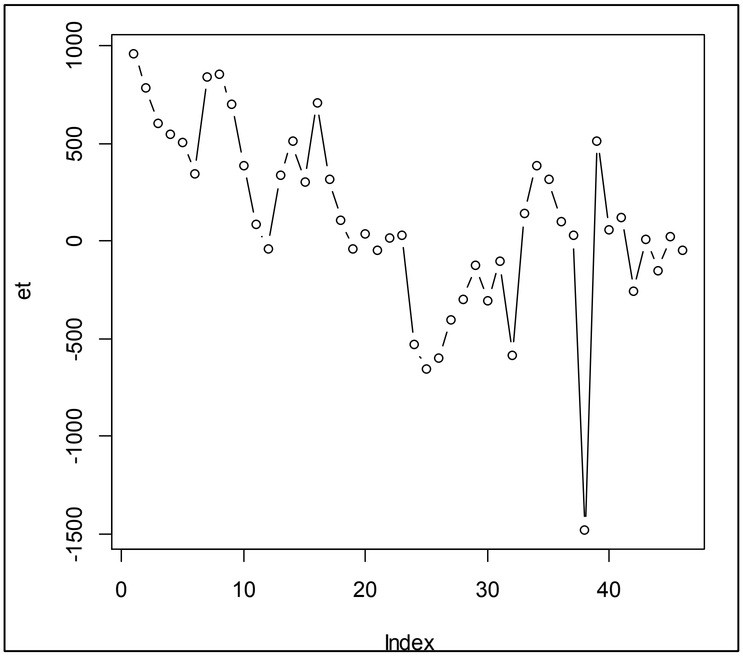

Fig. 3. Errors plot of proposed ridge regression estimator $\hat{\varepsilon}_{t(J)}$

5. CONCLUSION

In this study, we estimate different kind of ridge regression parameters than fit the model and calculate their mean square errors. Peanut production data has been used to analyze the results. The data has been taken peanut production and growth rate of Pakistan. The mean square error of the proposed estimator is compared with some existing ridge regression estimators. As the MSE of the proposed estimator is 2419443.5 which is minimum then the existing estimator of [5,6,7] and [3]. Hence, it is recommended to use proposed estimator.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Francis G, Dickson JH. Family likeness in stature. Proceedings of the Royal Society of London. 1986;242(40):42-73.
2. Yasin N, Karaibrahimoglu A, Genc A. Modified Ridge Regression Parameters: A Comparative Monte Carlo study. Hacettepe Journal of Mathematics and Statistics. 2013;43(5):827-841.
3. Hoerl E, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics. 1970;12(1):55-67.
4. Lius FI. A simulation study of ridge regression estimators with autocorrelated errors. Commun. Statistics. Simulation, Unniversidad de Santiago de chile. Santiago, Chile. 1989;18(2):673-702.
5. Prais S, Winsten C. Trend estimators and serial correlation. Cowles Commission Discussion Paper No. 383. Chicago, IL. Prais Cowles Commission Discussion Paper; 1954.
6. Hoerl AE, Kennard RW, Baldwin KT. Ridge Regression: Some Simulations. Journal of Commun Statist. 1975;4(2):105–123.
7. Lawless JF, Wang AP. Simulation Study of Ridge and Other Regression Estimators. Journal of Communication Statistics. 2010; 5(4):307–323.
8. Acer TS, Ozkale MR. Cross validation of ridge regression estimator in auto correlated linear regression models. Journal of Statistical Computation and Simulation. 2015;12(1).

9. Adwale FL, Ayinde K. Review and classification of the ridge parameter estimation techniques. Hacettepe Journal of Mathmatics and Statistics. 2017;46(5): 953-967.

10. Nasir J, Rind QM. Ridge Regression: A tool to forecast wheat area and production. Pak. J. Stat.Oper.Res. 2007;3(2):125-134.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/52893