International Journal of Engineering Research and Advanced Technology (IJERAT)

DOI: 10.31695/IJERAT.2019.3444

E-ISSN : 2454-6135

Volume.5, Issue 5
May -2019

Robust Regression Methods for Solving Non-Spherical Problem in Linear Regression

Nusirat Funmilayo Gatta[1], Waheed Babatunde Yahya[2] and Mohammed Kabir Garba[3]

[1-,3]Faculty of Physical Sciences

Department of Statistics, University of Ilorin, Ilorin,

Kwara State, Nigeria.

ABSTRACT

This study investigated the effects of non-spherical disturbance on the model parameters of some classical regression models. The aim was to examine the impacts of multicollinearity on the efficiency of classical Ordinary least squares (OLS) relative to the ridge regression (RR) and principal component regression (PCR) models. Data were simulated from a multivariate normal distribution with mean zero and variance-covariance matrixΣ at various sample sizes 25, 50, 100, 200, 500 and 1000. To assess the asymptotic efficiency and consistency of these regression models in the presence of multicollinearity, the evaluation criteria used were the Variance, Absolute bias, Mean Square Error (MSE) and Mean Square Error of Prediction (MSEP). Results from this work showed that the RR model had smaller variance, absolute bias and MSE when it was compared with OLS. Also, the ridge estimator had the least MSEP when compared to both the OLS and PCR models. Hence, it can be concluded that the ridge estimator performed better than the OLS and PCR when explanatory variables are highly correlated.

Keywords: Ordinary least squares, Principal Component Regression, Ridge Regression, Spherical Disturbance, Mean Square Error.

1. INTRODUCTION

Consider the population regression model with two explanatory variables,

$$Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad (1)$$

where $X = (1 \quad X_1 \quad X_2)$ and variance of β is:

$$\mathrm{V}(\beta) = \frac{\sigma^2}{n}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1} \quad (2)$$

The correlation ρ indicates collinearity between X_1 and X_2 , as this correlation approaches 1 the matrix X becomes singular and variance of a coefficient estimate $\sigma^2(1-\rho^2)^{-1}$ approaches infinity. If the predictors are not dependence that is the correlation coefficient for these variables are zeros, the Eigenvalues of the data matrix X are equal to one and matrix X is of full rank such variables are called orthogonal or uncorrelated variables. On the other hand, if the variables are nonorthogonal (correlated), at least one of the Eigenvalue will be close to zero. (see El-Dereny and Rashwan (2011))

Freund and Litell (2000), Batterham *et al* (1997), Wax (1992), Cavell *et al* (1998), Mofenson *et al* (1999), Elmstahl *et al* (1997), Parkin *et al* (2002) and Kinta *et al* (2002) indicated that collinearity leads to imprecise estimate of parameters, increases the estimate of standard error of coefficients, causing wider confidence interval and increasing the chance to reject the significance of the test statistic.

El-Dereny and Rashwan (2011) compare ordinary ridge regression (ORR), generalized ridge regression (GRR), direct ridge regression (DRR) and ordinary least square (OLS). He discussed the properties of ridge regression estimators and method of

selecting biased ridge regression parameter K (shrinkage parameter). He reported that all methods of RR are better than OLS method.

Ranjit (2013) described several methods for detecting multicollinearity that is, by observing the correlation matrix, VIF, Eigenvalues of the correlation matrix. He found out that the degree of multicollinearity is more severe as $/X^1X/$ tends towards zero, and that multicollinearity cannot be eliminated completely but can be reduced by adopting methods such as RR, principal components regression (PCR) etc.

Alabi *et al* (2008) reported that when the independent variables are correlated, the OLS estimates leads to the problem of large standard errors of the parameters which can cause low t-test value and result to acceptance of a null hypothesis.

Feng-jeng (2008) solved the difficult problem of multicollinearity in the fitted regression model and further discussed that the problem of multicollinearity arise when there are approximate linear relationships between two or more predictors. A new estimator for solving multicollinearity problem in terms of parameter estimation known as maximum entropy was developed by Akdeniz (2011).

Gorgees and Ali (2013) applied three different Ridge regressions namely, Ordinary ridge regression (ORR_1) and (ORR_2) and Generalized ridge regression (GRR) on a data set that suffers from multicollinearity problem. Using the standard Mean square Error and coefficient of the determinant (R^2) the result shows that the GRR outperforms the other methods. Dorugade and Kashid (2010) proposed a new method of selecting ridge parameter (turning parameter K) the method is evaluated through a simulation study in term of mean square error and compare the ratio of the average of MSE with the ridge parameter suggested by Hoerl and Kennard and Khalaf and Shukur. It was discovered that the technique developed is better than the other ridge parameters.

Using data set to examine the performance of the three biased regression estimators that is, principal component regression, partial Least squares and Ridge regression on prediction. It is shown that for prediction, PCR, PLS and RR gives the same results. (see Ying 2010). Paramveer *et al* (2013) compares the risk of Ridge Regression to a simple variant of Ordinary Least Square where data are projected onto a finite-dimensional subspace then, performs Ordinary Least Square OLS in the space. The result shows that the risk of Ordinary Least Square is within a constant factor of the Ridge Regression risk.

2. LINEAR REGRESSION MODEL

This study makes use of multiple regression model where n sample observations of a dependent variable Y, explanatory variable X and the relationship between X and Y are observed. The project commences with the case of a K regressors that is, k= 1, 2 …K. and this is written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

Where β_i (i=0, 1, 2 …k) are the regression coefficient and ε is the error term. Thus, the linear equation can be written in matrix form as:

$$Y = X\beta + \varepsilon$$

Where Y is a vector of n× 1 observations of the dependent variable, X is an n× (k+1) matrix of independent variables, β is a (k+1) ×1 vector of unknown parameters and ε is an n×1 vector of errors $\varepsilon \sim N(0, \sigma^2 I)$.

Ordinary Least Squares Method

The least square estimators $\hat{\beta}_i$ of β_i $(i = 0, 1, 2 \ \ldots k)$ are the ones that minimize the sum of squares

$$\hat{\beta}_{ols} = (X'X)\, X'y$$

The variance-covariance matrix of the parameter is

$$V(\beta) = \sigma^2 (X^1 X)^{-1}$$

$$\text{Absolute bias} = \frac{\sum_{i=1}^{k} |\beta_{OLS} - \beta|}{k}$$

DOI : 10.31695/IJERAT.2019.3444

$$\text{MSE} = \frac{\sum_{i=1}^{k}(\beta_i - \beta)^2}{k}$$

$$\text{MSEP} = \frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n}$$

3. RIDGE REGRESSION

Ridge regression is a method of tackling the threat of multicollinearity; RR coefficients β_R are the values of β that minimize a penalized residual sum of squares:

$$\beta_R = \min \left\{\sum_{i=1}^{n}(y - \beta_0 - \sum_{j=1}^{k}\beta_j X_{ij})^2 + \lambda \sum_{j=i}^{k}\beta^2\right\}$$

$$\hat{\beta}_R = (X^1X + \lambda I)^{-1}\ X^1Y$$

According to Hoerl and Kennard, the ridge parameter is:

$$\lambda = \frac{p\sigma^2}{\beta'\beta}$$

where $\sum_{j=i}^{k}\beta^2$ is called the shrinkage penalty and λ is the turning parameter. The shrinkage penalty can shrinks the parameter estimates towards zero but not exactly zero. When the parameter λ is zero, the ridge regression estimate will produce the same estimate as Ordinary Least square estimate. However as λ tends to infinity, the ridge parameter estimate approaches zero. Variance of ridge regression estimate is:

$$V(\hat{\beta}_R) = \sigma^2(X^1X + \lambda I)^{-1}X^1X\ (X^1X + \lambda I)^{-1}$$

$$\text{Absolute bias} = \frac{\sum_{i=1}^{k}|\beta_R - \beta|}{k}$$

$$\text{MSE of ridge} = \frac{\sum_{i=1}^{k}(\beta_R - \beta)^2}{k}$$

$$\text{MSEP} = \frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n}$$

4. PRINCIPAL COMPONENT REGRESSION (PCR)

Principal Component Regression is also one of the techniques used in handling multicollinearity problem, and it requires some mathematical computations that do not exist in Normal regression analysis. This method rewritten the linear regression models in terms of uncorrelated independent variables and the new variables formed as a result of the linear combination of the original independent variables is called principal components. From linear regression model in equation two above, suppose there exist a square matrix G such that $G'G = GG'$ equals identity matrix I (G is called the inverse of G' and the matrix G' is also an inverse G) and

Let $G'X'XG = \wedge$ where $\wedge$ is the diagonal matrix in order of decreasing Eigen values of $X'X$ that is, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_k \geq 0$, therefore equation (2) become

$Y = XGG'\beta + \varepsilon$, since GG' = I

$Y = Z\alpha + \varepsilon$, where

Z= XG= $[Z_1, Z_2, Z_3, \ldots Z_k,]$ these column of Z are called the principal components and $\alpha = G'\beta$ therefore,

$Z\alpha = XGG'\beta$ and $Z'Z = (XG)'(XG) = G'X'XG = G'G \wedge GG'$

The estimate of α will become:

$\hat{\alpha} = (Z'Z)^{-I}\ Z'Y$ or $\hat{\alpha} \wedge^{-I} Y$

DOI : 10.31695/IJERAT.2019.3444

The variance-covariance matrix is:

$V(\hat{\alpha}) = \sigma^2 \wedge^{-I}$

MSEP $= \frac{1}{n}(\hat{Y} - Y)'(\hat{Y} - Y)$

5. DATA GENERATING PROCEDURE

In this research work, five explanatory variables values were generated with the same sample size n**.** The sample sizes are 25, 50, 100, 200, 500 and 1000 each of this sample sizes were generated over 1000 iteration to inspect the effect of the estimators with respect to this sample sizes. Data were generated using equation (3.1) above but in this case, we make use of five predictors and the regression parameters β_0, β_1, β_2, β_3, β_4 and β_5 are set to be 20, 40, 30, 50, 80 and 55 respectively. The variables are generated using multivariate normal distribution and make use of two correlation structures, these are:

$$\begin{pmatrix} 1 & 0.7 & 0.7 & 0.7 & 0.8 \\ 0.7 & 1 & 0.8 & 0.8 & 0.93 \\ 0.7 & 0.8 & 1 & 0.93 & 0.93 \\ 0.7 & 0.8 & 0.93 & 1 & 0.95 \\ 0.8 & 0.93 & 0.93 & 0.95 & 1 \end{pmatrix}$$

The error term was generated using a normal distribution with mean zero and variance ten (10) and the relationship below is used to generate the regressand variable

$$y = 20 + 40X_1 + 30X_2 + 50X_3 + 80X_4 + 55X_5 + \epsilon_i$$

The purpose of all these simulation studies is to compare OLS and RR estimators through their absolute bias and MSE and so also to compare OLS, RR and PCR with respect to their mean square error of prediction (MSEP) to examine the predictive ability of each estimator.

5. RESULTS

Considering positive high correlation structure that is, r1=r2=r3=0.7; r4=r5=r6=0.8; r7=r8=r9=0.93; r10=0.95 the VIF for the simulated data set are as follows:

Table 1: VARIANCE INFLATION FACTOR (VIF) OF THE VARIABLES

SAMPLE SIZE	X1	X2	X3	X4	X5
n=25	5.747766	41.27515	16.20155	56.92161	275.6749
n=50	6.210466	41.5237	10.504	41.00265	211.9157
n=100	9.433733	56.20102	13.50942	60.03766	328.5111
n=200	5.978252	30.85184	9.088591	38.95834	174.0173
n=500	5.762636	33.79607	11.28923	41.32586	197.7226
n=1000	6.219904	37.52424	11.57944	44.70781	209.2167

From the above table, it can be seen that the variance inflation factor of the variables are more than ten (10) when the correlation between the explanatory variables was very high with different sample sizes except for X1 variable. Then, it is clearly shown that multicollinearity problem exists. Using method of OLS and RR to analyze the simulated data, the following results is obtained:

DOI : 10.31695/IJERAT.2019.3444

Table 2: Summary of the estimate of coefficient and variance using OLS and RR Method at simulated sample size 25-1000

SAMPLE SIZE	OLS ESTIMATE			RIDGE ESTIMATE		
		Coefficient	Variance		Coefficient	Variance
n=25	β_0 =20	20.09694	5.30926	β_0 =20	20.04515	5.086045
	β_1 =40	40.05853	34.64088	β_1 =40	40.05175	25.50986
	β_2 =30	30.95643	197.7324	β_2 =30	31.02001	123.4974
	β_3 =50	49.88632	59.07554	β_3 =50	50.17029	49.80548
	β_4 =80	80.58534	217.3549	β_4 =80	80.00637	140.6385
	β_5 =55	53.57662	1104.867	β_5 =55	53.70626	663.1037
n=50	β_0 =20	19.98742	2.315203	β_0 =20	19.96973	2.286983
	β_1 =40	40.04765	16.37619	β_1 =40	40.02747	13.34202
	β_2 =30	30.01891	85.73758	β_2 =30	30.04423	63.2876
	β_3 =50	50.1007	26.559	β_3 =50	50.20807	24.00618
	β_4 =80	79.86278	98.69653	β_4 =80	79.58739	74.36975
	β_5 =55	54.99462	490.1038	β_5 =55	55.09963	350.9956
n=100	β_0 =20	20.03455	1.058805	β_0 = 20	20.02519	1.052819
	β_1 =40	40.02255	7.35934	β_1 =40	40.00853	6.588757
	β_2 =30	30.08981	41.44856	β_2 =30	30.08108	35.9185
	β_3 =50	50.16944	13.20461	β_3 =50	50.21071	12.59018
	β_4 =80	80.02442	46.61918	β_4 =80	79.88036	40.81921
	β_5 =55	54.68129	233.1581	β_5 =55	54.77774	198.7391
n=200	β_0 =20	20.00226	0.531346	β_0 =20	19.99833	0.530053
	β_1 =40	40.02303	3.377697	β_1 =40	40.00405	3.179135
	β_2 =30	30.18862	18.5691	β_2 =30	30.15027	17.03515
	β_3 =50	50.12513	6.283709	β_3 =50	50.13887	6.088489
	β_4 =80	80.03548	21.53696	β_4 =80	79.9265	19.84494
	β_5 =55	54.67519	108.1894	β_5 =55	54.80605	98.41056

DOI : 10.31695/IJERAT.2019.3444

n=500	β_0 =20	20.00056	0.204958	β_0 =20	19.999	0.204897
	β_1 =40	39.92333	1.37536	β_1 =40	39.91996	1.339615
	β_2 =30	29.8807	7.152052	β_2 =30	29.87783	6.885544
	β_3 =50	49.92647	2.434517	β_3 =50	49.93565	2.400531
	β_4 =80	79.89837	8.592135	β_4 =80	79.86788	8.305911
	β_5 =55	55.34286	42.07386	β_5 =55	55.36432	40.38683
n=1000	β_0 =20	20.00803	0.088906	β_0 =20	20.00735	0.088851
	β_1 =40	39.99234	0.685176	β_1 =40	39.99013	0.676023
	β_2 =30	30.00141	3.816272	β_2 =30	29.99805	3.746377
	β_3 =50	49.99058	1.201453	β_3 =50	49.9938	1.193643
	β_4 =80	79.9891	4.683265	β_4 =80	79.97376	4.603021
	β_5 =55	55.03201	22.83449	β_5 =55	55.04663	22.38224

Both methods (OLS and RR) produced close estimates to the true value across the sample sizes. Taking a look at their variances the ridge regression estimate has minimum variance compared to the ordinary least squaresxs estimate. Increasing the sample sizes improve the variance of the two estimates. At a point in time, OLS estimates and RR estimates converge.

Table 3: Absolute bias of the estimators at various sample sizes

SAMPLE SIZES	ORDINARY LEAST SQUARES	RIDGE
n=25	10.22923	8.060476
n=50	6.853374	5.921017
n=100	4.701329	4.378599
n=200	3.267761	3.135145
n=500	2.026656	1.992275
n=1000	1.478323	1.466578

Table 3 present the Absolute bias of the estimators which is used to measure the consistency of the estimators, it was discovered that the RR has smaller absolute bias to that of the OLS estimate. Hence, this is an indication that ridge regression estimator is better than the Ordinary least square estimator though both are consistent and as the sample sizes become larger, the two estimators meet. The plot or graph line below shows the clearer picture of the presentations

DOI : 10.31695/IJERAT.2019.3444

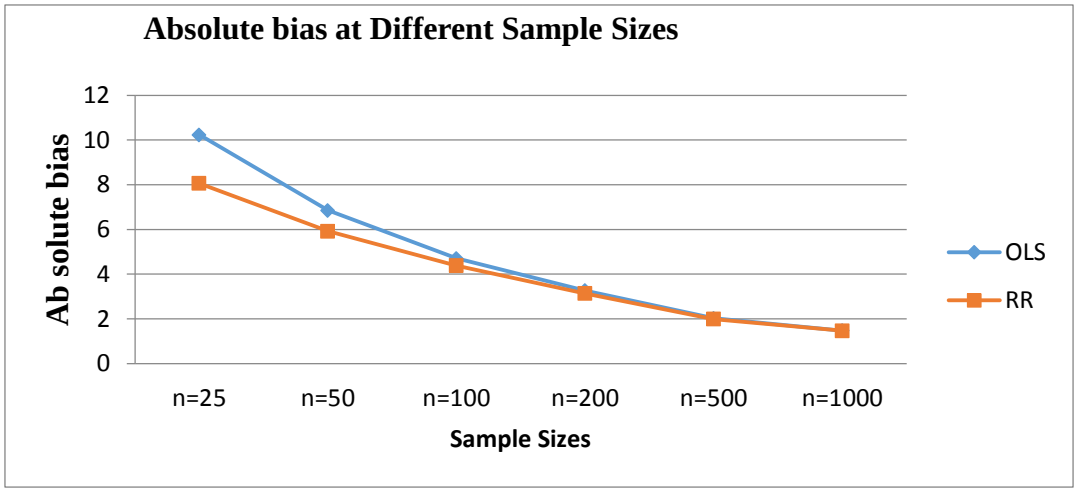

Fig 1.: Plot of Absolute bias at various sample sizes

Table 4: Mean square error of the estimators

SAMPLE SIZES	OLS MSE	RR MSE
n=25	270.1117	168.2302
n=50	119.8501	87.99783
n=100	57.10774	49.25471
n=200	26.41471	24.17136
n=500	10.32074	9.93991
n=1000	5.546268	5.443419

Table 4 presents the Mean Square Error MSE which is used to measure the efficiency of the estimators to the true values, it was observed that the RR estimator produce smaller MSE under the small, medium and large sample sizes compared to OLS estimator. While at large sample size the two estimators tends to produce similar mean square error (MSE) result.

Table 5: Mean square error of prediction the estimators

SAMPLE SIZE	**ORDINARY LEAST SQUARES MSEP**	**RIDGE MSEP**	**PC MSEP**
n=25	75.7604	168.2302	75.7604
n=50	88.4472	87.99783	88.4472
n=100	94.25558	49.25471	94.25558
n=200	96.79781	24.17136	96.79781
n=500	98.70971	9.93991	98.70971
n=1000	99.26722	5.443419	99.26722

DOI : 10.31695/IJERAT.2019.3444

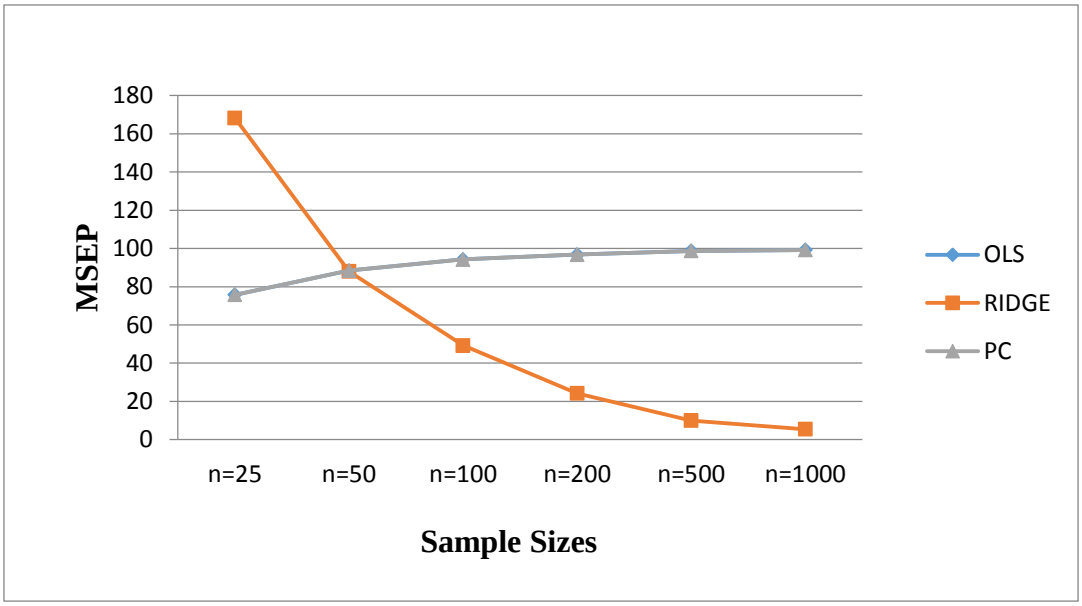

Fig 2.: Plot of MSE prediction at various sample sizes

Considering the predictive ability of the estimators, using their Mean Square Error of Prediction as a result were shown in table 4.5 above, It was observed that the OLS and PC estimators performed better than RR at sample size 25. For sample size 50 and above, the ridge regression estimator outperformed the other two estimators.

6. CONCLUSION

In the simulation of this study, it was observed that in both variance and MSE, the RR outperformed the OLS estimator and as the sample size increases the two estimators improved in their absolute bias and so also the MSE, thus this lessens the strength of multicollinearity. Accessing the predictive ability of the three estimators using mean square error of prediction (MSEP), it was examined that RR has lower MSEP compared to the other estimators and that OLS and PCR have the same value of MSEP.

REFERENCES

1. Akdeniz (2011): Generalized Maximum Entropy Estimators: Applications to the Portland Cement Dataset. *Open Statistics and Probalility Journal; 20011, vol.3 p13-23*

2. Alabi O.O, Ayinde, Kayode and Olatayo, T. O. (2008): Effect of Multicollinearity on power Rates of the Ordinary Least Square Estimators. *Journal of mathematics and Statistics; Vol.4, Issue 2, p75-81*

3. Batterham AM, Tolfrey K and George KP. (1997): Nevill's explanation of Kleiber's 0.75 mass exponent: an artifact of collinearity problems in least squares models? *Journal of Applied Physiology.;82 : 693–697.*

4. Cavell, A. C, Lydiate, D. J. and Parkin, I. A. P. (1998): Collinearity between a 30- centimorgan segment of Arabidopsis thaliana chromosome 4 and duplicated regions within the Brassica napus genome. Genome.;41 : 62–69.

5. Dorugade A. V. and Kashid D. N. (2010): Alternative Method for Choosing Ridge Parameter for Regression *Journal of Applied Mathematical Sciences, Vol. 4, no. 9, pg 447 – 456*

6. El-Dereny, M.and Rashwan, N. I.(2011): Solving Multicollinearity Problem using Ridge Regression Models, *Int. Journal. Contemp. Math. Sciences, Vol. 6, 2011, no. 12, 585 – 600.*

7. Feng-Jeng Lin(2008): Solving multicoolinearity in the process of fitting Regression Model using the Nested estimate procedure. Quality and quantity, June 2008, Vol.42 Issue 3, p417- 423.

8. Freund, R., and Littell, R. C. (2000): SAS System for Regression. 3rd ed. SAS Inst., Inc. Cary, NC.
9. Gorgees H. M. and Ali B. A. (2013): Employing Ridge Regression Procedure to Remedy the Multicollinearity Problem, *Jour. for Pure & Appl. Sci. Vol. 26 (1)*

DOI : 10.31695/IJERAT.2019.3444

10. Hoerl, A. E. and Kennard, R. W. (2000): Ridge Regression: Biased Estimation for non Orthogonal Problems, Techno metrics, 42, 80 – 86.

11. Hoerl, A. E. and Kennard, R.W. (1970): Ridge Regression: Biased Estimation for non orthogonal Problems, Technometrics, 12, 55 – 67

12. Kmita, M., Fraudeau, N. and Herault, Y. (2002)Serial deletions and duplications suggest a mechanism for the collinearity of Hoxd genes in limbs. Nature. 420:145–150.

13. Marquardt, D. W. (1970) Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation, Techno metrics, 12, No.3, 591-612

14. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2001): *Introduction to linear regression analysis*, 3rd edition, Wiley, New York.

15. Paramveer, S. D., Dean P. F., Sham M. K. and Lyle H. U. (2013): A Risk comparison of Ordinary Least Squares versus Ridge Regression. *Journal of Mechine learning Research vol 14, 1505- 1511*

16. Parkin, I. A. P., Lydiate, D. J. and Trick, M. (2002): Assessing the level of collinearity between Arabidopsis thaliana and Brassica napus for A. thaliana chromosome 5. Genome. 2002; 45:356– 366.

17. Ranjit K. P. (2013): Causes, Effects and Remedies of multicollinearity, department of Agricultural Statistics Roll No. 4405 I.A.S.R.I, Library Avenue, New Delhi-11001

18. Ying L. (2010); A Comparison Study of Principle Component Regression, Partial Least Squares Regression and Ridge Regression with Application to FTIR Data. Master thesis in statistics Faculty of Social Sciences Uppsala University, Sweden.

DOI : 10.31695/IJERAT.2019.3444