International Journal of Research Publication and Reviews
Journal homepage: www.ijrpr.com ISSN 2582-7421

Selection of Best Methods of estimation under Multicollinearity

Bature, Tajudeen Atanda[1*]; Lawal, Adekunle Yusuf[2]; Aji, David Adashu[3]; Habiba Danjuma[4]; Sulu, Abeeblahi Babajide[5]

[1, 5]Department of Statistics, Ahmadu Bello University, Zaria, Nigeria
[2] Nigerian Army College of Education, Ilorin, Nigeria
[3] Dept. of Mathematics &Statistics, Federal University Wukari, Nigeria
[4] Department of Statistics, Federal Polytechnic, Bali, Nigeria
Corresponding Email: tajudeenatanda56@yahoo.com

ABSTRACT

In this research, we tried to study the best estimator method under multicollinearity among ordinary least squared,Lasso Estimator, Robust-M Estimator, and Ridge Regression base on the assumption of multicollinearity.Also, the result shows the values of Variance of Coefficient estimates from the various estimators under low collinearity condition and Bias of Coefficient estimate from the various estimators under low collinearity condition.

KEYWORDS: Multicollinearity, Linear Regression Model, Dependent Variable, Independent Variable, BLUE and OLS Estimators.

1.0 INTRODUCTION

Multicollinearity is one of the problems of Ordinary Least Squares in regression analysis. Some estimators have been suggested as alternatives to the Ordinary Least Squares estimator to improve the accuracy of the parameter estimates in the linear regression model under multicollinearity. In this study, Robust-M, Ridge regression and Lasso estimators to handle the problem of multicollinearity are compared. The classical linear regression model and its assumptions, multicollinearity and its sources and consequences.

2.1 LITERATURE REVIEW

In this research the problem of multicollinearity in the linear regression model is considered. An overview on the multicollinearity diagnostics is also presented. Several methods that are used to detect the presence of multicollinearity are discussed. Besides, an overview on the methods to handle the problem of multicollinearity is presented. In particular, Robust-M, Ridge regression and Lasso estimators have been suggested as a means to improve the accuracy of the parameter estimate in the model when multicollinearity exists.

Ridge regression is a very common treatment for multicollinearity (Melonun*et al.*, 2002; Grewal *et al.*, 2004; Alauddin and Nghien, 2010; Alin, 2010).As it makes full use of the data and does not require the addition or removal of explanatory variables (Li *et al*, 2010). It was introduced by Hoerl and Kennard (1970). It is also a biased technique, with variance reduced in return for an introduction of some bias (Mason and Brown, 1975; Grewal *et al.*, 2004; Alin, 2010). Ridge regression works through the addition of small value K (which is a symmetric positive matrix) (Fourgeaud*etal*, 1984) to the correlation matrix of the variables in all the diagonal elements (thus creating the ridge which gives the regression its name).Where a normal regression estimation is based upon a standard X′X matrix, ridge regression uses an estimator defined as

$b = \left(X^T X + KI_p \right)^{-1} X^T Y$ (Hoerl and Kennard, 1970; Wan, 2002) where K=diag ($k_1,k_2,\ldots,k_p$) where k_i's are biasing parameters (Wan, 2002). It has been shown that defining $k_i = \sigma^2 / \gamma_i^2$ (where γ_i is the ithelement of γ) will minimize the mean square error (Wan, 2002).This addition of k allows ridge regression to have enough flexibility to reduce the inflated variances of OLS coefficients that arise from multicollinearity (Li *et al.*, 2010) and thus increase the reliability of point estimates (Butler and McNertney, 1991).

There have been various criteria put forward for enabling the most appropriate choice of k,such as using a variance inflation factor, generalized cross-validation and a ridge trace (Hoerl and Kennard, 1970; Li *et al.*, 2010).The introduction of a ridge parameter introduces bias into the model but also improves efficiency (Grewal *et al.*, 2004). It has been shown that the larger the value of k used, the smaller the variance will become (Hoerl and Kennard, 1970). However, this also causes more bias to arise thus creating a situation where a balance must be found to choose the appropriate level of k (Li *et al.*,2010).

The problem of robustness in statistics goes back to the beginning of statistics. Walker (1987)emphasizes the importance of applying robust estimators by showing the potential effects of multicollinearity on estimators.

Askin and Montgomerry (1980) introduced a family of estimators that combined robust M-estimation criteria with biased estimation constraints. Pfaffenberger and Dielman (1984) used similar approach to combat multicollinearity using M-estimation with LAV estimation. Lawrence and Marsh

(1984), Asking and Montgomery (1984), and Pfaffenberge and Dietman (1990) compared alternative combinations of ridge regression and robust regression techniques. Askin and Montgomerry, and Pfaffanberger and Dielman used designed experiments with Monte Carlo simulation, while Lawrence and Marsh used real data to predict fatalities in the US coal mining industry.

Rey (1983) notes that the Greek besiegers or antiquity switched from using the mean to a more robust measure, the median. Hampel*et al.*, (1986) pointed out that rejection of outliers was considered by Bernoulli (1777) and Bessel and Baeyer (1838). Formal rejection rules were given by Pierce (1852) and Chauvenet (1863).Through accounts of the early work can be found in papers by Harter (1974-1976), Huber (1972) and Stigler (1973), Box (1953) actually coined the term robustness and Tukey (1960) demonstrated the drastic no robustness of the mean and presented robust alternatives. In the 1960s papers by Huber (1964, 1965, and 1968) and Hampel (1968) formed the basis for the theory of robust estimation and extended this theory to applications such as regression.

3.0 RESEARCH METHODOLOGY

The research methodology consists Lasso Estimator, Robust-M Estimator, and Ridge Regression.

3.1 Robust-M Estimator

Procedure

M –estimation of location, instead of minimizing the sum of squares residuals, a robust regression M –estimator minimizes the sum of a less rapidly increasing function of residuals.

$$Min\sum_{i=1}^{n}\rho\left(y_i-\sum_{i=1}^{n}x_{ij}\beta_j\right)=Min\sum_{i=1}^{n}\rho\left(\varepsilon_i\right) \qquad .....(3.1)$$

Note that the residuals must be standardized by a robust estimate of their scale

$\hat{\sigma}_{\varepsilon}$, This is estimated simultaneously. Taking the derivatives of equation (2.19) and solving produces the score function

$$\sum_{i=1}^{n}\rho'\left(y_i-\sum\ x_{ij}\beta_j\left|\widehat{\sigma}\right.\right)x_{ik}=\sum_{i=1}^{n}\rho'\left(\varepsilon_i\left|\widehat{\sigma}_{\varepsilon}\right.\right)x_i=0. \qquad ....(3.20)$$

There is now a system of k+1 equations, for which p' is replaced by appropriate weights that decrease as the size of the residual increases

$$\sum_{i=1}^{n}w_i\left(\varepsilon_i\left|\widehat{\sigma}_{\varepsilon}\right.\right)x_i=0 \qquad .....(3.3)$$

An iterative reweighted least square (IRLS) is employed to find M –estimates for regression as follows

1. Setting the iteration counter at I=0, an OLS is fitted to the data, finding the initial estimates of the regression coefficients$\hat{\beta}^{(0)}$.

2. The residuals are extracted from the preliminary OLS regression,$\varepsilon_i^{(0)}$, and used to calculate initial estimates for the weights.

3. A weight function is then chosen and applied to the initial OLS residuals to create preliminary weights, W ($\varepsilon_i^{(0)}$).

4. The first iteration, I=1, uses weighted least squares (WLS) to minimize $\sum_{i=1}^{n}w_i^{(1)}\varepsilon_i^2$ and thus obtain $\hat{\beta}^{(1)}$.The solution is

$$\hat{\beta}^{(1)}\ =\ \left(X^TWX\right)^{-1}X^TWy \qquad .....(3.4)$$

5. The process continues by using the residuals from the initial WLS to calculate new weights$w_i^{(2)}$.

6. The new weights $w_i^{(2)}$ are used in the next iteration, I=2, of WLS to estimate $\hat{\beta}^{(2)}$.

7. Steps 4-6 are repeated until the estimate of $\hat{\beta}$ stabilizes from the previous iteration.

More generally, at each of the q iterations, the solution is

$\hat{\beta}^{(I)}=(X^TW_qX)^{-1}X^TW_qy,\ \ where\ W_q=diag\{w_i^{(I-1)}\}.$

The iteration continues until

$\hat{\beta}^{(I)}-\ \ \hat{\beta}^{(I-1)}\approx 0.$

3.2 LASSO ESTIMATOR

Procedure

Lasso estimator uses the same procedure with Ridge regression estimator above but the only difference is that the β_j^2 term in the ridge regression penalty has been replaced by $|\beta_j|$ in the lasso penalty

$$\hat{\beta}_L = \arg\min_{\beta} (Y - X\beta)^T (Y - X\beta) + K|\beta| I_P \quad \text{.....(3.5)}$$

$$= \arg\min_{\beta} (Y - X\beta)^T (Y - X\beta) + K\|\beta\|_1$$

Where also

$$X = \begin{pmatrix} x_{11} & . & . & . & x_{1p} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ x_{n1} & . & . & . & x_{np} \end{pmatrix}, \quad |\underline{\beta}| = \begin{pmatrix} |\beta_1| & 0 & . & . & 0 \\ 0 & |\beta_2| & & & . \\ . & & . & & . \\ . & & & . & . \\ . & & & . & 0 \\ 0 & . & . & . & |\beta_p| \end{pmatrix} = diag\left(|\beta_j|\right)$$

$$Y = \begin{pmatrix} y_1 \\ . \\ . \\ . \\ y_n \end{pmatrix} \text{ and } \quad I_p = \begin{pmatrix} 1 & 1 & 1 & . & . & . & 1 \\ 1 & 1 & 1 & . & . & . & . \\ 1 & . & & & & & . \\ . & . & & & & & .. \\ . & . & & & & & . \\ . & . & & & & & . \\ 1 & . & . & . & & & 1 \end{pmatrix}_{pxp}$$

Differentiating (2.17) with respect to β yields

$$\beta_{Lasso} = \left(X^T X + K B^{-1}\right)^{-1} X^T y \quad \text{....(3.6)} \cdot$$

Where B = Diag $\left(|\beta_1|, |\beta_2|, ..., |\beta_P|\right)$

3.3 Ridge Regression

Procedure

$$G(\beta, X, Y, K) = (Y - X\beta)^T (Y - X\beta) + \beta^2 I_p. \quad \text{.....(3.7)}$$

Minimizing G in equation (2.8) above we can get the best estimates of β which can be denoted $\hat{\beta}_R$

$$\hat{\beta}_R = \arg\min_{\beta} (Y - X\beta)^T (Y - X\beta) + \beta^2 I_p. \quad \text{....(3.8)}$$

$$= \arg\min_{\beta} (Y - X\beta)^T (Y - X\beta) + K\|\beta\|_2^2. \quad \text{.....(3.9)}$$

where $X = \begin{pmatrix} x_{11} & . & . & . & x_{1p} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ x_{n1} & . & . & . & x_{np} \end{pmatrix}$, y= $\begin{pmatrix} y_1 \\ . \\ . \\ . \\ y_n \end{pmatrix}$

$\underline{\beta^2}$ = $\begin{pmatrix} \beta_1^2 & 0 & . & . & . & 0 \\ 0 & \beta_2^2 & & & & . \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & \\ 0 & . & . & . & & \beta_p^2 \end{pmatrix} I_p$ = $\begin{pmatrix} 1 & 1 & . & . & . & 1 \\ 1 & 1 & & & & . \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & . \\ 1 & . & . & . & & 1 \end{pmatrix}_{pxp}$

Differentiating (2.9) with respect to β and equating the result to zero, we obtain

$$\left(X^T X + KI_p\right)\beta = X^T Y. \quad \text{.....}(3.10)$$ Yielding

$$\beta_R = \left(X^T X + KI_p\right)^{-1} X^T Y \quad \text{.....}(3.11)$$

$$= \left(X^T X + KI_p\right)^{-1} X^T X \beta^* \quad \text{. .....}(3.12)$$

Where the scalar k > 0 is called the ridge parameter which is choosing arbitrarily

$$E\left(\hat{\beta}_R\right) = \left(X^T X + KI_p\right)^{-1} X^T X E\left(\hat{\beta}\right)$$

$$= \left(X^T X + KI_p\right)^{-1} X^T X \beta^*. \quad \text{.....(3.13)}$$

$$Var\left(\hat{\beta}_R\right) = \left(X^T X + KI_p\right)^{-1} \text{var}\left(\hat{\beta}\right)\left[\left(X^T X + KI_p\right)^{-1} X^T X\right]^T$$

$= \boldsymbol{\sigma}^2(X^TX + KI_P)^{-1}(X^TX)(X^TX + KI_P)^{-1}$

Ridge estimator is biased

B $(\hat{\beta}_R)$ = E$(\hat{\beta}_R)$ – β*

= - k$(X^TX + KI_P)^{-1}$β* … (2.15)

≠ 0

However, each diagonal component of var $(\hat{\beta}_R)$ is always less than that of $\text{var}\left(\hat{\beta}\right)$.

MSE $(\hat{\beta}_R)$ = tr[$(\sigma^2 X^TX + KI_P)^{-1}X^TX(X^TX) + KI_P)^{-1}$] + $K^2\beta^{*T}((X^TX + KI_P)^{-2}\beta^*$ < MSE$(\hat{\beta})$

……(2.16)

However, the parameter k which minimizes MSE or any other nontrivial quadratic loss function depends on $\boldsymbol{\sigma}^2$ and unknown β.

3.4 METHODS OF ESTIMATION UNDER MULTICOLLINEARITY

i. Ordinary Least Squares (OLS)

ii. Ridge Regression Estimator

iii. Lasso Estimator

iv. Robust–M Estimator

i. Ordinary Least Squares (OLS)

The classic linear model is Y =**X**β + ε …. (2.1)

Where the design matrix **X**, is defined as annxp matrix of non-stochastic variables of rank p (<n) and Y as an nx1 vector of explained variable, β is defined as px1 parameters vector and ε is an nx1 vector of residuals with

E(ε)=0 ….(2.2)

Var(ε) =$\boldsymbol{\sigma}^2$ ….(2.3)

$$\mathbf{X}=\begin{pmatrix} x_{11} & . & . & . & x_{1p} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ x_{n1} & . & . & . & x_{np} \end{pmatrix}, \quad Y=\begin{pmatrix} y_1 \\ . \\ . \\ . \\ y_n \end{pmatrix} \varepsilon=\begin{pmatrix} \varepsilon_1 \\ . \\ . \\ . \\ \varepsilon_n \end{pmatrix}$$

The OLS aims to minimize

$$\sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T\varepsilon=\left(Y-X\beta\right)^T\left(Y-X\beta\right)$$

$$=Y^TT-Y^TX\beta-\beta^TX^TY+\beta^TX^TX\beta$$

at the minimum: $\frac{\delta}{\delta\beta}(\sum_{i=1}^{n}\varepsilon_i^2)$ = 0

$$= \frac{\delta}{\delta\beta}(Y^TY-Y^TX\beta-\beta^TX^T+\beta^TX^TX\beta)$$

$$=0 - X^TY-X^TY+2(X^TX)\beta$$

$$\Rightarrow X^TX\beta=X^T\beta \qquad \ldots\ldots\ldots\ldots(3.14)$$

OLS estimator $\hat{\beta}$ is obtain by solving the normal equation (5) above yielding

$$\hat{\beta}_{OLS}=\left(X^TX\right)^{-1}X^TY \qquad \ldots\ldots(3.15)$$

$$E\left(\hat{\beta}\right)=\beta^*. \qquad \ldots\ldots(3.16)$$

$$Var\left(\hat{\beta}\right)=E\left[\left(\hat{\beta}-X\beta^*\right)\left(\hat{\beta}-\beta^*\right)^T\right]=\sigma^2\left(X^TX\right)^{-1}. \qquad \ldots\ldots(3.17)$$

Where β^* is the true value of β.

OLS is a good estimation procedure if X′X is nearly a unit matrix. However, when multicollinearity occurs, OLS estimator will be sensitive.

4.0 DATA ANALYSIS

The results of the performance of the estimators at various levels of multicollinearity and at different sample sizes considered in this work are presented and discussed. The four estimators were assessed using each of the criteria stated in section 3.3. These estimators were ranked using the ranks 1, 2, 3, and 4 with rank 1 assigned to the best estimator that has lowest value of average variances, root-mean square errors and biases respectively. A rank 2 is assigned to the second-best estimator and so on.

4.1 PERFORMANCES OF ESTIMATORS UNDER LOW COLLINEARITY CONDITION

Table 4.1: Variance of Coefficient estimates from the various estimators under low collinearity condition

Sample Size n	Parameter	Estimator			
		OLS	Robust -M	Ridge	LASSO
$n = 20$	β_0	2860.688	3049.831	2633.21	2860.173
	β_1	3.038434	3.270756	2.392059	3.047398
	β_2	2.587179	2.670972	1.87163	2.57737
	β_3	1.440544	1.566129	1.068972	1.430239
	β_4	1.267413	1.346544	0.926946	1.246477
$n = 50$	β_0	1155.216	1217.949	1065.215	1142.891
	β_1	1.32167	1.404953	1.044026	1.321389
	β_2	0.950187	0.998828	0.689376	0.947196
	β_3	0.605084	0.640183	0.448442	0.604983
	β_4	0.517162	0.534456	0.37771	0.517284
$n = 200$	β_0	304.0445	321.6585	280.1351	301.0088
	β_1	0.308276	0.317417	0.242252	0.308324
	β_2	0.244536	0.257195	0.176236	0.243803
	β_3	0.147224	0.159128	0.109781	0.147215
	β_4	0.116124	0.124969	0.085125	0.116118

Table 4.1 shows the variances for all the four estimators considered when the collinearity level is low. Ridge Regression Estimator has slightly smaller variances than those of other estimators at all sample sizes. It can also be seen that, variances of all the parameters of four estimators decreases as sample size increases.

Table 4.2: Root Mean Square Error of Coefficient estimate from the various estimators under low collinearity condition

Sample Size n	Parameter	Estimator			
		OLS	Robust –M	Ridge	LASSO
$n = 20$	β_0	53.473	55.20033	153.8831	63.6876
	β_1	1.742789	1.809159	2.475933	1.806882
	β_2	1.60774	1.633494	7.744683	1.625762
	β_3	1.199629	1.25085	1.198569	1.210591
	β_4	1.125231	1.159906	2.183008	1.132821
$n = 50$	β_0	33.97728	34.89102	146.914	45.64032
	β_1	1.149723	1.185125	2.180562	1.229213
	β_2	0.974603	0.999342	7.632104	0.993802
	β_3	0.779683	0.802473	0.937044	0.786761
	β_4	0.719597	0.73196	2.027754	0.752005
$n = 200$	β_0	17.43682	17.93789	145.3538	36.26386
	β_1	0.555215	0.563333	1.942776	0.673001
	β_2	0.494359	0.507113	7.627081	0.547969
	β_3	0.38351	0.398762	0.691391	0.4226
	β_4	0.340854	0.353672	1.969686	0.394777

Table 4.2 shows the values of root mean square errors associated with the estimation of parameters of the model using the four methods of estimation and three sample sizes considered under low collinearity. From the table, it can be observed that Ordinary Least Squares (OLS) have slightly smaller values of RMSE of coefficients than the alternative estimators across all sample sizes while Ridge has slightly larger values of RMSE. It can also be seen that, root mean square errors of all the four estimators decrease as sample size increases.

5.0 CONCLUSION

Based on the general finding about the estimators presented and discussed earlier, the following conclusions are drawn:

1. That under severe collinearity condition, irrespective of the sample size, Ordinary Least Squares (OLS) is unbiased but inefficient.
2. That under severe collinearity condition, regardless of the sample size, Lasso estimator is the most efficient estimator but it is biased.
3. That sample size has little or no effect on the performance of the estimators across all the collinearity levels

5.1 RECOMMENDATION

Based on the above findings, the following are recommended:

1. That when collinearity level between the predictors is severe, Lasso Estimator will perform better regardless of the size of the data.
2. That the degree of multicollinearity between the predictors should be considered while estimating parameters of Regression models so as to avoid erroneous inferences.

REFRENCES

- Adichie, **J.** N. (1967), Estimation of Regression Coefficients Based on Rank Tests, *Annals of Mathematical Statistics*, 38,894-904.
- Akritas, M. G. (1991), Robust M Estimation in the Two-Sample Problem, *Journalof the American StatisticalAssociation, **86,** 201-204.*
- Alabi, O. O., Ayinde, Kayode, Olatayo, T. O. (2008). Effect of multicollinearity on power rates of the OLS Estimators. *Journal of mathematics and statistics; 2008, vol. 4 Issue 2, p75- 79.*
- Alauddin, M., & Nghiem, H. S. (2010). Do Instructional Attributes PoseMulticollinearity Problems? An Empirical Exploration. *EconomicAnalysis& Policy, 40, 351–361.*

- Pfaffenberger, R. C., and Dielman, T. E. (1990), A Comparison of Regression Estimators When Both Multicollinearity and Outliers Are Present, *Robust Regression: Analysis and Applications*,Lawrence, K. D. and Arthur, J. L. (Ed.), 243-270.
- Randolph, W. C. (1988). A transformation for heteroscedastic error components regression Models. *Economic Letters 27*, 349-354.
- Silvey, S. D. (1969). Multicollinearity and Imprecise Estimation. *Journal of theRoyal Statistical Society, 31, 539–552.*
- Stewart, G. W. (1987). Collinearity and Least Squares Regression. *StatisticalScience, 2, 68–84.*
- Tibshirani, R. (1996). *Regression shrinkage and selection via the LASSO*. J. Royal. Statist. Soc B., Vol. 58, No. 1, p 267-288)
- Vaughan, S. T. and Berry, E. K. (2005). Using Monte Carlo Techniques to Demonstrate the Meaning and Implications of Multicollinearity. *Journal of Statistics Education,13*(1).
- Yohai, V. J. (1987), High Breakdown-Point and High Efficiency Robust Estimates forRegression, *The Annals of Statistics*, 15, 642-656.
- Belsley, D. A. (1991). Conditioning Diagnostics Collinearity and Weak Data in Regression. New York: Wiley-Interscience.
- Belsley, D. A., Kuh, E., &Welsch, R. E. (1980). *Regression Diagnostics*. NewYork: John Wiley & Sons.
- Breusch, T. S. and Pagan, A. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica, 47*, 1287-1294.
- Brown, William G., Beattie, Bruce R. (1975): Improving Estimators of Economic Parameters by Use of Ridge Regression with Production function Applications. *Amerian Journal of Agricultural Economics; Feb75, Vol. 57 Issue 1, p21-26.*