Malaysian Journal of Computing, 6 (1): 698-714, 2021

eISSN: 2600-8238 online

INVESTIGATING THE IMPACT OF MULTICOLLINEARITY ON LINEAR REGRESSION ESTIMATES

Adewoye Kunle Bayo[1*], Ayinla Bayo Rafiu[2], Aminu Titilope Funmilayo[3], and Onikola Isaac Oluyemi[4]

[1,3,4]*Department of Statistics, The Federal Polytechnic, Offa, Kwara State, Nigeria*
[2]*Department of Statistics, University of Ilorin, Ilorin, Kwara State, Nigeria*
[1*]olakunleadewoye@yahoo.com, [2]atiku4sta@yahoo.com,
[3]telty4life@yahoo.com, [4]yemobenson2k@gmail.com,

ABSTRACT

The study was to investigate the impact of multicollinearity on linear regression estimates. The study was guided by the following specific objectives, (i) to examine the asymptotic properties of estimators and (ii) to compare lasso, ridge, elastic net with Ordinary Least Squares (OLS). The study employed Monte-Carlo simulation to generate set of highly collinear and induced multicollinearity variables with sample sizes of 25, 50, 100, 150, 200, 250, 1000 as a source of data in this research work and the data was analyzed with lasso, ridge, elastic net and ordinary least squares using statistical package. The study findings revealed that absolute bias of ordinary least squares was consistent at all sample sizes as revealed by past researched on multicollinearity as well while lasso type estimators fluctuated alternately. Also revealed that, mean square error of ridge regression outperformed other estimators with minimum variance at small sample size and OLS was the best at large sample size. The study recommended that OLS was asymptotically consistent at a specified sample sizes on this research work and ridge regression was efficient at small and moderate sample size.

Keywords: *Lasso & Elastic, Multicollinearity, Ridge*

Received for review: 28-08-2020; Accepted: 07-02-2021; Published: 23-02-2021

1. Introduction

Statistical analysis of econometric data (cross sectional) usually is based on the sets of fundamental assumptions such as homoscedasticity, normality, non-serial correlation, non–autocorrelation of error term. If any of these assumptions is violated, the effect of it on the parameter estimates usually leads to poor judgment on decision making. Henceforth, t-test or test of hypothesis, standard error and confidence interval that make use of parameter estimate become invalid.

A basic assumption in multiple linear regression model is that the rank of the matrix of observations on explanatory variables is the same as the number of explanatory variables. In other words, such matrix is of full column rank. This in turn implies that all the explanatory variables are independent, i.e., there is no linear relationship among the explanatory variables. It is termed that the explanatory variables are orthogonal. In many situations in practice, the explanatory variables may not remain independent due to various reasons. The situation where the explanatory variables are highly intercorrelated is referred to as multicollinearity.

Zakari *et al.* (2018) compares the Partial Least Squares Regression (PLSR), Ridge Regression (RR) and Principal Component Regression (PCR) as an alternative procedure for handling multicollinearity problem. A Monte Carlo simulation study was used to evaluate the effectiveness of these three procedures. Mean Squared Errors (MSE) was calculated for comparison purposes. From the results of their work, it shows that the RR is more efficient when the number of regressors is small, while the PLSR is more efficient than the others when the number of regressors is moderate or high. Esra & Suleyman (2015) compared partial least squares regression, principal component regression, ridge regression and multiple linear regression methods in modelling and predicting daily mean PM10 concentrations on the base of various meteorological parameters obtained for the city of Ankara, in Turkey. The analysed period is February 2007. Their results show that while multiple linear regression and ridge regression yield somewhat have better results for fitting to this dataset, principal component regression and partial least squares regression are better than both of them in terms of prediction of PM10 values for future datasets. In addition, partial least squares regression was the remarkable method in terms of predictive ability as it had a close performance with principal component regression even with a smaller number of factors.

Esra & Semra (2016) compared the performance of robust biased Robust Ridge Regression (RRR), Robust Principal Component Regression (RPCR) and RSIMPLS methods with each other and their classical versions known as Ridge Regression (RR), Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) in terms of predictive ability by using trimmed Root Mean Squared Error (TRMSE) statistic in case of both of multicollinearity and outlier's existence. Hence, the aim of this study is to investigate the impact of multicollinearity in linear regression estimates while the specific objectives are to examine the asymptotic properties of multicollinearity on how it affects the stability of parameter estimates and to compare lasso, ridge and elastic net estimators with ordinary least squares.

2. Methodology

The term multicollinearity is being defined as multi-implies many and collinearity implies linear dependence. In other words, it is the existence of near-linear relationships among the set of independent variables. The presence of multicollinearity causes all kinds of problems with regression analysis (Gujrati, 2004). Multicollinearity refers to a situation in which there is an exact (or nearly exact) linear relation among two or more of the input variables Hawking (1983). Exact relations usually arise by mistake or lack of understanding.

2.1 Types of Multicollinearity

2.1.1 Perfect Multicollinearity

In case of perfect multicollinearity (in which one independent variable is exact linear combination of the others) the design matrix X has less than full rank, and therefore the moment matrix $X'X$ cannot be inverted. Under these circumstances, for general linear model $Y_i = X\beta + \varepsilon$, the ordinary least-squares estimator $\hat{\beta}_{ols} = (X'X)^{-1}X'Y$ does not exist. Mathematically, a set of variables is perfectly multicollinear if there exist one or more exact linear relationships among some of the variables. It is expressed as

$$\lambda_0 + \lambda_1 x_{1i} + \lambda_2 x_{2i} + \dots + \lambda_k x_{ki} = 0 \tag{1}$$

Holding for all observations i, where λ_j are constants and X_{ij} is the ith observation on the jth explanatory variables.

Perfect multicollinearity is fairly common when working with raw datasets, which frequently contain redundant information. Once redundancies are identified and removed, however, nearly multicollinear variables often remain due to correlations inherent in the system being studied. In such a case, instead of the above equation holding, we have that equation in modified form with an error term.

$$\lambda_0 + \lambda_1 x_{1i} + \lambda_2 x_{2i} + \dots + \lambda_k x_{ki} + v_i = 0 \tag{2}$$

2.1.2 Imperfect Multicollinearity

In this case, there is no exact linear relationship among the variables, but the X_j variables are nearly perfectly multicollinear if the variance of v_i is small for some set of values for the λ's. In this case, the matrix $X'X$ has an inverse but is ill-conditioned so that a given computer algorithm may or may not be able to compute an approximate inverse, and if it does so the resulting computed inverse may be highly sensitive to slight variations in the data (due to magnified effects of either rounding error or slight variations in the sampled data points) and so may be very inaccurate or very sample-dependent.

2.2 Ordinary Least Square (OLS) Method

Recall that regression equation is given as:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pn} + \varepsilon_i \tag{3}$$

where Y_i is the explained variable, $X_1, X_2, X_3, \ldots, X_p$ are the explanatory variables, $\beta_1, \beta_2, \beta_3, \ldots, \beta_p$ are the regression coefficients and ε_i is the error term.

In matrix notation, we write the above model as multiple regression:

$$Y = X\beta + \varepsilon \tag{4}$$

where Y is nx1 vector of responses, X is an nxp matrix of the regressor variables, β is a px1 vector of unknown constants, and ε is an nx1 vector of random errors, with $\varepsilon_j \sim IID(0, \sigma^2)$ It will be convenient to assume that the regressor variables are standardized. Consequently, X^1X is a pxp matrix of correlations between the regressors and the response. Introducing normal function of normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(Y-X\beta)^1(Y-X\beta)} \tag{5}$$

$$f(\tilde{\beta}, \sigma^2/XY) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^{=2}}(Y-X\tilde{\beta})^1(Y-X\tilde{\beta})}$$

Using maximum likelihood function to obtain Least Squares Regression (OLS)

$$L(\tilde{\beta}, \sigma^2/XY) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(Y-X\tilde{\beta})^1(Y-X\tilde{\beta})} \tag{6}$$

Taking logarithm of both sides

$$LogL(\tilde{\beta}, \sigma^2/XY) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}log(\sigma^2) - \frac{1}{2\sigma^2}(Y - X\tilde{\beta})^1(Y - X\tilde{\beta}) \tag{7}$$

Differentiate with respect to $\tilde{\beta}$

$$\frac{\partial \log L(\beta, \delta^2/XY)}{\partial \beta} = \frac{-1}{2\delta^2}\left[-2X^1(Y - X)\tilde{\beta}\right]$$

Since

$$\frac{\partial LogL(\tilde{\beta}, \sigma^2/XY)}{\partial \tilde{\beta}} = 0 \tag{8}$$

$$X^1(Y - X\tilde{\beta}) = 0$$

$$\tilde{\beta} = (X^1X)^{-1}X^1Y \tag{9}$$

The coefficients of OLS $\hat{\beta}$ is proved to be unbiased estimate.
When differentiate with respect to σ^2

$$\frac{\partial LogL(\tilde{\beta},\sigma^2/XY)}{\partial\sigma^2} = -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\left[(Y-X\tilde{\beta})^1(Y-X\tilde{\beta})\right] \quad (10)$$

$$\frac{\partial LogL(\tilde{\beta},\sigma^2/XY)}{\partial\sigma^2} = 0$$

$$\frac{1}{2\sigma^4}\left[(Y-X\tilde{\beta})^1(Y-X\tilde{\beta})\right] = \frac{n}{2\sigma^2}$$

$$V(\tilde{\beta}_{OLS}) = \sigma^2(\mathrm{X}^1\mathrm{X})^{-1} \quad (11)$$

2.3 Ridge Regression Estimator

The Ridge Regression technique proposed by Hoerl & Kennard (1970) has become a common tool for analysis of data characterized with high multicollinearity. Addition of small positive quantities to the diagonal elements of the X^1X matrix prior to inverting it has also been suggested. For ridge regression can be expressed as:

$$\hat{\beta}_R = (X^1X + \lambda \mathrm{I})^{-1}X^1Y \quad (12)$$

Where $\hat{\beta}_R$ are ridge estimates of parameter vector, though the estimators are biased, they have more precision in terms of mean square error than the OLS estimators and λ is a positive value.

To estimate the coefficient of ridge regression as:

$$\hat{\beta} = (Y-X\tilde{\beta})^1(Y-X\tilde{\beta}) \text{ subject } \tilde{\beta}^1\tilde{\beta} \leq \mathrm{M} \quad \text{i.e } \tilde{\beta}^1\tilde{\beta}\text{-M}\leq 0 \quad (13)$$

$$f\,(\tilde{\beta},\sigma^2/XY) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}\{(Y-X\tilde{\beta})^1(Y-X\tilde{\beta})+\lambda(\tilde{\beta}^1\tilde{\beta}-\mathrm{M}\}} \quad (14)$$

Using maximum likelihood function to obtain Ridge Regression

$$L(\tilde{\beta},\sigma^2/XY) = (2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\{(Y-X\tilde{\beta})^1(Y-X\tilde{\beta})+\lambda(\tilde{\beta}^1\tilde{\beta}-\mathrm{M})\}} \quad (15)$$

$$= (2\pi)^{-\frac{n}{2}}(\sigma^2)^{\frac{1}{2}}e^{-\frac{1}{2\sigma^2}\{(Y-X\tilde{\beta})^1(Y-X\tilde{\beta})+\lambda(\tilde{\beta}^1\tilde{\beta}-\mathrm{M})\}}$$

Taking logarithm of both sides

$$logL\left(\tilde{\beta},\sigma^2/XY\right) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}log(\sigma^2) - \frac{1}{2\sigma^2}\{(Y-X\tilde{\beta})^1(Y-X\tilde{\beta}) + \lambda(\tilde{\beta}^1\tilde{\beta}-\mathrm{M})\Big) \quad (16)$$

Differentiate with respect to $\tilde{\beta}$

$$\frac{\partial \log L(\beta,\delta^2/XY)}{\partial\beta} = \frac{-1}{2\delta^2}\left[-2X^1\left(Y-X\tilde{\beta}\right)+2\lambda\tilde{\beta}\right] \quad (17)$$

Since $$\frac{\partial LogL(\tilde{\beta},\sigma^2/XY)}{\partial\tilde{\beta}} = 0$$

$$-\frac{1}{2\sigma^2}[-2X^1(Y-X\tilde{\beta}) + 2\,\lambda\tilde{\beta}]=0$$

$$-2X^1(Y-X\tilde{\beta}) + 2\,\lambda\tilde{\beta} = 0$$

$$2\,\lambda\tilde{\beta} = 2X^1(Y-X\tilde{\beta})$$

$$\tilde{\beta} = (X^1X + \lambda I)^{-1}X^1Y \tag{18}$$

Ridge regression estimator $E(\tilde{\beta}) = K\beta$ is biased
When differentiate with respect to σ^2

$$\frac{\partial \log L\left(\beta, \tilde{\delta^2}/XY\right)}{\partial \delta^2} = \frac{n}{2\delta^2} + \frac{1}{2\delta^4}\left[\left(Y - X\tilde{\beta}\right)^1\left(Y - X\tilde{\beta}\right) + \lambda\left(\tilde{\beta}^1\tilde{\beta} - M\right)\right] \tag{19}$$

$$\frac{\partial LogL(\tilde{\beta}, \sigma^2/XY)}{\partial \sigma^2} = 0$$

$$MSE(\tilde{\beta}_{Ridge}) = V(\tilde{\beta}_{Ridge}) + \left[\left(bias(\tilde{\beta})\right)^2\right] = \sigma^2 Tr[(X^1X + \lambda I)X^1X(X^1X + \lambda I)^{-1}] + \lambda^2\tilde{\beta}^1(X^1X + \lambda I)^{-2}\tilde{\beta}$$

$$RMSE = \sqrt{MSE(\tilde{\beta}_{Ridge})} \tag{20}$$

2.4 The Least Absolute Shrinkage and Selection Operator

Proposed by Chenlei *et al.* (2006) is a popular technique for model selection and estimation in linear regression models. It employs an L1-type penalty on the regression coefficients which tends to produce sparse models, and thus is often used as a variable selection tool. They showed that under appropriate conditions, the Lasso estimators are consistent for estimating the regression coefficients, and the limit distribution of the Lasso estimators can have positive probability mass at 0 when the true value of the parameter is 0. It has been demonstrated in that the Lasso is more stable and accurate than traditional variable selection methods such as best subset selection.

The Lasso estimate is the solution to

$$\hat{\boldsymbol{\beta}}_{(\mathbf{Lasso})} = \min_{\beta}(y - x\beta)'(y - x\beta), \quad s.t. \sum_{j=1}^{d}|\beta j| \le t. \tag{21}$$

Here $t \ge 0$ is a turning parameter. Let $\widehat{\beta^0}$ be the ordinary least square (OLS) estimate and $t_0 = \sum |\beta j^0|$. Values of $t < t_0$ will shrink the solutions toward 0. As shown in Chenlei *et al.* (2006) the Lasso gives sparse interpretable models and has excellent prediction accuracy. An alternative formulation of the Lasso is to solve the penalized likelihood problem

$$\hat{\boldsymbol{\beta}}_{(\mathbf{Lasso})} = \min_{\beta} \frac{1}{n}(y - x\beta)'(y - x\beta) + \lambda \sum_{j=1}^{d}|\beta j| \tag{22}$$

The formulation (21) and (22) are equivalent in the sense that, for any given λ Є [0, α), there exists a $t \ge 0$ such that the two problems have the same solution, and vice versa.

Introducing normal function for normal distribution

$$f(\tilde{\beta}, \sigma^2/XY) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(Y - X\tilde{\beta})^1(Y - X\tilde{\beta})} \tag{23}$$

To estimate coefficient of lasso regression are the solutions to the L1 optimization problem:

$$\text{minimize}(Y - X\tilde{\beta})^1(Y - X\tilde{\beta}) \text{subject to } \tilde{\beta} \le M \text{ i.e., } \tilde{\beta} - M \le 0 \tag{24}$$

$$f(\tilde{\beta}, \sigma^2/XY) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\{(Y - X\tilde{\beta})^1(Y - X\tilde{\beta}) + \lambda(\tilde{\beta} - M\}} \tag{25}$$

Using maximum likelihood function to obtain Lasso Regression

$$L(\tilde{\beta}, \sigma^2/XY) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\{(Y - X\tilde{\beta})^1(Y - X\tilde{\beta}) + \lambda(\tilde{\beta} - M)\}} \tag{26}$$

$$= (2\pi)^{-\frac{n}{2}}(\sigma^2)^{\frac{1}{2}}e^{-\frac{1}{2\sigma^2}\{(Y-X\tilde{\beta})^1(Y-X\tilde{\beta})+\lambda(\tilde{\beta}-M)\}}$$

Taking logarithm of both sides

$$\log L\left(\tilde{\beta},\delta^{2/}/XY\right)=\frac{-n}{2}\log(2\pi)-\frac{1}{2}\log(\delta^2)-\frac{1}{2\delta^2}\left\{\left(Y-X\tilde{\beta}\right)^1\left(Y-X\tilde{\beta}\right)+\lambda\left(\tilde{\beta}-M\right)\right. \quad (27)$$

Differentiate penalized residual sum of squares with respect to $\tilde{\beta}$

$$\frac{\partial LogL(\tilde{\beta},\sigma^2/XY)}{\partial\tilde{\beta}} = -\frac{1}{2\sigma^2}[-2X^1(Y-X\tilde{\beta})+\lambda]$$

Since $$\frac{\partial LogL(\tilde{\beta},\sigma^2/XY)}{\partial\tilde{\beta}} = 0 \quad (28)$$

$$-\frac{1}{2\sigma^2}[-2X^1(Y-X\tilde{\beta})+\lambda]=0$$

$$\tilde{\beta} = (X^1X)^{-1}\left(X^1Y-\frac{1}{2}\lambda\right) \quad (29)$$

When differentiate with respect to σ^2

$$\frac{\partial\log L\left(\beta,\delta^{\tilde{2}}/XY\right)}{\partial\delta^2}=\frac{n}{2\delta^2}+\frac{1}{2\delta^4}\left[\left(Y-X\tilde{\beta}\right)^1\left(Y-X\tilde{\beta}\right)+\lambda\left(\tilde{\beta}-M\right)\right] \quad (30)$$

$$\frac{\partial LogL(\tilde{\beta},\sigma^2/XY)}{\partial\sigma^2} = 0$$

$$\frac{1}{2\sigma^4}\left[(Y-X\tilde{\beta})^1(Y-X\tilde{\beta})+\lambda(\tilde{\beta}-M)\right]=\frac{n}{2\sigma^2}$$

$$MSE(\tilde{\beta}_{Lasso}) = V(\tilde{\beta}_{Lasso}) + \left(bias(\tilde{\beta})\right)^2$$

$$RMSE = \sqrt{MSE(\tilde{\beta}_{Lasso})} \quad (31)$$

It performs automatic variable selection, but it produces biased estimates for the large coefficients but ignore other group of variables.

2.5 Elastic Net

The elastic net proposed a compromise between the two that attempts to shrink and do a sparse selection simultaneously. A new regularization of the lasso for the unknown group of variables and for the multicollinear predictors. The elastic net method over comes the limitations of the Lasso method which uses a penalty function based on

$$\|\beta\|_1 = \sum_{j=1}^{p}|\beta j| \quad (32)$$

Use of this penalty function has several limitations. For instance, in the "large p, small n" case the Lasso selects at most n variables before it saturates. Also, if there is a group of highly correlated variables, then the Lasso tends to select one variable from a group and ignore the others. To overcome these limitations, the elastic net adds a quadratic part to the penalty ($\|\beta\|^2$), which when used alone is Ridge Regression (known also as Tikhonov regularization). The elastic net estimator can be expressed as:

$$\hat{\beta}_{EN} = arg\min_{\beta}(y - x\beta)'(y - x\beta) + \lambda_1 \sum_{j=1}^{p}|\beta j| + \lambda_2 \sum_{j=1}^{p}|\beta j|^2 \quad (33)$$

where λ_1 and λ_2 are turning parameters. As a result, the elastic net method includes the Lasso and Ridge Regression: in other words, each of them is a special case where $\lambda_1=\lambda$, $\lambda_2=0$ or vice versa. Introducing normal function for normal distribution

$$f\ (\tilde{\beta}, \sigma^2/XY) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(Y - X\tilde{\beta})^1(Y - X\tilde{\beta})} \quad (34)$$

To estimate coefficient of elastic net regression are the solutions to penalized L1 and L2 optimization problem:

minimize $(y - x\beta)'(y - x\beta)$ subject to $\lambda_1 \sum_{j=1}^{p}|\beta j| + \lambda_2 \sum_{j=1}^{p}|\beta j|^2$

$$f\ (\tilde{\beta}, \sigma^2/XY) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\{(Y - X\tilde{\beta})^1(Y - X\tilde{\beta}) + \lambda_1(\tilde{\beta} - M) + \lambda_2(\tilde{\beta}^1\tilde{\beta} - M)\}} \quad (35)$$

Using maximum likelihood function to obtain Lasso Regression

$$L(\tilde{\beta}, \sigma^2/XY) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\{(Y - X\tilde{\beta})^1(Y - X\tilde{\beta}) + \lambda_1(\tilde{\beta} - M) + \lambda_2(\tilde{\beta}^1\tilde{\beta} - M)\}} \quad (36)$$

$$logL(\tilde{\beta}, \sigma^2/XY) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}log(\sigma^2) - \frac{1}{2\sigma^2}\{(Y - X\tilde{\beta})^1(Y - X\tilde{\beta}) + \lambda_1(\tilde{\beta} - M) + \lambda_2(\tilde{\beta}^1\tilde{\beta} - M)\}$$

Differentiate penalized residual sum of squares with respect to $\tilde{\beta}$

$$\frac{\partial LogL(\tilde{\beta}, \sigma^2/XY)}{\partial\tilde{\beta}} = -\frac{1}{2\sigma^2}[-2X^1(Y - X\tilde{\beta}) + \lambda_1 + 2\lambda_2\tilde{\beta} \quad (37)$$

Since

$$\frac{\partial LogL(\tilde{\beta}, \sigma^2/XY)}{\partial\tilde{\beta}} = 0 \quad (38)$$

$$-\frac{1}{2\sigma^2}[-2X^1(Y - X\tilde{\beta}) + \lambda_1 + 2\lambda_2\tilde{\beta}] = 0$$

$$\tilde{\beta}_{EN} = (X^1X + \lambda_2 I)^{-1}(X^1Y - \frac{1}{2}\lambda)$$

When differentiate with respect to σ^2

$$\frac{\partial LogL(\tilde{\beta}, \sigma^2/XY)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\left[(Y - X\tilde{\beta})^1(Y - X\tilde{\beta}) + \lambda_1(\tilde{\beta} - M) + \lambda_2(\tilde{\beta}^1\tilde{\beta} - M)\right] \quad (39)$$

$$\frac{\partial LogL(\tilde{\beta}, \sigma^2/XY)}{\partial\sigma^2} = 0 \quad (40)$$

$$\frac{1}{2\sigma^4}\left[(Y - X\tilde{\beta})^1(Y - X\tilde{\beta}) + \lambda_1(\tilde{\beta} - M) + \lambda_2(\tilde{\beta}^1\tilde{\beta} - M)\right] = \frac{n}{2\sigma^2}$$

$$MSE(\tilde{\beta}_{EN}) = V(\tilde{\beta}_{EN}) + \left(bias(\tilde{\beta})\right)^2 \quad (41)$$

$$RMSE = \sqrt{MSE(\tilde{\beta}_{EN})} \quad (42)$$

2.6 Measurement Criteria

2.6.1 Bias of an Estimator

Let $\hat{\beta}$ be an estimator of a parameter β we say that $\hat{\beta}$ is a bias estimator of β if it becomes more likely to be close to the true value as the sample size increases. This intuition attribute is formalized in the notion of consistency.

It is defined that an estimator $\hat{\beta}$ of β is said to be consistent, if for any arbitrary ε, δ > 0 there exists a value N such that for n > N, the probability of the estimator $\hat{\beta}$. That is

$$Pr\{(|\hat{\beta} - \hat{\beta}|) < \varepsilon\} > 1 - \delta \quad (43)$$

Since the condition implies that

$$\lim_{n\to\infty} Pr(|\hat{\beta} - \hat{\beta}| < \varepsilon) = 1 \quad (44)$$

Consistency is alternatively referred to as convergences in probability. Accordingly, we say that $\hat{\beta}$ has β as its probability limit which we write $Pr \lim_{n\to\infty} \hat{\beta} = \beta$

$$\text{Biased} = |E(\hat{\beta}) - \beta| \quad (45)$$

A biased is said to be consistent if absolute value of biased estimates increases or decreases as the sample sizes increases.

2.6.2 Mean Square Error Estimator

If we have a scalar parameter β to be estimated and the statistic T is used as an estimator, then the mean square error is given as:

$$MSE\ (T,\beta) = E((T - \beta)^2, \beta) \quad (46)$$

$$= var(T,\beta) + (b\ \beta)^2$$

$$MSE = E(T - \beta)^2$$

$$= E(T - E(T) + E(T) - \beta)^2$$

$$\boldsymbol{MSE} = var(T) + [b(\beta)]^2 \quad (47)$$

Meaning that mean square error is the sum of variance estimator and squared of bias estimator and is efficient if smaller value is obtained compare to other estimators. In the case of unbiased estimators, it is just the ratio of their variance and the one with smaller variance will be more efficient if among all the unbiased estimators of $\hat{\beta}$ is the one with the smallest variance then it been called the most efficient estimator of β, that is for two unbiased estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ for the parameter β with variance $V(\hat{\beta}_1)$ and $V(\hat{\beta}_2)$ respectively, the efficiency of $\hat{\beta}_1$ relative $\hat{\beta}_2$ is defined by

$$e(\hat{\beta}_1, \hat{\beta}_2) = \frac{V(\hat{\beta}_2)}{V(\hat{\beta}_1)} \quad (48)$$

It is also explained about the fit of the model to data set used and defined as:

$$MSE = \frac{\sum_i^n (y_i - \hat{y}_i)^2}{n - p} = \frac{\sum_i^n e_i^2}{n - p} \quad (49)$$

2.6.3 Root of Mean Square Error Estimator

The root mean square error of an estimator $\hat{\beta}$ is said to be more efficient if compared with root mean square error of another estimator and defined as:

$$RMSE = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n-p}} \tag{50}$$

2.6.4 Predictive Mean Square Error Estimator

The predictive mean square error is defined as the average error in prediction y given x for future case not used in the construction of a prediction equation. There are two regression situations, x-random and x-controlled. In the case that x is random, both y and x are randomly selected. In the controlled situation, design matrices are selected by experimenters and only y is random. For case of presentation, simply consider only the x-random case. In x-random situations, the data (x_i, y_i) are assumed a random sample from its parent distribution (X, Y). Then, if $\hat{\mu}(X)$ is a prediction procedure constructed using the present data, the prediction error can be decomposed as

$$PE(\hat{\mu}) = E(Y - \hat{\mu}(X))^2 \tag{51}$$

where the expectation

$$PMSE = mean((Y - X\beta)^2) \tag{52}$$

2.7 Data Generation Processes

A simulation study is performed to examine the magnitude of bias, mean square error, root of mean square error and predictive mean square error due to the presence of multicollinearity amongst explanatory variables in multiple linear regression model given as

$$Y = X\beta + \varepsilon \tag{53}$$

Let the sample size n equals to 25, 50, 100, 125, 150, 200, 250 and 1000 using Monte-Carlo simulation to replicate each sample size n in 1000 times, then data sets were generated for highly collinear variables X_1, X_2 and X_3 over uniform distribution with minimum 0 and maximum 1 and induce multicollinearity in the variable X_4 and X_5 with assume error $e_1 = 0.2$ and $e_2 = 0.1$.

Case one: to consider the case where x_1, x_2 and x_3 are correlated with one another

$$X_i = runif(n, 0{,}1), \quad i = 1{,}2{,}3 \tag{54}$$

Case two: to consider the case where x_2 and x_3 are sum together with error e_1=0.2 to generate random values of X_4.

$$X_4 = X_2 + X_3 + e_1 \tag{55}$$

Case three: to consider the case where X_4 and error e_2=0.1 are sum together to generate random values of X_5.

$$X_5 = X_4 + e_2 \tag{56}$$

And the response variable Y is generated from covariate X with the true parameter and error term given as

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_5 + \beta_5 X_5 + rnorm(n) \tag{57}$$

Where β_i are 1.2, 2.3, 1.7, 2.5, 3.2 and 1.6, $i = 1{,}2{,}3, \dots, n$

3. Data Analysis, Results and Interpretation

This section depicted the analysis, results and interpretation of comparison of lasso type estimators with ordinary least squares (OLS) estimator when the covariates were multicollinear. The correlation between variables with multicollinearity X_4 and X_5 regressed with variables X_1, X_2 and X_3.

Table 1: Correlation matrix of multicollinear covariates

Rij	**X1**	**X2**	**X3**	**X4**	**X5**
X1	1.0000	-0.0593	0.07106	-0.0106	-0.0158
X2	-0.0593	1.0000	0.0491	0.6463	0.6285
X3	0.07106	0.0491	1.0000	0.6668	0.6513
X4	-0.0106	0.6463	0.6668	1.0000	0.9778
X5	-0.0158	0.6285	0.6513	0.9778	1.0000

From Table 1 above, showed that multicollinearity exist in a correlation matrix, where R_{ij} that had value above 0.9 shows that multicollinearity is present in the variable X_4 anX_5.

Table 2: Variance Inflation Factor (VIF) for the detection of multicollinearity

Variables	**25**	**50**	**100**	**125**	**150**	**200**	**250**	**1000**
X1	1.1920	1.0322	1.0143	1.0072	1.0024	1.0024	1.0024	1.0119
X2	2.5147	1.0322	3.7077	2.6136	3.1517	3.1517	3.1517	3.1313
X3	3.4871	4.0314	3.7515	2.6493	3.3550	3.3550	3.3550	3.3019
X4	25.5446	21.0742	26.9947	27.8870	26.0804	26.0804	26.0804	28.0683
X5	21.3060	18.6614	21.5356	23.2851	21.5203	21.5203	21.5203	22.8219

In Table 2 above depicted the detection of multicollinearity in the covariates using VIF, this implies that X_4 and X_5 had severe multicollinearity due to the high VIF that exceeded 10 which was the standard threshold.

Objective 1: To examine the asymptotic properties of estimators

Table 3: Absolute Bias and Mean Square Error of OLS Regression Estimate

Sample sizes	**Absolute Bias of OLS Regression Estimate Coefficients**						**Mean Square Error of OLS Regression Estimate Coefficients**					
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
25	0.0152	0.0327	0.0138	0.0214	0.0161	0.0272	0.7911	0.5467	1.3516	1.7367	8.1384	7.0188
50	0.0154	0.0434	0.0190	0.0233	0.0399	0.0242	0.1939	0.2478	0.9481	0.9463	2.0942	1.7269
100	0.0328	0.0309	0.0204	0.0292	0.0426	0.0360	0.1364	0.1309	0.5409	0.4276	1.3841	1.0393
125	0.0058	0.0183	0.0405	0.0028	0.0594	0.0413	0.0784	0.0978	0.3156	0.2602	1.0295	0.8009
150	0.0042	0.0048	0.0144	0.0103	0.0347	0.0243	0.0682	0.0837	0.219	0.2339	0.9148	0.7329
200	0.0004	0.0012	0.0049	0.0074	0.0197	0.0155	0.0578	0.0704	0.1893	0.1679	0.6200	0.5264
250	0.0009	0.0008	0.0128	0.0086	0.0049	0.0131	0.0438	0.0582	0.1366	0.1368	0.4253	0.3551
1000	0.0006	0.0029	0.0060	0.0028	0.0012	0.0013	0.0096	0.0124	0.0376	0.0365	0.1270	0.1006

Sample sizes	**Absolute Bias of OLS Regression Estimate Coefficients**						**Mean Square Error of OLS Regression Estimate Coefficients**					
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Minimum	0.0004	0.0008	0.0060	0.0028	0.0012	0.0013	0.0096	0.0124	0.0376	0.0365	0.1270	0.1006

From Table 3 above, the study observed that there are increments in absolute bias from sample size 25 to 100 when considering β_0 and β_3 coefficients, also as the sample sizes increases the absolute bias for β_0 and β_3 coefficients began to decrease, that is consistency of estimator at larger sample sizes, sample above 100 could be regarded as benchmark for user of statistics when facing with the problem of multicollinearity. When considering β_1 coefficient there is increment in absolute bias from sample size 25 to 50 as the sample sizes increases the absolute bias for β_1 coefficient began to decrease, that is consistency of estimator at larger sample sizes, sample above 50 could be regarded as benchmark for user of statistics when facing with the problem of multicollinearity. When considering β_2 and β_4 coefficients there are increments in absolute bias from sample size 25 to 125 as the sample sizes increases the absolute bias for β_2 and β_4 coefficients began to decrease, that is consistency of estimator at larger sample sizes, sample above 125 could be regarded as benchmark for user of statistics when facing with the problem of multicollinearity. When considering β_5 coefficient there is decrement in absolute bias from sample size 25 to 50 as the sample sizes increases the absolute bias for β_5 coefficient began to increase and decreases, that is inconsistency of estimator at larger sample sizes due to severe multicollinearity.

The study also observed from Table 3, the coefficients β_4 and β_5 that had severe multicollinearity among the MSE of coefficients β_0, β_1, β_2, β_3, β_4 and β_5 also decreases as sample sizes increases, this attest to the fact that the study observed efficiency of the estimator and conformity to the law of large number.

Table 4: Absolute Bias and Mean Square Error of Lasso Regression Estimate

Sample sizes	**Absolute Bias of Lasso Regression Estimate Coefficients**						**Mean Square Error of Lasso Regression Estimate Coefficients**					
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
25	0.0431	0.0019	0.0939	0.0489	0.0759	0.0449	0.7873	0.5678	1.2815	1.7455	7.4391	5.9181
50	0.0715	0.0575	0.0794	0.0660	0.0916	0.1259	0.1852	0.2525	0.9594	0.9285	1.944	1.6554
100	0.0507	0.0252	0.0871	0.0700	0.0958	0.0495	0.1328	0.1284	0.5689	0.4532	1.3866	1.0100
125	0.0609	0.0565	0.0701	0.0377	0.0495	0.0339	0.0858	0.1087	0.3110	0.2459	1.0534	0.8338
150	0.0655	0.0425	0.0754	0.0937	0.0983	0.0570	0.0718	0.0841	0.2448	0.2661	0.9547	0.7166
200	0.0527	0.0273	0.0833	0.0740	0.0487	0.0098	0.0615	0.0689	0.1935	0.1967	0.6263	0.5215
250	0.0582	0.0421	0.0967	0.0889	0.0648	0.0118	0.0402	0.0534	0.1498	0.1594	0.4847	0.3760
1000	0.0494	0.0326	0.0734	0.0635	0.0690	0.0354	0.0126	0.0134	0.0452	0.0435	0.1242	0.0971
Minimum	0.0431	0.0019	0.0701	0.0377	0.0487	0.0098	0.0126	0.0134	0.0452	0.0435	0.1242	0.0971

From Table 4 above, the study observed that there are increment and decrement in absolute bias varying from sample size to sample size when considering β_0, β_1, β_2, β_3, β_4 and β_5 coefficients as the sample sizes increases the absolute bias of coefficients began to decrease and increases that is inconsistency of estimator at varying sample sizes.

The study also observed from coefficients β_4 and β_5 that had severe multicollinearity among the MSE of coefficients β_0, β_1, β_2, β_3, β_4 and β_5 also decreases as sample sizes increases, this attest to the fact that the study observed efficiency of the estimator and conformity to the law of large number.

Table 5: Absolute Bias and Mean Square Error of Ridge Regression Estimate

Sample sizes	**Absolute Bias of Ridge Regression Estimate Coefficients**						**Mean Square Error of Ridge Regression Estimate Coefficients**					
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
25	0.4603	0.3114	0.1669	0.0648	0.6910	0.5160	0.8519	0.5275	0.7616	0.7614	0.9059	0.7268
50	0.2667	0.1310	0.2215	0.0941	0.7874	0.4361	0.2293	0.2056	0.4296	0.3398	0.7770	0.3453
100	0.3243	0.2388	0.1384	0.0603	0.7592	0.4483	0.2091	0.1613	0.2109	0.1585	0.6410	0.2687
125	0.2937	0.1967	0.2124	0.0855	0.7983	0.4520	0.1552	0.1205	0.2012	0.1277	0.6842	0.2500
150	0.3209	0.2063	0.1670	0.1056	0.8024	0.4538	0.1636	0.1119	0.1426	0.1270	0.6870	0.2456
200	0.3521	0.2244	0.0955	0.0300	0.7547	0.4583	0.1752	0.1076	0.0895	0.0853	0.6016	0.2406
250	0.2942	0.1969	0.1853	0.1255	0.8067	0.4541	0.1186	0.0823	0.0937	0.0793	0.6754	0.2286
1000	0.2871	0.1975	0.1837	0.1384	0.8063	0.4555	0.0904	0.0490	0.0506	0.0373	0.6564	0.2136
Minimum	0.2667	0.1310	0.0955	0.0300	0.7547	0.4361	0.0904	0.0490	0.0506	0.0373	0.6016	0.2136

From Table 5 above, the study observed that there are increment and decrement in absolute bias varying from sample size to sample size when considering β_0, β_1, β_2, β_3, β_4 and β_5 coefficients as the sample sizes increases the absolute bias of coefficients began to decrease and increases that is inconsistency of estimator at varying sample sizes.

The study observed from coefficients β_0, β_1, β_2, β_3, β_4 and β_5 that had severe multicollinearity the mean square error of coefficients β_0, β_1, β_2, β_3, β_4 and β_5 also decreases as sample sizes increases, these attest to the fact that the study observed efficiency of the estimator and conformity to the law of large number.

Table 6: Absolute Bias and Mean Square Error of Enet Regression Estimate

Sample sizes	**Absolute Bias of Elastic Net Coefficients**						**Mean Square Error of Elastic Net Coefficients**					
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
25	0.0090	0.0022	0.0110	0.0098	0.0045	0.0010	0.7283	0.5413	1.338	1.7424	7.6695	6.2743
50	0.0546	0.0272	0.0447	0.0486	0.1118	0.1248	0.1668	0.2394	0.9246	0.9287	1.7669	1.5108
100	0.0562	0.0405	0.0178	0.0261	0.0483	0.0341	0.1343	0.1296	0.4977	0.4187	1.0495	0.8113
125	0.0570	0.0459	0.0228	0.0432	0.1382	0.1339	0.0802	0.1009	0.3098	0.2425	0.7751	0.5927
150	0.0613	0.0372	0.0322	0.0552	0.0572	0.0599	0.0711	0.0832	0.2287	0.2489	0.7431	0.5701
200	0.0518	0.0258	0.0502	0.0439	0.0715	0.0800	0.0614	0.0684	0.1822	0.1864	0.5025	0.4301
250	0.0575	0.0368	0.0483	0.0456	0.1069	0.1129	0.0422	0.0557	0.1345	0.1443	0.3601	0.3008
1000	0.0585	0.0390	0.0430	0.0418	0.0876	0.0908	0.0132	0.0141	0.0370	0.0377	0.1021	0.0829
Minimum	0.0090	0.0022	0.0110	0.0098	0.0045	0.0010	0.0132	0.0141	0.0370	0.0377	0.1021	0.0829

From Table 6 above, the study observed that there are increment and decrement in absolute bias varying from sample size to sample size when considering β_0, β_1, β_2, β_3, β_4 and β_5 coefficients as the sample sizes increases the absolute bias of coefficients began to decrease and increases that is inconsistency of estimator at varying sample sizes.

The study observed from coefficients β_0, β_1, β_4 and β_5 that had severe multicollinearity, the mean square error of coefficients β_0, β_1, β_2, β_3, β_4 and β_5 also decreases as sample sizes increases, these attest to the fact that the study observed efficiency of the estimator and conformity to the law of large number.

Objective 2: **Compare lasso, ridge and elastic net estimators with ordinary least squares and interpretation.**

Table 7: Comparison of the Three Methods with OLS For Absolute Bias and MSE at n= 25

Sample size 25	**Estimators**	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Absolute Bias	**Ols**	0.0152	0.0327	0.0138	0.0214	0.0161	0.0272
	Lasso	0.0431	0.0019	0.0939	0.0489	0.0759	0.0449
	Ridge	0.4603	0.3114	0.1669	0.0648	0.6910	0.5160
	Enet	0.0090	0.0022	0.0110	0.0098	0.0045	0.0010
MSE	**Ols**	0.7911	0.5467	1.3516	1.7367	8.1384	7.0188
	Lasso	0.7873	0.5678	1.2815	1.7455	7.4391	5.9181
	Ridge	0.8519	0.5275	0.7616	0.7614	0.9059	0.7268
	Enet	0.7283	0.5413	1.338	1.7424	7.6695	6.2743

From Table 7 above, considering, the absolute bias of coefficients β_0, β_1, β_2, β_3, β_4 and β_5, the study observed that Elastic net is consistent and outperformed other estimators with minimum values and mean square error of Ridge is efficient. These served as benchmark regarded sample size of 25 when facing with problem of multicollinearity.

Table 8: Comparison of the Three Methods with OLS For Absolute Bias and MSE at n=50

Sample size 50	**Estimators**	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Absolute Bias	**Ols**	0.0154	0.0434	0.0190	0.0233	0.0399	0.0242
	Lasso	0.0715	0.0575	0.0794	0.0660	0.0916	0.1259
	Ridge	0.2667	0.1310	0.2215	0.0941	0.7874	0.4361
	Enet	0.0546	0.0272	0.0447	0.0486	0.1118	0.1248
MSE	**Ols**	0.1939	0.2478	0.9481	0.9463	2.0942	1.7269
	Lasso	0.1852	0.2525	0.9594	0.9285	1.9440	1.6554
	Ridge	0.2293	0.2056	0.4296	0.3398	0.7770	0.3453
	Enet	0.1668	0.2394	0.9246	0.9287	1.7669	1.5108

From Table 8 above, considering, the absolute bias of coefficients β_0, β_1, β_2, β_3, β_4 and β_5, the study observed that Ordinary least squares is consistent and outperformed other estimators with minimum values and mean square error of Ridge is efficient. These served as benchmark regarded sample size of 50 when facing with problem of multicollinearity.

Table 9: Comparison of the Three Methods with OLS For Absolute Bias and MSE at n=100

Sample size 100	**Estimators**	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Absolute Bias	**Ols**	0.0328	0.0309	0.0204	0.0292	0.0426	0.0360
	Lasso	0.0507	0.0252	0.0871	0.0700	0.0958	0.0495
	Ridge	0.3243	0.2388	0.1384	0.0603	0.7592	0.4483
	Enet	0.0562	0.0405	0.0178	0.0261	0.0483	0.0341
MSE	**Ols**	0.1364	0.1309	0.5409	0.4276	1.3841	1.0393
	Lasso	0.1328	0.1284	0.5689	0.4532	1.3866	1.0100
	Ridge	0.2091	0.1613	0.2109	0.1585	0.6410	0.2687
	Enet	0.1343	0.1296	0.4977	0.4187	1.0495	0.8113

From Table 9 above, considering, the absolute bias of coefficients β_0, β_1, β_2, β_3, β_4 and β_5, the study observed that Ordinary least squares is consistent and outperformed other estimators with minimum values and mean square error of Ridge is efficient. These served as benchmark regarded sample size of 100 when facing with problem of multicollinearity.

Table 10: Comparison of the Three Methods with OLS For Absolute Bias and MSE at n=125

Sample size 125	**Estimators**	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Absolute Bias	**Ols**	0.0058	0.0183	0.0405	0.0028	0.0594	0.0413
	Lasso	0.0609	0.0565	0.0701	0.0377	0.0495	0.0339
	Ridge	0.2937	0.1967	0.2124	0.0855	0.7983	0.4520
	Enet	0.0570	0.0459	0.0228	0.0432	0.1382	0.1339
MSE	**Ols**	0.0784	0.0978	0.3156	0.2602	1.0295	0.8009
	Lasso	0.0858	0.1087	0.3110	0.2459	1.0534	0.8338
	Ridge	0.1552	0.1205	0.2012	0.1277	0.6842	0.2500
	Enet	0.0802	0.1009	0.3098	0.2425	0.7751	0.5927

From Table 10 above, considering, the absolute bias of coefficients β_0, β_1, β_2, β_3, β_4 and β_5, the study observed that Ordinary least squares is consistent and outperformed other estimators with minimum values and mean square error of Ridge is efficient. These served as benchmark regarded sample size of 125 when facing with problem of multicollinearity.

Table 11: Comparison of the Three Methods with OLS For Absolute Bias and MSE at n=150

Sample size 150	**Estimators**	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Absolute Bias	**Ols**	0.0042	0.0048	0.0144	0.0103	0.0347	0.0243
	Lasso	0.0655	0.0425	0.0754	0.0937	0.0983	0.0570
	Ridge	0.3325	0.2239	0.1606	0.1408	0.8168	0.4477
	Enet	0.0727	0.0429	0.0309	0.0148	0.1020	0.0754
MSE	**Ols**	0.0682	0.0837	0.2190	0.2339	0.9148	0.7329
	Lasso	0.0718	0.0841	0.2448	0.2661	0.9547	0.7166
	Ridge	0.1743	0.1218	0.1361	0.1379	0.7072	0.2427
	Enet	0.0750	0.0807	0.2110	0.2505	0.6986	0.5609

From Table 11 above, considering, the absolute bias of coefficients β_0, β_1, β_2, β_3, β_4 and β_5, the study observed that Ordinary least squares is consistent and outperformed other estimators with minimum values and mean square error of Ridge is efficient. These served as benchmark regarded sample size of 150 when facing with problem of multicollinearity.

Table 12: Comparison of the Three Methods with OLS For Absolute Bias and MSE at n=200

Sample size 200	**Estimators**	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Absolute Bias	**Ols**	0.0004	0.0012	0.0049	0.0074	0.0197	0.0155
	Lasso	0.0527	0.0273	0.0833	0.0740	0.0487	0.0098
	Ridge	0.3625	0.2422	0.1088	0.0291	0.7656	0.4615
	Enet	0.0593	0.0382	0.0383	0.0484	0.1195	0.1207
MSE	**Ols**	0.0578	0.0704	0.1893	0.1679	0.6200	0.5264
	Lasso	0.0615	0.0689	0.1935	0.1967	0.6263	0.5215
	Ridge	0.1827	0.1135	0.0962	0.0778	0.6192	0.2451
	Enet	0.0649	0.0769	0.1936	0.1858	0.5341	0.4461

From Table 12 above, considering, the absolute bias of coefficients β_0, β_1, β_2, β_3, β_4 and β_5, the study observed that Ordinary least squares is consistent and outperformed other estimators with minimum values and mean square error of Ridge is efficient. These served as benchmark regarded sample size of 200 when facing with problem of multicollinearity.

Table 13: Comparison of the Three Methods with OLS for Absolute Bias and MSE at n=250

Sample size 250	Estimators	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Absolute Bias	**Ols**	0.0009	0.0008	0.0128	0.0086	0.0049	0.0131
	Lasso	0.0582	0.0421	0.0967	0.0889	0.0648	0.0118
	Ridge	0.2942	0.1969	0.1853	0.1255	0.8067	0.4541
	Enet	0.0575	0.0368	0.0483	0.0456	0.1069	0.1129
MSE	**Ols**	0.0438	0.0582	0.1366	0.1368	0.4253	0.3551
	Lasso	0.0402	0.0534	0.1498	0.1594	0.4847	0.3760
	Ridge	0.1186	0.0823	0.0937	0.0793	0.6754	0.2286
	Enet	0.0422	0.0557	0.1345	0.1443	0.3601	0.3008

From Table 13 above, considering, the absolute bias of coefficients β_0, β_1, β_2, β_3, β_4 and β_5, the study observed that Ordinary least squares is consistent and outperformed other estimators with minimum values and mean square error of Ridge is efficient. These served as benchmark regarded sample size of 250 when facing with problem of multicollinearity.

Table 14: Comparison of the Three Methods with OLS for Absolute Bias and MSE at n=1000

Sample size 1000	Estimators	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Absolute Bias	**Ols**	0.0006	0.0029	0.006	0.0028	0.0012	0.0013
	Lasso	0.0494	0.0326	0.0734	0.0635	0.0690	0.0354
	Ridge	0.2871	0.1975	0.1837	0.1384	0.8063	0.4555
	Enet	0.0585	0.0390	0.0430	0.0418	0.0876	0.0908
MSE	**Ols**	0.0096	0.0124	0.0376	0.0365	0.1270	0.1006
	Lasso	0.0126	0.0134	0.0452	0.0435	0.1242	0.0971
	Ridge	0.0904	0.0490	0.0506	0.0373	0.6564	0.2136
	Enet	0.0132	0.0141	0.0370	0.0377	0.1021	0.0829

From Table 14 above, considering, the absolute bias of coefficients β_0, β_1, β_2, β_3, β_4 and β_5, the study observed that OLS is consistent and outperformed other estimators with minimum values and mean square error of Ordinary least squares is efficient. These served as benchmark regarded sample size of 1000 when facing with problem of multicollinearity.

Table 15: Comparison of the Three Methods with OLS for PMSE

Estimator	25	50	100	125	150	200	250	1000
Ols	1.3065	0.9387	1.2422	0.9978	0.9962	1.0642	0.9975	1.0005
Lasso	1.2955	0.9312	1.2465	0.7129	1.0293	1.0105	0.9550	1.0007
Ridge	1.3110	0.9584	1.2594	0.8919	1.0096	1.1552	1.0369	1.0278
Enet	1.3039	0.9276	1.2449	0.9149	0.9319	1.0294	0.9122	1.0021

Table 15 above depicted the comparison of the predictive mean square error of three different methods with OLS using different sample sizes. It showed that at n=25, Lasso is efficient for precision, at n=50, Elastic net is efficient for precision and at n=100 and 1000, OLS is good for precision.

4. Findings

This study investigated the impact of investigating the multicollinearity on linear regression estimates of lasso, ridge and elastic net compared with ordinary least squares measure on criteria of absolute bias, mean square error and predictive mean square error. From the analysis, it can be clearly observed that the three methods applied can be used to solve the problem of multicollinearity. Objective one is to examine the asymptotic properties of estimators. From Table 3, the study observed that, there are increment in absolute bias of OLS from one sample size to another when considering parameter estimates as the sample sizes is increasing the absolute bias began to decrease showed that OLS is consistent. It can be observed that, there is decrement in mean square error of OLS as sample sizes increases when considering parameter estimates that is efficiency of an estimator. From Table 4, 5 and 6, the study observed that, there are increment and decrement in absolute bias of lasso, ridge and elastic net as the sample sizes increases when considering parameter estimates that are fluctuate alternately showed inconsistency of the estimators. From Table 4, 5 and 6, the study observed that, there are decreases in mean square error of lasso, ridge and elastic net as the sample sizes increases when considering parameter estimates, it showed that estimators were efficient. The objective two, to compare the three lasso type estimators with OLS. From Table 7, the study observed that, absolute bias of elastic net is consistent and mean square error of ridge is efficient when considering parameter estimates with minimum value at sample sizes of 25. From Table 8, 9, 10, 11, 12 and 13, the study observed that, absolute bias of OLSs is consistent and mean square error of ridge is efficient when considering parameter estimates with minimum value at sample sizes of 50, 100, 125, 150, 200 and 250 respectively.

5. Conclusion

This study presents the impact of multicollinearity on the methods of estimating the parameters of a regression model. From the analysis, it can be clearly observed that some methods applied can be used to solve the problem of multicollinearity when the sample size is small and moderate. Judging from the tables, it can be observed that: In examining the asymptotic properties of an estimator, absolute bias of OLS is consistent and mean square error for lasso, ridge, elastic net and OLS asymptotically decrease as the sample size increases thus, they are asymptotically efficient. To compare the three methods of lasso type estimators with OLS in solving the problem of multicollinearity showed that ridge outperform other estimators followed by elastic net. Focusing on multicollinearity, this study seeks to contribute to recent efforts to improve researchers' methodological approaches to the analysis of linear regression model in some particular areas such as: a) When focusing on small and moderate sample sizes in the presence of multicollinearity, ridge regression is recommended and OLS at large sample sizes i.e., assumption of ordinary least square is valid and b) To overview the major variables that may influence multicollinearity.

References

Chenlei, L., Yi, L. and Grace W. (2006). A note on the LASSO and related procedures in model selection. Published by Institute of Statistical Science, Academia Sinical. Pp.1273-1284.

Esra, P and Semra, T. (2016). The comparison of classical and Robust Biased Regression methods for determining unemployment rate in Turkey. *Journal of data science: JDS,* pp 739-768.

Esra, P and Suleyman, G. (2015). The comparison of partial least Squares Regression, Principal Component Regression and Ridge Regression with Multiple Linear Regression for predicting PM10 concentration level Based on Meteorological parameter. *Journal of Data science*, 663-692.

Gujrati, D. N. (2004). Basic econometrics 4th edition, Tata McGraw-Hill, New Delhi. *McGraw Companies*

Hawking, R. R. and Pendleton, O. J. (1983). The regression dilemma, Commun. Stat. *Theo. Meth*, 12, 497-527.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for Nonorthogonal problems. *Technometrics,* 12, 55-67.

Hoerl, A. E. and Kennard, R. W. (1976). Ridge regression: Iterative estimation of biasing parameter, Commun. Stat. *Theo. Meth,* 5, 77-88.

Hoerl, A. E., Kennard, R. W. and Baldwin, K. F. (1975). Ridge regression: some simulations, Commun. Stat. *Theo. Meth*., 4, 105-123.

Zakari Y., Yau S. A. and Usman, U. (2018). Handling multicollinearity; A comparative study of the prediction performance of some methods based on some probability distribution. *Annals. Computer Science Series pp 15-21*.