

**Research Article**

Copyright © All rights are reserved by Afeez Mayowa BABALOLA

Modelling Colinearity in the Presence of Non-Normal Error: A Robust Regression Approach

Afeez Mayowa BABALOLA*¹ and Maxwell Obubu²¹Department of Statistics, University of Ilorin, Nigeria²Department of Statistics, Nnamdi Azikwe University, Nigeria***Corresponding author:** Afeez Mayowa BABALOLA, Department of Statistics, University of Ilorin, Nigeria.**Received Date:** July 15, 2019**Published Date:** July 26, 2019**Abstract**

Multicollinearity and non-normal errors, which lead to unwanted effect on the least square estimator, are common problems in multiple regression models. It would therefore seem important to combine estimation techniques for addressing these problems. In the presence of multicollinearity and non-normal errors, different estimation techniques were examined, namely, the Ordinary Least Squares (OLS), Ridge Regression (R), Weighted Ridge regression (WR), Robust M-estimation (M), and Robust Ridge Regression with emphasis on M-estimation (RM). When compared with the condition of Collinearity, the results of a simulated study shows that Robust Ridge (RM) provides a more efficient estimate than the other estimators considered. When both Collinearity and non-normal errors were considered, the M-estimators (M) produces a more efficient and precise estimates.

Keywords: Multicollinearity; Ordinary least squares (OLS); Regressor; Simulation; M-estimator; Estimates**Introduction**

Regression analysis is “robust” in that it will typically provide estimates that are reasonably unbiased and efficient even when one or more of the assumptions is not completely met. However, a large violation of one or more assumptions will result in poor estimates and, consequently, the wrong conclusions being drawn. Two important problems are considered in regression analysis: multicollinearity and non-normal error distribution. Multicollinearity is the term used to describe cases in which the regressors are correlated among themselves. Another common problem in regression estimation methods is that of non-normal errors. The term simply means that the error distributions have fatter tails than the normal distribution. These fat-tailed distributions are more prone than the normal distribution to produce outliers, or extreme observations in the data. The Ordinary Least Squares estimators (LS) of coefficients are known to possess certain optimal properties when explanatory variables are not correlated among themselves, and the disturbances of the regression equation are independently, identically distributed normal random variables. This study focuses on violation of the assumption that there is no linear relationship between the explanatory variables and the disturbances distribution is non-normal. If this assumption

is violated, multicollinearity ensues which leads to inflation of Standard Error (SE) of the coefficients of the estimated parameters [1-6]. Multicollinearity is a case of multiple regression in which the predictor variables are themselves highly correlated, meaning that one can be linearly predicted from the others with a non-trivial degree of accuracy. If there is no linear relationship between the regressors, they are said to be orthogonal. When the regressors are orthogonal, the inferences such as;

- (i) identifying the relative effects of the regressor variables,
- (ii) prediction and/or estimation, and
- (iii) selection of an appropriate set of variables for the model

Can be made easily. Unfortunately, in most applications of regression, the regressors are not orthogonal. Sometimes the lack of orthogonality is not serious. However, in some situations the regressors are nearly perfectly linearly related, and in such case the inferences based on the regression model can be misleading or erroneous. When there are near-linear dependencies among the regressors, the problem of multicollinearity is said to exist. Non-normal error is a situation whereby the errors have fatter tails than the normal distribution tail. The Gauss-Markov Theorem

says that OLS estimates for coefficients are BLUE when the errors are normal and homoscedastic. When errors are nonnormal, the 'E' property (Efficient) no longer holds for the estimators and in small samples, and the standard errors will be biased. When errors are heteroscedastic, the standard errors become biased. Thus, we typically examine the distribution of the errors to determine whether they are normal.

M – Estimation

The theory of robustness developed by Huber and Hampel (1960) laid the foundation for finding practical solutions to many problems, when statistical concepts were vague to serve the purpose. Robust regression analyses have been developed as an improvement to least squares estimation in the presence of outliers and to provide us information about what a valid observation is and whether this should be thrown. The primary purpose of robust regression analysis is to fit a model which represents the information in the majority of the data. Robust regression is an important tool for analyzing data that are contaminated with outliers. It can be used to detect outliers and to provide resistant results in the presence of outliers. Many methods have been developed for these problems. Many researchers have worked in this field and described the methods of robust estimators. The class of robust estimators includes M-, L- and R-estimators. The M-estimators are most flexible ones, and they generalize straightforwardly to multiparameter problems, even though they are not automatically scale invariant and have to be supplemented for practical applications by an auxiliary estimate of scale any estimate [6-10].

Motivation

This research work was motivated by the idea that in the presence of substantial collinearity and non-normal error in a regression model, the results obtained from Ordinary Least Square Estimator are unreliable. The effect and remedy to such influences needs to be study.

Material and Methods

Ridge regression estimation

The basic requirement of the LS estimation of a linear regression is that $(X^T X)^{-1}$ exists. The two major reasons that the inverse does not exist is (1) when $n > p$, (2) multicollinearity. When multicollinearity exists the design matrix is "ill conditioned" and invertible. A simple way to guarantee the invertibility is adding a constant to the diagonal matrix $(X^T X)$ before estimating the coefficients. Hoerl and Kennard (1970) proposed the ridge estimator

$$\beta_R = (X^T X + k)(X^T X)^{-1} X^T Y; k > 0$$

Another way of understanding the ridge estimator is; recall that the LS estimator tends to minimize the residual sum square i.e. $\sum_{i=1}^n (Y_i - X_i \beta)^2$; or say it choose $\min_{\beta} \sum_{i=1}^n (Y_i - X_i \beta)^2$; but when collinearity exist then $\hat{\beta}$ is sensitive to small change in the data set and we uses ridge estimator that penalizes the value of β by minimizing

$$\min_{\beta} \left\{ \sum_{i=1}^n (Y_i - X_i \beta)^2 + k \|\beta\|^2 \right\};$$

The first term of equation (3.1.5) is regarded as the *loss* and the second term of the equation is the *penalty*. Note that: with larger k the penalty of β becomes stronger [10-14].

Robust regression methods

Robust Regression estimators have been proven to be more reliable and efficient than Least Squares Estimator especially when disturbances are non-normal. "Non-normal disturbances" are disturbances distribution that have heavy and fatter tails than the normal distribution and are prone to produce outliers. Since outliers greatly influence the estimated coefficients, standard errors and test statistics, the usual statistical procedure may be most inefficient as the precision of the estimator has been affected. Several different classifications of robust regression exist. Two of the most commonly considered group are: L-estimators and M-estimators. The L-estimators is the earliest one. One important member of regression L-estimators is called Weighted Least Squares estimator (WLS).

Weighted Least Squares Estimator

Given a regression model: $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + e_i$ where e_i are random errors that is independent and identically distributed with mean zero and variance σ^2 . Suppose we have

$$E(Y/X = x_i) = \beta' x_i$$

and we assume

$$V(Y/X = x_i) = \text{Var}(e_i) = \sigma^2 / w_i$$

where w_1, w_2, \dots, w_n are known positive numbers.

The model can now be re-written as

$$Y = X\beta + e$$

$$\text{Var}(e) = \sigma^2 w^{-1}$$

$$\text{RSS}(\beta) = (Y - X\beta)' W (Y - X\beta)$$

$$\sum w_i (y_i - x_i' \beta)^2$$

The Weighted Least Squares estimator also known as the Generalized Least Squares estimator is given by:

$$\hat{\beta}_{WLS} = (X'WX)^{-1} X'WY$$

where W is a diagonal matrix with diagonal elements. The diagonal elements of W matrix are set equal to:

$$w_{ii} = \begin{cases} \frac{1}{\hat{e}_i} & \text{if } \hat{e}_i \neq \text{zero} \\ 1 & \text{if } \hat{e}_i = \text{zero} \end{cases}$$

M - Estimation

The linear model:

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + e_i$$

This can also be written in matrix form as:

$$Y = X\beta + \epsilon$$

The general M-estimator minimizes the objective function

$$\sum_{i=1}^n \rho(e_i) = \sum_{i=1}^n \rho(y_i - x_i' \beta)$$

Where the function ρ gives the contribution of each residual to the objective function

Robust ridge regression estimates

The following formula is used to compute robust ridge estimates:

$$\hat{\beta}_{RobustRidge} = (X'X + KI)^{-1} X'X \hat{\beta}_{Robust} \dots\dots\dots (3.3.1)$$

where $\hat{\beta}_{Robust}$ denotes the coefficient estimates from the robust estimators. Many methods of selecting appropriate k values have been proposed in the literature. In this study, the method proposed by Hoerl, Kennard and Baldwin (HKB) (1975), based on Least Squares estimator, has been used for the selection of the k values, building on robust estimators:

$$k = \frac{ps_{Robust}^2}{\hat{\beta}'_{Robust} \hat{\beta}_{Robust}} \dots\dots\dots (3.3.2)$$

where p is the number of regressors and s_{Robust}^2 is the robust scale estimator.

Weighted ridge estimator (Wr)

Weighted Ridge Estimator $\hat{\beta}_{WR}$ can be computed using the following formula:

$$\hat{\beta}_{WR} = (X'WX + KI)^{-1} X'WY$$

where k is determined from the data using:

$$k = \frac{ps_w^2}{\hat{\beta}'_W \hat{\beta}_W}$$

$$s_w^2 = \frac{(Y - X \hat{\beta}_W)' (Y - X \hat{\beta}_W)}{n - p} \text{ and}$$

$\hat{\beta}_W$ denotes the coefficient estimates from the Weighted Least Square estimators.

Ridge M-Estimator (Rm)

The Ridge M-estimator of the parameter β is given by:

$$\hat{\beta}_{RM} = (X'X + KI)^{-1} X'Y$$

where the biasing parameter is given as

$$k = \frac{ps_M^2}{\hat{\beta}'_M \hat{\beta}_M}$$

where the M-estimator procedure of is used rather than Least Square estimator in computing and values in order to reduce the effect of non-normality on the value choosen for , and the value can be written as: [16-18]

$$s_M^2 = \frac{(Y - X \hat{\beta}_M)' (Y - X \hat{\beta}_M)}{n - p}$$

Results and Discussion

Data simulation

The data utilized for this work were simulated from R (www.cran.r-project.org) statistical package. The three predictor variables are generated form the multivariate normal package as follows:

$\beta = [\beta_0 = 25, \beta_1 = 65, \beta_2 = 15, \beta_3 = 45]$; is the vector of the true population value that the data set is simulated from.

$$\rho_{ij} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix}; \text{ is the specified correlation and it varies}$$

between Tolerable Non-Orthogonal Correlation Point and the extreme correlation

$\sigma_j^2 = [\sigma_1^2 = 2, \sigma_2^2 = 3, \sigma_3^2 = 4]$ is the variance vector of the predictors.

$\mu_j = [\mu_1 = 10, \mu_2 = 15, \mu_3 = 20]$ is the specified mean vector of the predictors.

$$Sigma = \left\{ \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix} \right\} \text{ Variance - covariance matrix}$$

$$X = \text{mvrnorm}(n, c(\mu_1, \dots, \mu_3), Sigma)$$

$$X = [X_{i1}, X_{i2}, X_{i3}]$$

One important factor in this study is the disturbances distribution. Therefore, the residual term was simulated from two distributions, namely:

- The univariate normal distribution with mean 0 and standard deviation $\sigma = 10$, and
- Cauchy distribution with median zero and scale parameter one

$$e1 = \text{rnorm}(n * p, 0, \sigma)$$

$$e2 = \text{rcauchy}(n * (1 - p), 0, 1)$$

$$e = c(e1, e2)$$

Where p represents the proportion of error from the normal distribution.

$$Y = X\beta + \epsilon$$

And the response is simulated with the relationship given below

$$Y_i = 25 + 65X_{i1} + 15X_{i2} + 45X_{i3} + \epsilon$$

$Y = [Y_1, Y_2, \dots, Y_n]$ is the vector of the generated dependent variable.

Correlation structure

For the TNCP (tolerable non orthogonal correlation)

$$\begin{bmatrix} 1 & 0.0005 & 0.0001 \\ 0.0005 & 1 & 0.0009 \\ 0.0001 & 0.0009 & 1 \end{bmatrix}$$

For the severe correlation

$$\begin{bmatrix} 1 & 0.9910 & 0.9509 \\ 0.9372 & 1 & 0.9809 \\ 0.9509 & 0.9910 & 1 \end{bmatrix}$$

In order to compare the performance of the estimators, the Mean Square Error (MSE) was computed for the data set at various sample sizes and at different iteration. In each case, the average MSEs for each sample were reported. The true parameter values were compared with the regression estimates at different sample sizes (Table 1).

Table 1: Summary of Estimate of Coefficients and Standard Error at Various Sample Size for Tolerable Non-Orthogonal Correlation.

Sample size	True value	OLS		RIDGE		Weighted Ridge		M-Estimator		Robust Ridge	
		Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err
n = 20	$\beta_0 = 25$	25.48	22.59	26.72	17.08	4.33	576.13	25.44	17.65	22.50	15.19
	$\beta_1 = 65$	65.00	1.06	64.96	0.78	64.50	121.40	65.01	0.81	65.06	0.76
	$\beta_2 = 15$	14.98	0.86	14.97	0.66	15.18	57.27	14.98	0.68	15.05	0.63
	$\beta_3 = 45$	44.98	0.75	44.95	0.57	46.03	35.76	44.98	0.59	45.05	0.54
n = 50	$\beta_0 = 25$	24.87	10.06	25.37	10.76	28.79	1645.82	24.94	11.06	23.57	10.23
	$\beta_1 = 65$	65.01	0.47	65.00	0.49	65.75	145.27	65.02	0.51	65.04	0.49
	$\beta_2 = 15$	14.99	0.38	14.99	0.41	12.74	114.00	14.99	0.43	15.03	0.41
	$\beta_3 = 45$	44.99	0.33	44.98	0.35	45.78	57.25	44.99	0.36	45.02	0.34
n = 100	$\beta_0 = 25$	24.91	6.55	25.16	7.53	-1131	50483.14	24.92	7.78	24.24	7.34
	$\beta_1 = 65$	64.99	0.30	64.99	0.35	69.01	247.98	64.99	0.36	65.01	0.35
	$\beta_2 = 15$	14.99	0.25	14.99	0.29	86.39	3161.31	15.00	0.29	15.01	0.28
	$\beta_3 = 45$	45.00	0.21	44.99	0.24	53.28	299.90	45.00	0.25	45.01	0.24
n = 200	$\beta_0 = 25$	24.91	5.42	25.03	5.42	26.39	1237.53	24.92	5.57	24.56	5.35
	$\beta_1 = 65$	65.00	0.25	64.99	0.26	67.25	72.45	65.00	0.26	65.00	0.26
	$\beta_2 = 15$	15.00	0.20	15.00	0.20	15.49	24.51	15.00	0.20	15.00	0.20
	$\beta_3 = 45$	45.00	0.18	44.99	0.17	43.51	49.04	45.00	0.18	45.00	0.17
n = 500	$\beta_0 = 25$	24.95	3.30	25.00	3.32	137.62	4273.49	24.93	3.41	24.81	3.30
	$\beta_1 = 65$	65.00	0.15	64.99	0.16	59.84	236.86	65.00	0.16	65.00	0.16
	$\beta_2 = 15$	15.00	0.12	15.00	0.12	14.19	73.26	15.00	0.13	15.00	0.12
	$\beta_3 = 45$	44.99	0.11	44.99	0.11	42.69	61.61	44.99	0.11	45.00	0.11
n = 1000	$\beta_0 = 25$	25.06	2.36	25.08	2.36	440.68	18471.47	25.05	2.42	24.99	2.35
	$\beta_1 = 65$	65.00	0.11	65.00	0.11	61.54	127.69	65.00	0.11	65.00	0.11
	$\beta_2 = 15$	14.99	0.09	14.99	0.09	13.05	108.28	14.99	0.09	14.99	0.09
	$\beta_3 = 45$	44.99	0.07	44.99	0.07	29.67	696.97	44.99	0.08	44.99	0.07

The Variance Inflation Factor for all the variables is as follows:

$vif(X_1) = 1.000000$; $vif(X_2) = 1.000001$; $vif(X_3) = 1.000001$ Distribution of Error: 100% Normal.

From Table 1 it is observed that the value of estimated regression coefficients for all the estimators considered is an unbiased estimate of the population parameter when the regressors are orthogonal, and the disturbances of the regression equation

are independent, identically distributed normal random variables, except for Weighted Ridge which is over estimating the parameter (Table 2).

Table 2: Mean Square Error of the estimators at various sample sizes.

Sample Size	Ols	Ridge	Weighted Ridge	M-Estimator	Robust Ridge
20	74.3973	74.0518	5.7278	78.2738	73.4124
50	29.2738	29.1267	123.271	30.729	31.6666
100	14.3101	14.2662	31.7915	15.201	32.8128

200	7.3932	7.3782	3807.788	7.8055	4.5684
500	2.7757	2.7734	446.7766	2.9294	10.0698
1000	1.4071	1.4076	222.1396	1.4745	0.0914

From Table 2, it is observed that the MSEs asymptotically tend to zero as the sample sizes become large, which is indicating the consistency of the estimators except that of the Weighted Ridge which is not stable (Table 3).

Table 3: Summary of Estimate of Coefficients and Standard Error At Various Sample Size for Severe Correlation Level.

Sample size	True value	OLS		RIDGE		Weighted Ridge		M-Estimator		Robust Ridge	
		Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err
n = 20	$\beta_0=25$	24.83	34.39	19.65	30.20	-7.87	411.46	24.62	34.12	14.63	24.23
	$\beta_1=65$	65.11	21.56	55.82	12.28	7.81	1389.87	65.02	21.45	56.91	15.29
	$\beta_2=15$	14.89	28.01	27.39	15.11	36.44	610.84	14.99	27.96	24.88	20.10
	$\beta_3=45$	45.03	10.49	40.55	6.30	59.49	1013.19	45.01	10.48	42.14	8.00
n = 50	$\beta_0=25$	24.86	23.95	20.61	19.80	25.99	654.75	24.97	21.47	18.28	17.85
	$\beta_1=65$	64.83	15.01	58.24	9.69	56.16	440.02	64.79	13.47	59.73	11.28
	$\beta_2=15$	15.19	19.50	24.03	12.18	28.85	595.92	15.24	17.52	21.42	14.70
	$\beta_3=45$	44.94	7.30	41.82	4.72	38.97	266.41	44.91	6.53	43.14	5.62
n = 100	$\beta_0=25$	25.14	15.53	22.06	14.08	50.17	2364.85	25.02	14.88	20.98	13.32
	$\beta_1=65$	65.20	9.73	60.66	7.71	97.70	2853.07	65.21	9.36	62.01	8.46
	$\beta_2=15$	14.67	12.65	20.76	9.84	-41.39	3727.43	14.64	12.19	18.58	11.03
	$\beta_3=45$	45.13	4.73	42.99	3.77	70.48	1328.16	45.15	4.60	44.00	4.20
n = 200	$\beta_0=25$	24.87	9.96	22.99	10.18	172.90	6263.96	24.89	10.62	22.52	9.91
	$\beta_1=65$	64.97	6.24	62.27	5.93	129.90	2793.06	65.01	6.67	63.19	6.25
	$\beta_2=15$	15.01	8.11	18.63	7.65	-78.81	3775.57	14.97	8.70	17.19	8.15
	$\beta_3=45$	45.00	3.04	43.73	2.87	76.02	1542.98	45.02	3.25	44.37	3.06
n = 500	$\beta_0=25$	24.98	6.47	24.12	6.51	123.51	3760.35	24.99	6.77	23.96	6.44
	$\beta_1=65$	64.92	4.05	63.70	4.00	96.37	1558.78	64.92	4.30	64.15	4.09
	$\beta_2=15$	15.11	5.26	16.75	5.14	-3.51	1733.39	15.11	5.54	16.06	5.27
	$\beta_3=45$	44.94	1.97	4.37	1.89	38.61	485.40	44.95	2.03	44.67	1.94
n = 1000	$\beta_0=25$	25.07	4.61	24.62	4.57	69.41	2559.44	25.08	4.72	24.55	4.55
	$\beta_1=65$	64.90	2.89	64.26	2.84	100.78	1695.05	64.90	2.97	64.51	2.87
	$\beta_2=15$	15.14	3.75	15.99	3.68	-42.77	2339.81	15.15	3.86	15.62	3.73
	$\beta_3=45$	44.93	1.40	44.63	1.36	67.76	971.89	44.92	1.43	44.79	1.38

From Table 3, it is observed that the result obtained from Weighted Ridge is inconsistent with true parameter and result obtained from other estimators considered in this study. It is also observed that as the sample sizes become large, the result obtained from other estimators except Weighted Ridge converges to the true parameter (Table 4).

Table 4: Mean Square Error of the estimators at various sample sizes.

Sample Size	Ols	Ridge	Weighted Ridge	M-Estimator	Robust Ridge
20	592.7164	404.2302	141.8168	628.9521	158.385
50	235.1698	203.2862	11573.14	247.9795	92.6825
100	113.7335	108.3811	11509.46	119.7368	44.1539
200	57.8141	57.9948	2532.823	60.951	28.425
500	23.1925	23.6267	78656.25	24.7985	10.0074
1000	11.391	11.5671	136089.3	12.046	0.4928

From Table 4, it is observed that the MSEs asymptotically tend to zero as the sample sizes become large, which is indicating the consistency of the estimators except that of the Weighted Ridge

which is not stable. Where the MSE of Robust Ridge clearly show that it is more efficient than other estimator considered (Figure 1).

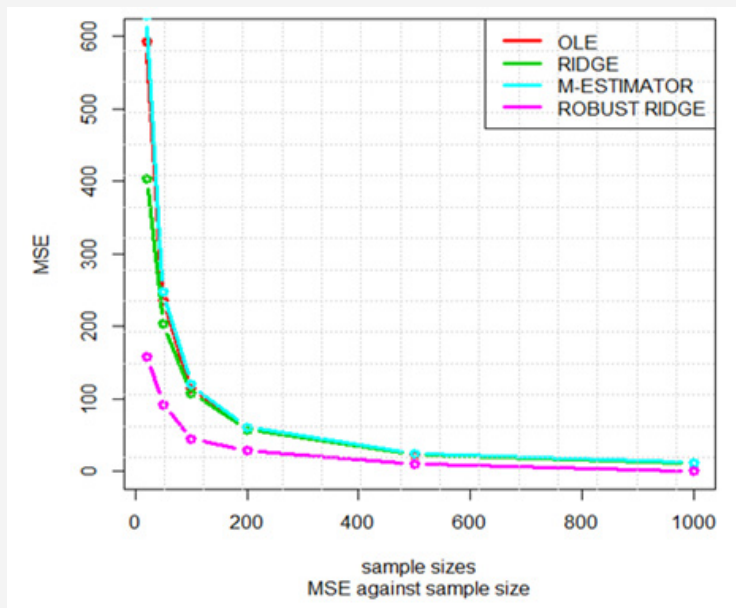


Figure 1: MSEs Plot for Various Estimators across the Selected Sample Sizes.

From the plot above, it is observed that the MSE asymptotically tend to zero as the sample sizes becomes large, which indicate the consistency of the estimators (Table 5).

Table 5: Summary of Estimate of Coefficients and Standard Error At Various Sample Size for Tolerable Non-Orthogonal Correlation.

Sample size	True value	OLS		RIDGE		Weighted Ridge		M-Estimator		Robust Ridge	
		Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err
n = 20	$\beta_0=25$	42.23	5.85	65.63	211.06	16.09	137.09	25.26	9.44	19.05	15.45
	$\beta_1=65$	64.59	0.27	63.52	7.62	64.96	18.83	64.97	0.45	64.44	4.86
	$\beta_2=15$	14.40	0.22	14.71	2.85	15.14	10.10	14.99	0.37	15.50	2.93
	$\beta_3=45$	44.83	0.19	43.96	5.20	45.25	9.65	45.00	0.31	45.16	1.96
n = 50	$\beta_0=25$	64.19	26.80	76.95	252.58	16.54	298.22	24.88	4.65	18.36	15.56
	$\beta_1=65$	64.70	1.26	63.18	8.41	67.10	73.37	65.00	0.21	64.31	6.00
	$\beta_2=15$	14.46	1.03	14.59	3.28	13.34	74.14	15.00	0.17	15.49	3.48
	$\beta_3=45$	43.66	0.89	43.68	5.93	45.78	34.12	45.00	0.15	45.12	2.95
n = 100	$\beta_0=25$	29.54	12.28	67.11	206.49	24.98	268.64	25.01	3.11	18.36	15.68
	$\beta_1=65$	64.82	0.57	63.40	7.76	64.31	15.16	65.00	0.14	64.44	4.84
	$\beta_2=15$	14.79	0.47	14.61	3.38	15.17	14.26	14.99	0.11	15.43	3.48
	$\beta_3=45$	45.00	0.40	43.96	5.41	45.23	8.74	45.00	0.10	45.24	2.52
n = 200	$\beta_0=25$	25.24	20.50	69.13	212.38	10.50	1003.29	25.00	2.06	18.02	15.81
	$\beta_1=65$	65.73	0.96	63.40	7.74	65.84	65.43	64.99	0.10	64.45	4.96
	$\beta_2=15$	15.17	0.78	14.76	2.85	15.98	58.00	15.00	0.08	15.63	3.76
	$\beta_3=45$	44.55	0.68	43.82	5.40	45.01	27.82	44.99	0.06	45.11	2.58
n = 500	$\beta_0=25$	22.55	87.06	64.47	198.39	139.29	5408.07	24.94	1.29	18.05	15.94
	$\beta_1=65$	65.04	4.09	63.60	7.14	66.20	101.30	65.00	0.06	64.59	4.49
	$\beta_2=15$	14.81	3.34	14.58	3.30	14.48	104.57	15.00	0.05	15.31	3.22
	$\beta_3=45$	45.27	2.89	44.06	4.82	39.35	229.38	45.00	0.04	45.31	2.41

n = 1000	$\beta_0=25$	18.28	10.07	66.23	207.10	17.25	462.07	24.96	0.88	18.43	15.84
	$\beta_1=65$	65.87	0.47	63.50	7.62	65.31	39.88	65.00	0.04	64.49	5.14
	$\beta_2=15$	15.10	0.38	14.60	3.06	14.92	20.02	14.99	0.03	15.39	3.42
	$\beta_3=45$	44.84	0.33	43.99	5.24	45.15	20.60	45.00	0.03	45.22	2.45

From the above table, it is observed that the result obtained from Robust M-estimation is an unbiased estimate of the true parameter, while result obtained from other estimators considered

in this study are biased which is as a result of non-normal error (Table 6).

Table 6: Mean Square Error of the estimators at various sample sizes.

Sample Size	Ols	Ridge	Weighted Ridge	M-Estimator	Robust Ridge
20	77988.4306	11568.1473	40.3088	22.4051	0.3960
50	954096.0362	16647.0311	27.7816	5.4475	9.6257
100	46585.9499	11124.3117	398.0189	2.4362	6.8527
200	11766.5161	11783.7524	19.8095	1.0661	4.2687
500	19429.5031	10246.5889	108.0036	0.4216	109.0263
1000	75786.3191	11167.8509	538.9687	0.1979	1.3369

From the Table above, it is observed that the MSE of M-estimation asymptotically tend to zero as the sample sizes become large, which is indicating the consistency of the estimator compared to other

estimators considered in this study, while the result of MSE based on OLS is the worst among the estimators. Robust M-estimator is more efficient than other estimator considered (Table 7).

Table 7: Summary of Estimate of Coefficients and Standard Error at Various Sample Size for Severe Correlation Level.

Sample size	True value	OLS		RIDGE		Weighted Ridge		M-Estimator		Robust Ridge	
		Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	Std Err
n = 20	$\beta_0=25$	49.49	152.51	47.32	184.34	16.59	312.61	24.71	9.61	15.57	15.80
	$\beta_1=65$	42.54	95.61	56.66	10.88	50.12	256.00	64.96	5.90	56.62	12.91
	$\beta_2=15$	49.67	124.21	25.01	11.19	32.37	348.60	15.06	7.67	24.17	14.01
	$\beta_3=45$	28.97	46.52	40.51	6.85	39.59	142.62	44.97	2.89	42.59	6.33
n = 50	$\beta_0=25$	23.64	12.51	48.26	183.32	-0.72	1070.22	24.93	4.49	15.40	15.44
	$\beta_1=65$	59.18	7.84	56.51	10.58	47.98	659.26	65.05	2.76	56.59	12.57
	$\beta_2=15$	23.80	10.19	25.23	10.92	34.29	721.30	14.93	3.58	24.41	14.00
	$\beta_3=45$	41.39	3.81	40.43	6.80	40.72	174.24	45.02	1.33	42.55	6.06
n = 100	$\beta_0=25$	13.71	5.54	42.25	151.56	2.44	822.40	25.01	3.01	15.87	16.08
	$\beta_1=65$	56.90	3.47	56.70	9.96	55.49	391.08	65.02	1.85	56.90	12.17
	$\beta_2=15$	27.61	4.51	25.27	10.56	31.15	434.26	14.97	2.39	24.04	13.77
	$\beta_3=45$	40.17	1.69	40.60	6.44	38.50	241.57	45.00	0.89	42.68	6.17
n = 200	$\beta_0=25$	14.68	14.50	49.12	177.18	21.13	599.28	24.98	1.99	15.70	15.90
	$\beta_1=65$	57.49	9.09	56.44	10.76	66.13	362.45	64.96	1.26	56.58	12.75
	$\beta_2=15$	24.59	11.81	25.13	10.76	13.60	453.08	15.05	1.64	24.18	13.86
	$\beta_3=45$	42.09	4.42	40.49	6.89	45.70	158.37	44.97	0.61	42.75	6.21
n = 500	$\beta_0=25$	49.36	6.86	44.16	177.24	3.07	2536.91	25.00	1.20	15.96	16.08
	$\beta_1=65$	84.28	4.30	56.76	10.34	82.29	1381.72	65.00	0.77	56.93	12.75
	$\beta_2=15$	-10.60	5.59	25.08	10.87	-16.01	1486.49	15.00	0.99	23.96	14.18
	$\beta_3=45$	53.39	2.09	40.64	6.53	60.02	544.96	44.99	0.37	42.68	6.37

n = 1000	$\beta_0 = 25$	50.49	8.27	40.06	151.17	147.41	6141.39	25.00	0.88	15.43	16.45
	$\beta_1 = 65$	88.19	5.18	56.53	10.44	161.74	4535.72	65.00	0.54	56.83	12.82
	$\beta_2 = 15$	-16.10	6.74	25.12	10.76	-100.28	5484.48	14.99	0.71	23.74	13.79
	$\beta_3 = 45$	55.41	2.52	40.84	6.33	76.59	1541.79	45.00	0.26	42.89	6.42

The Variance Inflation Factor for all the variables is as follows: $vif(X_1) = 164.7018$; $vif(X_2) = 416.9834$; $vif(X_3) = 78.0039$, Distribution of Error: 100%

From the above table, it is observed that the result obtained from Robust M-estimation is an unbiased estimate of the true parameter, while result obtained from other estimators considered

in this study are biased which is as a result of effect of collinearity and non-normal error (Table 8).

Table 8: Mean Square Error of the estimators at various sample sizes.

Sample Size	Ols	Ridge	Weighted Ridge	M-Estimator	Robust Ridge
20	1014136.5510	8736.6686	468.1883	48.6734	268.4810
50	64818.9575	8652.7089	208.9701	10.6160	80.6441
100	175584.9516	5926.6005	877.8945	4.7518	65.8274
200	111942.0554	8109.8389	54.4042	2.1653	22.1353
500	1219251.7116	8056.3711	1358.4787	0.7986	11.5062
1000	733579.6179	5881.7098	230.1928	0.4148	9.9903

From the table above, it is observed that the MSE of M-estimation asymptotically tend to zero as the sample sizes become large, which is indicating the consistency of the estimator compared to other

estimators considered in this study, while the result of MSE based on OLS is the worst among the estimators. Robust M-estimator is more efficient than other estimator considered (Figure 2).

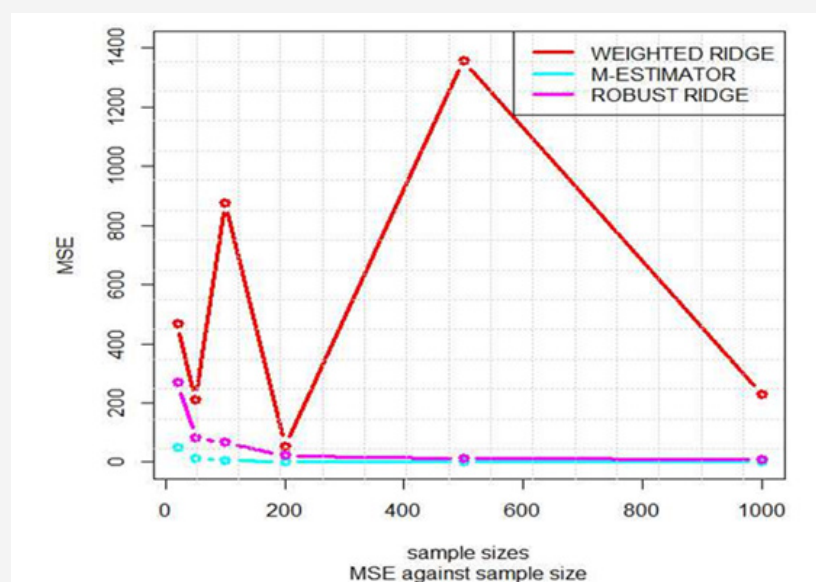


Figure 2: MSEs Plot for Various Estimators across the Selected Sample Sizes.

Conclusion

From the simulation studies it can be concluded to a reasonable extent that In the presence of both multicollinearity and non-normal error the OLS estimators produces relatively large Mean Squares Errors, hence it gives unstable parameter estimates. the sample size is inversely proportional to the mean square error of estimate i.e. the larger the sample size, the lower the mean square error and vice versa. To make any reasonable predictions and inferences in any multiple linear regressions, one must be certain that the predictors are non-orthogonal and the percentages of data that are outliers are minima before the commencement of the

analysis. This study reveal that the consequence of collinearity in the presence of non-normal error is more severe compared to when the error distribution is normal. Finally, the results of comparisons shows that for the condition of collinearity, Robust Ridge estimates are more efficient than the other estimators considered. While for the condition of collinearity and non-normal error M-estimator produces estimates that were more efficient and precise.

Recommendation

This research work will help present and future researchers in choosing or determining the appropriate estimator for estimating the regression parameters when the traditional method for

estimating the parameters is unreliable in the presence of both multicollinearity and non-normal error problem in a data set.

Acknowledgement

None.

Conflict of Interest

No conflict of interest

References

1. Askin GR, Montgomery DC (1980) Augmented robust estimators. *Technometrics* 22(3): 333-341.
2. Belsley DA, Kuh E, Welsch RE (1980) *Regression Diagnostics: Identifying influential Data and Sources of Collinearity*. John Wiley & Sons, New York.
3. Babalola AM, Obubu M, Otekunrin OA, Obiora-Ilono H (2018) Selection and Validation of Comparative Study of Normality Test. *American Journal of Mathematics and Statistics* 8(6): 190-201.
4. Babalola AM, Obubu M, George OA, Ikediwuwa CU (2018) Reprospective Analysis of Some Factors Responsible for Infant Mortality in Nigeria: Evidence from Nigeria Demographic and Health Survy (NDHS). *American Journal of Mathematics and Statistics* 8(6): 184-189.
5. Moawad El-Fallah Abd El-Salam (2013) The Efficiency of some Robust Ridge Regression for Handling Multicollinearity and Non-normals Errors Problems. *Applied Mathematical Sciences* 7(77): 3831-3846.
6. Farrar, Glauber (1967) Multicollinearity in Regression Analysis. *Review of Economics and Statistics* 49: pp. 92-107
7. Gujarati DN (2004) *Basic Econometric* 4th edition. Tata McGraw-Hill, New Delhi, India.
8. Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics* 12: pp. 55-67.
9. Hoerl AE, Kennard RW (1970) Ridge regression: Applications to non-orthogonal problems. *Technometrics* 12: pp. 69-82.
10. Hoerl AE, Kennard RW, Baldwin KF (1975) Ridge regression: Some simulations. *Communications in Statistical Theory* 4(2): 105-123.
11. Holland PW (1973) Weighted ridge regression: Combining ridge and robust regression methods. *NBER Working Paper* 11: 1-19.
12. Huber PJ (1964) Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35(1): 73-101.
13. Montgomery DC, eck EA (1982) *Introduction to Linear Regression Analysis*. Wiley, New York, USA.
14. Rawlings JO, Pantula SG, Dickey DA *Applied Regression Analysis. A Research Tool*.
15. Samkar Hatice, Alpu Ozlem (2010) Ridge Regression Based on Some Robust Estimators. *Journal of Modern Applied Statistical Methods* 9(2): 17.
16. Silvapulla MJ (1991) Robust ridge regression based on an M-estimator. *Australian Journal of Statistics* 33(3): 319-333.
17. Obubu M, Babalola AM, Ikediwuwa CU, Amadi PE (2019) Parametric Survival Modeling of Maternal Obstetric Risk; a Censored Study. *American Journal of Mathematics and Statistics* 9(1): 17-22.
18. Obubu M, Babalola AM, Ikediwuwa CU, Amadi PE (2018) Modelling Count Data; A Generalized Linear Model Framework. *American Journal of Mathematics and Statistics* 8(6): 179-183.