Performance Evaluation of Some Estimators of Linear Models with Collinearity and Non–Gaussian Error

W. B. Yahya[*]; M. K. Garba; A. G. Ajayi[¶]; K. A. Dauda[†]; O. R. Olaniran; N. F. Gatta

Department of Statistics,
University of Ilorin,
Ilorin, Nigeria.
e-mail: dr.yah2009@ymail.com[*]; adefemiajayi@ymail.com[¶]

[†]Department of Statistics and Mathematical Sciences,
Kwara State University,
Malete, Nigeria

Abstract **- Among typical challenges in numerous multiple linear regression models are those of multicollinearity and non–normal disturbances which have created undesirable consequences for the ordinary least squares (OLS) estimator which is the popular and naïve technique for estimating linear models. Thus, it appears so critical to combine strategies for estimating regression models in order to muddle through while these challenges are present. In this study, the strength of some methods of estimating classical linear regression model in the presence of multicollinearity and non-normal error structures were investigated. The conventional Least Squares (LS), Ridge Regression (RR), Weighted Ridge (WR), Robust M-estimation (M) and Robust Ridge Regression (RRR) methods taking into accounts M-estimation procedures were considered in this study. Results from Monte-Carlo study revealed the superiority of the RRR estimator over others using Mean Squared Errors (MSE) of parameter estimates and Absolute Bias (AB) as assessment criteria among others over various considerations for the distribution of the disturbance term and levels of multicollinearity. The study concluded that whenever linear regression modeling is intended and multicollinearity among the regressors and non-spherical disturbance structure on the response variable are suspected in a data set, the RRR estimator should be adopted in order to ensure optimal efficiency.**

Keywords: *Non-normal disturbances, Collinearity, Robust M-estimation, Ridge and Weighted Ridge Regression*

I. Introduction

A regression estimator is said to be "robust" the moment it is capable of typically providing parameters estimates that are reasonably unbiased and efficient even when one or more of the assumptions underlining its usage is not completely met. Conversely, a large violation of one or more required assumptions might results into poor estimates of the parameters which may consequently lead to wrong conclusions and inference [1, 2].

Two major problems are often of paramount concern in regression analysis: multicollinearity and non–normal error structure.

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 [3]. 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 easily made.

Another common problem in regression estimation methods is that of non–normal errors. The term simply means that the distribution of errors have fatter tails than that of the normal distribution. These fat–tailed distributions are more prone than the normal distribution to produce outliers, or extreme observations in the data [4].

The Gauss-Markov Theorem says that Ordinary Least Squares (OLS) estimates for coefficients are Best Linear Unbiased Estimator (BLUE) when the errors are normal and homoscedastic. When errors are non-normal, the 'E' property of Efficiency will no longer holds for the estimators and the standard errors of estimate will be biased.

The Ordinary Least Squares (OLS) estimators 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 and identically distributed normal random variables. It is not every time we would have ideal situation like this. This work examines the effects of non–normal error terms on the efficiency of the estimates of the marginal regression coefficients in multiple regression modellingThe aim of this study is to examine the effect of multicollinearity and non-normal error problem.

II. RESEARCH METHODOLOGY

A. Model Identification

The classical multiple linear regression model is defined by the equation,

$$\boldsymbol{y} = \mathbf{1}\beta_0 + \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where, $\boldsymbol{y}$is a $n \times 1$ vector of observations on the dependent variable, β_0 is an unknown constant, $\boldsymbol{X}$ is a $n \times p$ matrix consisting of n observations on p variables, $\boldsymbol{\beta}$ isa $p \times 1$ vector of unknown regression coefficients, and $\boldsymbol{\varepsilon}$ is a $n \times 1$ vector of errors identically and independently distributed with mean zero and variance σ^2.

1. ***Ordinary Least Squares***

When the matrix $\boldsymbol{X}$has a full rank of p, the OLS estimator $\hat{\beta}_{OLS}$ can be obtained by minimizing the sum of squared residuals,

$$\hat{\varepsilon}'\hat{\varepsilon} = (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \tag{2}$$

hence,

$$\widehat{\boldsymbol{\beta}}_{OLS} = (X'X)^{-1}(X'y) \tag{3}$$

The variance and the Mean Squared Error (MSE) of the OLS estimator are given in equation (4) and (5)

$$Var(\hat{\beta}) = (X'X)^{-1}\sigma^2 \tag{4}$$

$$MSE = E(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \tag{5}$$

2. ***Ridge Regression***

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 of matrix $(X'X)$ before estimating the coefficients. Reference [5] and [6] proposed the ridge estimator

$$\beta_R = (X'X + kI)^{-1}(X'y) \qquad k > 0 \tag{6}$$

while k is the biasing parameter that we need to choose.

Several authors have proposed a number of procedures for estimating the value of k. Reference [7] have suggested that an appropriate choice of k is

$$k = \frac{p\hat{\sigma}^2}{\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\beta}}} \tag{7}$$

where $\widehat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are obtained from the least squares solution.

The variance and the mean squared error of the ridge estimator are given in (8) and (9)

$$Var(\hat{\beta}) = (X'X + kI)^{-1}X'X(X'X + kI)^{-1}\sigma^2 \tag{8}$$

$$MSE = E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} - \boldsymbol{\beta})'(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}} - \boldsymbol{\beta}) \tag{9}$$

3. ***M-Estimation***

Robust regression is an important tool for analyzing data that are contaminated with outliers. M-estimation introduced by [8] is not as vulnerable as least squares to unusual data. Consider the linear model in (1), the general M-estimator minimizes the objective function,

$$\sum_{i=1}^{n} \rho(e_i) = \sum_{i=1}^{n} \rho(y_i - x_i'\beta) \tag{10}$$

An iterative solution (called iterative reweighted least-squares, IRLS) is required to obtain the parameter estimates.

The variance and the mean squared error of the M-estimator are given in equation (11) and (12)

The asymptotic covariance matrix of $\hat{\beta}$ is

$$Var(\hat{\beta}) = \frac{E(\psi^2)}{[E(\psi')]^2}(X'X)^{-1} \tag{11}$$

using$\sum[\psi(e_i)]^2$ to estimate $E(\psi^2)$, and $\sum[\psi'(e_i)/n]^2$ to estimate $[E(\psi')]^2$ produces the estimated asymptotic covariance matrix where $\psi = \rho'$ be the derivative of ρ.

$$MSE = E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{M}} - \boldsymbol{\beta})'(\widehat{\boldsymbol{\beta}}_{\boldsymbol{M}} - \boldsymbol{\beta}) \tag{12}$$

Reference [9] discussed augmented robust estimates as a way of combining biased and robust regression techniques. This combined procedure is based on the fact that robust estimates can be computed using Weighted Least Squares procedure. When both outliers and multicollinearity occur in a data set, it would be preferred to combine methods for dealing with these problems simultaneously. This technique was adopted while fitting the robust ridge regression model in this work.

4. ***Weighted Ridge Estimator*** (WR)

Weighted Ridge Estimator$\widehat{\boldsymbol{\beta}}_{WR}$, can be computed using the formula in (13)

$$\widehat{\boldsymbol{\beta}}_{WR} = (X'WX + kI)^{-1}(X'Wy) \tag{13}$$

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

$$w_{ii} = \begin{cases} \dfrac{1}{\hat{e}_i} & if\, \hat{e}_i \neq 0 \\ 1 & if\ \hat{e}_i = 0 \end{cases} \tag{14}$$

k is determined from the data using:

$$k = \frac{pS_{WR}^2}{\widehat{\boldsymbol{\beta}}_{WR}{}'\widehat{\boldsymbol{\beta}}_{WR}} \quad (15)$$

The variance and the mean squared error of weighted ridge estimator are given in equation (16) and (17)

$$S_{WR}^2 = \frac{(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{WR}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{WR}})}{n-p} \quad (16)$$

$$MSE = E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{WR}} - \boldsymbol{\beta})'(\widehat{\boldsymbol{\beta}}_{\boldsymbol{WR}} - \boldsymbol{\beta}) \quad (17)$$

5. ***Robust Ridge Regression Estimator*** (RRR)

The robust ridge regression estimate of the parameter $\boldsymbol{\beta}$ is given by:

$$\widehat{\boldsymbol{\beta}}_{RRR} = (X'X + kI)^{-1}(X'y) \quad (18)$$

k is determined from the data using:

$$k = \frac{pS_M^2}{\widehat{\boldsymbol{\beta}}_M{}'\widehat{\boldsymbol{\beta}}_M} \quad (19)$$

where $\widehat{\boldsymbol{\beta}}_M$ is the estimated parameter vector using the M-Estimation method.

The variance and the mean squared error of the robust ridge regression estimator are given in equation (20) and (21)

$$S_M^2 = \frac{(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{M}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{M}})}{n-p} \quad (20)$$

$$MSE = E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{RRR}} - \boldsymbol{\beta})'(\widehat{\boldsymbol{\beta}}_{\boldsymbol{RRR}} - \boldsymbol{\beta}) \quad (21)$$

B. **Simulation Study**

For the purpose of this study, three sets of predictors $\boldsymbol{X} = (X_1, X_2, X_3)$were simulated from multivariate normal distribution.

Six sample sizes which include; [n = 20, 50, 100, 200, 500, 1000] were considered to verify the consistency of the estimators. One important factor in this study is the distribution of the models' error term. Therefore, the residual term ε_i was simulated from a fat-tailed distribution,the Cauchy distribution with median zero and scale parameter one other than from the usual normal distribution.

The response variable y was simulated with the relationship given below:

$$y_i = 25 + 65X_{1i} + 15X_{2i} + 45X_{3i} + \varepsilon_i$$

The correlation structure imposed among the three predictors is of two forms as presented under Case I and Case II below:

CASE I (No Multicollinearity)

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

and

CASE II (Presence of Multicollinearity)

$$\rho_{ij} = \begin{bmatrix} 1 & 0.9910 & 0.9509 \\ 0.9910 & 1 & 0.9809 \\ 0.9509 & 0.9809 & 1 \end{bmatrix}$$

to capture the two cases of orthogonality (non-correlated structure) and the presence of severe collinearity (non-orthogonality) respectively among the predictors.

III. Results

CASE I: The Variance Inflation Factor (VIF) for all the predictors are as follow:
$vif(X_1) = 1.0000$; $vif(X_2) = 1.0000$; $vif(X_3) = 1.0000$

Error Distribution: Cauchy

Table 1: Summary of Estimates of Coefficients and Standard Errors at Various Sample Sizes

Sample size	True value	Least Squares		Ridge		Weighted Ridge		M-Estimator		Robust Ridge	
		Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
n = 20	β_0=25	42.23	5.85	65.63	211.06	16.09	137.09	**25.26**	**9.44**	19.05	15.45
	β_1=65	64.59	0.27	63.52	7.62	64.96	18.83	**64.97**	**0.45**	64.44	4.86
	β_2=15	14.40	0.22	14.71	2.85	15.14	10.10	**14.99**	**0.37**	15.50	2.93
	β_3=45	44.83	0.19	43.96	5.20	45.25	9.65	**45.00**	**0.31**	45.16	1.96
n = 50	β_0=25	64.19	26.80	76.95	252.58	16.54	298.22	**24.88**	**4.65**	18.36	15.56
	β_1=65	64.70	1.26	63.18	8.41	67.10	73.37	**65.00**	**0.21**	64.31	6.00
	β_2=15	14.46	1.03	14.59	3.28	13.34	74.14	**15.00**	**0.17**	15.49	3.48
	β_3=45	43.66	0.89	43.68	5.93	45.78	34.12	**45.00**	**0.15**	45.12	2.95
n = 100	β_0=25	29.54	12.28	67.11	206.49	24.98	268.64	**25.S01**	**3.11**	18.36	15.68
	β_1=65	64.82	0.57	63.40	7.76	64.31	15.16	**65.00**	**0.14**	64.44	4.84
	β_2=15	14.79	0.47	14.61	3.38	15.17	14.26	**14.99**	**0.11**	15.43	3.48
	β_3=45	45.00	0.40	43.96	5.41	45.23	8.74	**45.00**	**0.10**	45.24	2.52
n = 200	β_0=25	25.24	20.50	69.13	212.38	10.50	1003.29	**25.00**	**2.06**	18.02	15.81
	β_1=65	65.73	0.96	63.40	7.74	65.84	65.43	**64.99**	**0.10**	64.45	4.96
	β_2=15	15.17	0.78	14.76	2.85	15.98	58.00	**15.00**	**0.08**	15.63	3.76
	β_3=45	44.55	0.68	43.82	5.40	45.01	27.82	**44.99**	**0.06**	45.11	2.58

Table 2: Mean Square Error of the estimators considered for non-normal disturbances problem

Sample size	OLS	Ridge	Weighted Ridge	M - Estimator	Robust RR
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	10.0263
1000	75786.3191	11167.8509	538.9687	0.1979	1.3369

CASE II: The Variance Inflation Factor (VIF) for all the predictors are as follow:
$vif(X_1) = 164.0718$; $vif(X_2) = 416.9834$; $vif(X_3) = 78.0039$

Error Distribution: Cauchy

Table 3: Summary of Estimates of Coefficients and Standard Errors at Various Sample Size

Sample size	**True value**	**Least Squares**		**Ridge Regression**		**Weighted Ridge**		**M-Estimator**		**Robust Ridge**	
		Coef	**SE**	**Coef**	**SE**	**Coef**	**SE**	**Coef**	**SE**	**Coef**	**SE**
n = 20	β_0=25	49.49	152.51	47.32	15.57	15.80	312.61	15.57	15.80	**24.71**	**9.61**
	β_1=65	42.54	95.61	56.66	56.62	12.91	256.00	56.62	12.91	**64.96**	**5.90**
	β_2=15	49.67	124.21	25.01	24.17	14.01	348.60	24.17	14.01	**15.06**	**7.67**
	β_3=45	28.97	46.52	40.51	42.59	6.33	142.62	42.59	6.33	**44.97**	**2.89**
n = 50	β_0=25	23.64	12.51	48.26	15.40	15.44	1070.22	15.40	15.44	**24.93**	**4.49**
	β_1=65	59.18	7.84	56.51	56.59	12.57	659.26	56.59	12.57	**65.05**	**2.76**
	β_2=15	23.80	10.19	25.23	24.41	14.00	721.30	24.41	14.00	**14.93**	**3.58**
	β_3=45	41.39	3.81	40.43	42.55	6.06	174.24	42.55	6.06	**45.02**	**1.33**
n = 100	β_0=25	13.71	5.54	42.25	151.56	2.44	822.40	15.87	16.08	**25.01**	**3.01**
	β_1=65	56.90	3.47	56.70	9.96	55.49	391.08	56.90	12.17	**65.02**	**1.85**
	β_2=15	27.61	4.51	25.27	10.56	31.15	434.26	24.04	13.77	**14.97**	**2.39**
	β_3=45	40.17	1.69	40.60	6.44	38.50	241.57	42.68	6.17	**45.00**	**0.89**
n = 200	β_0=25	14.68	14.50	49.12	177.18	21.13	599.28	15.70	15.90	**24.98**	**1.99**
	β_1=65	57.49	9.09	56.44	10.76	66.13	362.45	56.58	12.75	**64.96**	**1.26**
	β_2=15	24.59	11.81	25.13	10.76	13.60	453.08	24.18	13.86	**15.05**	**1.64**
	β_3=45	42.09	4.42	40.49	6.89	45.70	158.37	42.75	6.21	**44.97**	**0.61**
n = 500	β_0=25	49.36	6.86	44.16	177.24	3.07	2536.91	15.96	16.08	**25.00**	**1.20**
	β_1=65	84.28	4.30	56.76	10.34	82.29	1381.72	56.93	12.75	**65.00**	**0.77**
	β_2=15	-10.60	5.59	25.08	10.87	-16.01	1486.49	23.96	14.18	**15.00**	**0.99**
	β_3=45	53.39	2.09	40.64	6.53	60.02	544.96	42.68	6.37	**44.99**	**0.37**
n = 1000	β_0=25	50.49	8.27	40.06	151.17	147.41	6141.39	15.43	16.45	**25.00**	**0.88**
	β_1=65	88.19	5.18	56.53	10.44	161.74	4535.72	56.83	12.82	**65.00**	**0.54**
	β_2=15	-16.10	6.74	25.12	10.76	-100.28	5484.48	23.74	13.79	**14.99**	**0.71**
	β_3=45	55.41	2.52	40.84	6.33	76.59	1541.79	42.89	6.42	**45.00**	**0.26**

.

Table 4: Mean Square Error of the estimators considered

Sample size	OLS	Ridge	WR	M - Estimator	RRR
20	1014136.55	8736.668	468.188	268.4810	**48.6734**
50	64818.957	8652.708	208.970	80.6441	**10.6160**
100	175584.95	5926.600	877.894	65.8274	**4.7518**
200	111942.05	8109.838	54.4042	22.1353	**2.1653**
500	1219251.7	8056.371	1358.47	11.5062	**0.7986**
1000	733579.61	5881.709	230.192	9.9903	**0.4148**

I. Discussions

When the model is suffering from only the problem of non-normal error without collinear predictors, M-estimator provided good estimates of the regression parameters than any of the other estimators considered as shown in Table 1. Indeed, the results in Table 1 showed that the estimated parameter values by the M-estimator converge to their true values as the sample sizes become large.

Results in Table 2 further showed the superiority of M-estimator, with the least MSEs) over others as the sample sizes become large. This simply indicates the relative consistency of the M-estimator compared to other estimators considered in this study. Among all the estimators considered in this study, the performance of OLS was the worst based on MSE. Robust M-estimation is more efficient than other estimator considered when the error term is not Gaussian.

In the presence of the twin problems of collinearity and non-Gaussian error, the results from Table 3 showed that the Robust Ridge Regression (RRR) estimator provided regression estimates that are quite close to the true parameters values than any other estimators considered. The poor performance of other estimators was majorly due to the effects of collinearity and non-normal error structure imposed on the model, and unlike others, the Robust ridge estimator is robust to the twin problems of collinearity and deviation from Gaussian errors.

It can generally be observed from the results in Table 3 that as the sample sizes become large, the result obtained from Robust Ridge Regression based on M-estimation converges to the true parameter values with relatively smaller standard error of estimates when compared to other estimators.

In term of the Mean Square Error (MSE), it can be observed from the results in Table 4 that the Robust Ridge Regression yielded the least MSE at all the chosen sample sizes while its MSEs monotonically tend towards zero as the sample sizes become large. This performance level of RRR underscores its relative consistency compared to other estimators considered in this study.

V. Conclusion

In this study, a simple way of modelling collinear data in the presence of twin problems of multicollinearity and non–normal error using Monte Carlo simulation approach is presented. Results from Monte-Carlo study revealed the superiority of the Robust Ridge Regression (RRR) estimator over others based on Mean Squared Errors of parameter estimates and Absolute Bias as assessment criteria among others.

Finally, the consequence of collinearity in the presence of non-normal error is more severe compared to when the error distribution is normal.

References

[1] Gujarati, D.N. (2004). *Basic Econometrics*. McGraw-hill Companies.

[2] Yahya WB and Olaifa JB (2014): A note on Ridge Regression Modeling Techniques. *Electronic Journal of Applied Statistical Analysis,* **7**(2):343-361.

[3] Yahya WB, Adebayo SB, Jolayemi ET, Oyejola BA and Sanni OOM (2008): Effects of non-orthogonality on the efficiency of seemingly unrelated regression (SUR) models". *InterStat Journals*, 1-29.

[4] El-Fallah, M. and El-Salam A (2013). The efficiency of some robust ridge regression for handling multicollinearity and non-normals errors problems. Applied Mathematical Sciences, 7(77), 3831–3846.

[5] Hoerl, A.E., and Kennard, R.W. (1970a). Ridge regression: Biased estimation for non-orthogonal problems. Technometrics, 12, 55–67.

[6] Hoerl, A.E., and Kennard, R.W. (1970b). Ridge regression: Applications to non-orthogonal problems. Technometrics, 12, 69–82.

[7] Gujarati, D.N. (2004). *Basic Econometrics*. McGraw-hill Companies.

[8] Yahya WB and Olaifa JB (2014): A note on Ridge Regression Modeling Techniques. *Electronic Journal of Applied Statistical Analysis,* **7**(2):343-361.

[9] Yahya WB, Adebayo SB, Jolayemi ET, Oyejola BA and Sanni OOM (2008): Effects of non-orthogonality on the efficiency of seemingly unrelated regression (SUR) models". *InterStat Journals*, 1-29.