STATISTICAL COMPARISON OF DIFFERENT METHODS OF ESTIMATION (THE GENERALIZED LEAST SQUARES, WEIGHTED RIDGE AND, WEIGHTED LEAST SQUARES) IN THE PRESENCE OF HETEROSCEDASTICITY AND NON-NORMAL ERRORS

Igwe, N. O.,
Madu, C.,
Okeahialam, A. H.,
Elem-Uche, O.
Department of Maths/Statistics, Akanu Ibiam Federal Polytechnic,
Unwana, Ebonyi State, Nigeria.

Okereke, C. U.,
Department of Statistics, Michael Okpara University of Agriculture,
Umudike Umuahia-Abia State.

Chinyere Theresa Madubuike
Department of Mathematics, Abia State Polytechnic Aba, Abia State.

ndyigwe@gmail.com, chukwunonsomadu37@gmail.com, kettbanky09@yahoo.com, ohossanna@yahoo.com, urchstat@gmail.com,vinc56@yahoo.com

Corresponding Author's Phone Number: 08064090157

Other Authors: 08039508685, 08068730625, 08030893317, 07062872120, 08032927653.

ABSTRACT

Common problems in multiple regression models are heteroscedasticity and non-normal errors, which produce undesirable effects on the least squares estimators. This study saw reasons to combine different methods of estimation (The Ordinary Least Squares, Weighted Least Squares and Weighted Ridge Regression) to deal with these problems. From a simulation study, the results of comparisons show that for the condition of heteroscedasticity, Weighted Least Squares (WLS) estimates are more efficient than the other estimators considered. This is because its values in root mean square error (0.4788), residual standard error (2,7519) and residual mean absolute deviation (0.1167) has the best linear unbiased estimates. For the condition of heteroscedasticity and non-normal errors, Weighted Least Squares produces estimates that were more efficient and precise.

Keywords: Estimation Methods, Heteroscedasticity, Non-normal errors.

INTRODUCTION

Regression analysis is concerned with the study of the relationship between the explained variable and one or more other explanatory variables. It provides 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 (For given Xs, the mean value of the disturbance term μ_i is zero. For given Xs, the variance of the disturbance term μ_i is constant or homoscedastic) will result in poor estimates and consequently the wrong conclusions being drawn (Asukwo, 2019).

Two important problems are considered in regression analysis: heteroscedasticity and non-normal errors distribution. Heteroscedasticity is the term used to describe cases where the error terms do not have constant variance. The existence of heteroscedasticity is a major concern in the application of regression analysis, as it can invalidate statistical tests of significance that assume that the modeling errors are uncorrelated and uniform-hence that their variances do not vary with the effects being modeled. Heteroscedasticity often occurs when there is a large difference among the sizes of the observations. 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 (Welsch, 2018).

The Gauss-Markov theorem says that Ordinary Least Squares estimates for coefficients are BLUE when errors are normal and homoscedastic. When errors are non-normal, the “E” property (Efficient) no longer holds for the estimators and in small samples, and the standard errors will be biased. The classical assumption needed for the Ordinary Least Squares estimator to be efficient requires that the variance of the error term to be consistent and the same for all observations. In other words, the term is expected to be homoscedastic. When the assumption of homoscedasticity is violated and the variance is different for different observations, we have a problem referred to as heteroscedasticity (Gujarati, 2004).

Heteroscedasticity entails that the Ordinary Least Squares estimators of the population parameters are unbiased and consistent, but the usual standard errors become biased and inconsistent. In the case of heteroscedasticity, observations expected to have error terms with large variances are given a smaller weight than observations thought to have error terms with small variances. Specifically, coefficients are selected which minimize Ordinary Least Squares (It’s a special case of Generalized Least Squares) when the variance of all residuals is the same for all cases. The smaller the error variance, the more heavily the case is weighted. Intuitively, this makes sense; the observations with the smallest error variances should give the best information about the position of the true regression line. Generalized Least Squares (GLS) estimation can be a bit complicated. However, under certain conditions, estimators’ equivalent to those generated by GLS can be obtained using a Weighted Least Squares (WLS) procedure utilizing OLS regression on a transformed version of the original regression model. This method corrects for heteroscedasticity without altering the values of the coefficients. This method may be superior to regular OLS if heteroscedasticity is present, however, if the data is homoscedastic, the standard errors are equivalent to conventional standard errors estimated by OLS. Several modifications of the White method of computing heteroscedasticity-consistent standard errors have been proposed as corrections with superior finite sample properties (Chikezie, 2014). Thus, we typically examine the distribution of the errors to determine whether they are normal.

One of the assumptions of the classical linear regression model is that there is no heteroscedasticity. Breaking this assumption means OLS estimators are not the Best Linear Unbiased Estimators (BLUE) and their variance is not the lowest of all other unbiased estimators. Heteroscedasticity does not cause ordinary least squares coefficient estimators to be biased, although it can cause ordinary least squares estimates of the variance (and thus, standard errors) of the coefficients to be biased, possibly above or below the true or population variance (Richard, 2018). This study focuses on violation of the assumption that there is no linear relationship between the explanatory variables and the disturbances distribution. It is non-normal and there is non-constant variance. This research is to give the idea that in the presence of heteroscedasticity and non-normal error in a regression model, the results obtained from Ordinary Least Squares Estimator are unreliable. The effects and remedy of such influences needs to be studied.

The aim of this work is to examine the effect of heteroscedasticity and non-normal error problem on weighted least squares. The objectives are:

1. To examine some estimators which are resistant to the combined problems of heteroscedasticity and non-normality.
2. To fit the appropriate regression model on data having non-normality.
3. Performing a diagnostic test to evaluate the performance of the estimated model.
4. To compare the result with that of Ordinary Least Squares.

This research will be limited to study the effect of heteroscedasticity and non-normal error problem on the traditional weighted least squares estimator using a set of collinear data set simulated from multivariate normal distribution when the disturbances violated the normality assumption. It help future researchers in choosing the appropriate estimator for estimating the regression parameters when the traditional method for estimating the parameters is unreliable in the presence of both heteroscedasticity and non-normal error problem in a data set.

LINEAR REGRESSION MODEL

According to Okenna (2020), a regression model is said to be linear if it is linear both in variables and parameters. That is, viewing it on a graph, it can be estimated by a straight line. A linear regression model is simple if it contains only one independent variable.

That is, $Y = \beta_0 + \beta_1 X + e$ …….. (1)

A regression equation containing more than one independent variable (the regression equation contains K-independent variables where k > 1 is termed multiple regression model.

That is, $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + e$ ……… (2)

Where β_j; j = 0,1,………, k are unknown parameters, e is the error term.

Suppose that the data X_{ij} have been collected on the variables $X_1, X_2, \ldots, X_k$ and also Y_i has also been collected on Y, then we can re – write the model as follows:

$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + e_i$; i = 1, 2, … , n ………. (3)

ASSUMPTIONS OF CLASSICAL LINEAR REGRESSION MODEL

1. The regression model is linear in parameters.
2. The expected value of the residual given any value of the explanatory variable is zero (i.e. $E(e_i / x_i) = 0$).
3. The variance of the residual term given any value of the explanatory variable are equal (i.e. $Var(e_i / x_i) = \sigma^2$) . This is known as homoscedasticity.
4. The values of the explanatory variable are fixed in repeating sampling.
5. There is no correlation between the residual term given any value of the explanatory variable (i.e. $Cor(e_i , e_j / x_i , x_j) = 0$).
6. There should be no specification bias.
7. The number of sample must be greater than the number of parameters to be estimated.
8. There is no linear relationship between the residual and the explanatory variable (i.e. $Cor(e_i / x_i) = 0$).
9. There is no exact linear relationship between the explanatory variable (i.e. the explanatory variables are not collinear).
10. The residual term is normally distributed with mean of zero and variance σ^2 (i.e. $e_i \sim N(0, \sigma^2)$). Gujarati (2004).

HETEROSCEDASTICITY

If the error terms do not have constant variance, they are said to be heteroscedastic. The term heteroscedasticity which means 'deferring variance' and comes from the Greek 'hetero' (different) and 'skedasis' (dispersion). When the assumptions of homoscedasticity are violated and the variance is different for different observations, we have a problem of heteroscedasticity (Favour, 2018).

CAUSES OF HETEROSCEDASTICITY

According to Okeke (2016), the factors that can cause disturbances to be heteroscedastic are:

1. **Influence of the size of an explanatory variable in the size of the disturbance:** That is, observations expected to have error terms with large variances are given a smaller weight than observations thought to have error terms with small variances. The smaller the error variance, the more heavily the case is weighted. Intuitively, this makes sense; the observations with the smallest error variances should give the best information about the position of the true regression line.

2. **The presence of outliers can cause heteroscedasticity:** An outlier is an observation generated apparently by a different population to that generating the remaining sample observations. When the sample size is small, the inclusion or exclusion of such an observation can substantially alter the results of regression analysis and cause heteroscedasticity.

3. **Data transformation:** One of the solutions to solve the problem of multicollinearity consisted in transforming the model, taking ratios with respect to a variable (say x_{ij}), that is, dividing both sides of the model by x_{ij}. Therefore, the disturbance will now be u_i/x_{ij}, instead of u_i. Assuming that u_i fulfills the heteroscedastic assumption, the disturbances of the transformed model (u_i/x_{ij}) will no longer be homoscedastic but heteroscedastic.

DETECTION OF HETEROSCEDASTICITY

To detect heteroscedasticity, graphical or statistical test can be applied.

1. **Graphical test:** In this method, we are interested in the error term and its variation, so we plot a scatter plot that comes from a simple linear regression to detect possible deviations from homoscedasticity.

2. **Statistical test:** Three most common statistical test procedures to identify a problem of heteroscedasticity are the Goldfeld-Quandt test, the Breusch-Pagan test and the White test.
   (i) **The Goldfeld-Quandt test (GQ):** This test works under the assumption that the error term is homoscedastic. When this is true, the variance of one part of the must be the same as the variance of another part of the sample, independent on how the sample is sorted. If this is not the case we must conclude that the data is heteroscedastic.
   (ii) The Breusch-Pagan test (BP): This test is slightly more general than the Goldfeld-Quandt test, since it allows for more than one variable at a time to be tested. The starting point is a set of explanatory variables that we believe drives the size of the variance of the error term.
   (iii) White's test: This test is similar to the BP-test, but does not assume any prior knowledge of the heteroscedasticity, but instead examines whether the error variance is affected by any of the regressors, their squares r cross products. It is a large sample test, but does not depend on any normality assumption. Hence, this test is more robust than the other two test procedures described above.

REMEDYING HETEROSCEDASTICITY

Gabriel (2019) stated the following as ways to remedy heteroscedasticity

1. **Respecify the model:** This is based on the fact that sometimes heteroscedasticity results from improper model specification.
2. **Transform the dependent variable(s):** The use of transformation helps to stabilize the variance. Popular transformations of the dependent variable include the square root, log and reciprocal transformations. The log transform can be problematic if you want to exponentiate the results to present the results.
3. **The use of weighted least squares:** OLS in the presence of heteroscedasticity is LUE and not BLUE, Since it assigns equal weights (importance) to each observation. It is best to use an estimation method such that observations with greater variability are given less weight than those with smaller variability.

NON-NORMALITY OF ERROR

Various transformations are used to correct non-normally distributed data. Correlation, least squares regression, factor analysis, and related linear techniques are relatively robust against non-extreme deviations from normality provided errors are not severely asymmetric (Kasu, 2019). Severe asymmetry might arise due to strong outliers. Log-linear analysis, logistic regression, and related techniques using maximum likelihood estimation are even more robust against moderate departures from normality. Likewise, Monte Carlo simulations show that t-test is robust against moderate violations of normality (Boneau, 2019).

If the assumption that 'e' is distributed normally is called into question, we cannot use any of the t-test, F-test or R-square because these test are based on the assumption that 'e' is distributed normally. The result of these tests becomes meaningless.

To detect non-normality in errors, normal probability plot or Shapiro-Wilk test is used. In many cases of non-normal condition where the checks suggest that the data is normally distributed, there are options: Transform the dependent variable (reapeating the normality checks on the transformed data). Common transformations include taking the log or square root of the dependent variable (Stephen, 2001).

CAUSES FOR NON-NORMALITY

Basil (2017) stated that there are six reasons that are frequently to blame for non-normality.

(i) Extreme Values
(ii) Overlap of two or more processes
(iii) Insufficient data discrimination
(iv) Sorted data
(v) Values close to zero or a natural limit
(vi) Data follows a different distribution.

Extreme values: Too many extreme values in a data set will result in a skewed distribution. Normality of data can be achieved by cleaning the data. This involves determining measurement errors, data entry errors and outliers, and removing them from the data for valid reasons. It is important that outliers are identified as truly special causes before they are eliminated. Note: The nature of normally distributed data is that a small percentage of extreme values can be expected; not every outlier is caused by a special reason. Extreme values should only be explained and removed from the data if there are more of them than the expected under normal condition.

Overlap of two or more processes: Data may not be normally distributed because it actually comes from more than one process, operator or shift, or from a process that frequently shifts. If two or more data sets that would be normally distributed on their own are overlapped, data may look bimodal or multimodal. The remedial action for these situations is to determine which X's cause bimodal or multimodal distribution and then stratify the data. The data should be checked again for normality and afterward the stratified processes can be worked with separately.

Insufficient data discrimination: Round-off errors or measurement devices with poor resolution can make truly continuous and normally distributed data look discrete and not normal. We can overcome insufficient data discrimination by using more accurate measurement systems or by collecting more data.

Sorted data: Collected data might not be normally distributed if it represents simply a subset of the total output a process produced. This can happen if data is collected and analyzed after sorting.

Values close to zero or a natural limit: If a process has many values close to zero or a natural limit, the data distribution will skew to the right or left. In this case, a transformation such as the Box-Cox power transformation may help make data normal. In this method, all data is raised or transformed to a certain exponent indicated by lambda value. When comparing transformed data, everything under comparison must be transformed in the same way.

Data follow a different distribution: There are many data types that follow a non-normal distribution by nature. Examples include Weibull distribution (found with life data such as survival times of a product), Log-normal distribution (found with length data such as heights), Exponential distribution (found with growth data such as bacterial growth), Poisson distribution (found with rare events such as number of accidents), and Binomial distribution (found with proportion data such as percentage defectives). If data follow one of these different distributions, it must be dealt with using the same tools as with data that cannot be made normal.

METHODS: The data used in this research was simulated by the application of Monte Carlo approach. Generalized least square which is OLS on the transformed variables that satisfy the standard least squares assumptions shall be employed in the analysis. The estimators thus obtained are known as GLS estimators and it is these estimators that are BLUE.

The Weighted Least Squares estimator also known as the Generalized Least Squares estimator is given by: $\hat{\beta}_{WLS} = (X^{\iota}WX)^{-1} X^{\iota}WY$ where W is a diagonal matrix with diagonal elements w_{ii} . The diagonal elements of W matrix are set equal to $w_{ii} = \int_{1}^{\frac{1}{\hat{e}i}} \begin{matrix} if\ \hat{e}i \neq 0 \\ if\ \hat{e}i = 0 \end{matrix}$ where the $\hat{e}_i$ are residuals from an initial least squares fit to the data. The weight w_{ii} is applied to the observations and are intended to down-weight the extreme observations. Thus, the Weighted Least Squares estimated can be computed by applying Least Squares to the transformed observation $\sqrt{w_{ii}y_i}$ and $\sqrt{w_{ii}x_i}$.

The Weighted Ridge estimator can be computed using the formula: $\hat{\beta}_{WLS} = (X^{\iota}WX + KI)^{-1} X^{\iota}WY$, where is determined from the data using $K = \frac{PS_W^2}{\hat{\beta}_w^{\iota}\hat{\beta}_w}$.

$$S_W^2 = \frac{(Y - X\hat{\beta}_w)^{\iota}(Y - X\hat{\beta}_w)}{n-p}$$

and $\hat{\beta}_w$ denotes the coefficient estimates from the Weighted Least Squares estimators.

Samples of 275 independent variables were generated randomly, while the error term was generated to follow a gamma distribution with shape parameter 0.8 and scale parameter 0.25. These randomly generated values were used to estimate the values of the dependent variable using a multiple regression equation. Independent variables were also generated. We therefore compare the between estimators to ascertain the method which gives the best fit.

Table 1: The simulated data are as follows:

Serial number 1-40

Y	X1	X2	X3	Ei	Y	X1	X2	X3	Ei
10.25	1	2	3	2.50628	98.55	22	12	9	1.02849
21.00	3	4	6	2.88376	71.20	13	16	18	3.40459
31.75	5	6	9	0.11960	87.00	19	15	10	0.35767
42.50	7	8	12	0.87806	32.25	4	15	15	0.22429
53.25	9	10	15	0.31264	14.15	1	14	9	0.05469
64.00	11	12	18	2.08087	80.80	12	11	30	0.12005
74.75	13	14	21	0.09237	27.55	1	10	21	0.28793
85.50	15	16	24	1.37403	84.75	15	4	21	3.68984
96.25	17	18	27	2.17981	54.30	6	13	28	0.23479
107.00	19	20	30	1.07171	59.35	8	12	25	3.58392
113.05	27	8	3	0.38811	44.50	7	19	16	0.26534
90.15	15	19	29	0.52988	106.85	22	2	15	1.40919
62.80	11	18	18	0.33300	117.25	23	12	23	1.13184
83.85	19	15	7	0.25689	51.90	11	20	8	0.74122
75.05	12	3	23	1.89243	65.35	13	19	13	1.39283
52.35	6	7	25	7.01011	47.25	4	3	27	0.63790
51.00	10	15	10	0.67418	67.90	11	3	20	1.01339
134.55	26	17	29	1.76874	128.55	27	4	17	0.82274
73.10	17	13	4	1.95166	38.10	6	10	12	1.56090
110.45	24	3	11	0.28749	59.75	7	11	29	0.28022

Serial number 41-80

Y	X1	X2	X3	Ei	Y	X1	X2	X3	Ei
43.05	4	3	23	0.63209	50.55	9	13	13	1.50019
82.90	16	7	16	3.27700	40.45	6	14	15	0.32293
123.05	23	4	27	8.20782	62.30	11	10	16	0.00379
134.65	26	6	27	2.80154	127.55	27	9	17	0.26221
109.90	22	13	20	0.16629	74.70	15	7	12	1.34944
40.20	8	8	6	0.71459	125.15	23	4	29	0.82923
76.80	14	8	18	2.15934	101.70	20	14	20	0.02690
100.00	21	11	14	0.37947	84.75	20	20	5	1.59284
71.75	10	11	29	0.01921	49.05	8	11	15	0.09081
109.10	22	17	20	2.07468	95.55	19	9	17	0.27295
59.75	11	7	13	0.01226	99.65	22	17	11	1.11964
25.95	3	16	13	0.22687	106.60	24	17	10	1.37293
59.25	12	19	11	2.00247	79.30	13	7	24	0.75820
69.10	10	19	28	0.55399	49.40	5	7	26	0.92580
73.00	15	5	10	1.96895	112.65	22	15	23	0.18415
32.70	2	20	24	1.06113	55.40	12	12	6	3.28219
125.60	26	4	18	0.14929	107.70	19	6	28	1.71087
109.80	22	3	18	1.40400	119.00	26	16	14	1.49706
74.25	11	8	27	0.18412	102.55	21	14	17	0.69957
91.95	15	10	29	0.08348	128.55	24	7	29	1.63697

Serial number 81-120

Y	X1	X2	X3	Ei	Y	X1	X2	X3	Ei
57.70	10	13	16	2.77464	34.25	3	6	19	1.38977
121.90	25	13	20	0.16875	88.20	20	8	6	1.99884
66.55	13	13	13	1.37635	70.45	12	5	19	0.13251
110.25	23	5	15	0.06192	64.05	15	13	3	0.80199
108.20	19	14	30	1.15127	107.30	20	7	24	1.55347
30.50	5	7	8	1.11631	83.35	15	11	21	0.18558
38.30	8	7	4	2.21850	94.10	19	11	16	0.09397
119.50	23	6	24	7.85650	47.75	9	6	9	1.21517
16.50	3	16	4	2.23800	64.90	15	14	4	0.40587
58.10	7	14	28	1.92493	52.80	9	7	14	0.25252
126.10	25	13	24	2.40304	120.55	24	5	21	2.17714
54.85	7	4	23	1.16630	106.20	23	20	14	0.26343
40.90	7	16	12	6.24027	57.60	7	6	26	0.44963
130.20	24	4	30	0.07772	64.25	11	16	19	0.94960
81.40	18	2	6	4.75151	29.15	5	19	9	2.47589
103.40	21	15	18	5.21733	83.35	15	11	21	1.48732
121.25	25	11	19	1.29970	66.30	9	13	28	0.10306
106.45	20	6	23	2.09390	135.10	26	9	28	1.46494
59.40	7	18	30	0.04543	35.10	3	7	20	2.31370
78.80	16	17	14	0.13114	120.25	25	16	19	0.34795
57.70	10	13	16	2.77464	34.25	3	6	19	1.38977

Serial number 121-160

Y	X1	X2	X3	Ei	Y	X1	X2	X3	Ei
110.10	20	14	28	3.39827	66.10	12	11	16	0.22684
58.40	11	19	14	0.34740	13.70	1	11	8	2.31812
70.00	12	2	18	0.11493	110.95	20	15	29	3.38007
57.80	13	20	6	1.12052	56.05	8	18	23	1.13853
57.75	10	18	17	2.82058	31.95	6	4	5	0.77180
93.90	22	9	4	5.26920	89.45	18	9	15	3.75368
14.35	2	12	5	0.79718	31.20	1	18	26	0.68307
24.80	3	6	10	1.24019	109.75	21	20	25	1.60452
83.30	14	7	24	0.76378	66.35	9	18	29	6.91675
110.95	22	13	21	0.02191	92.90	18	18	20	0.16444
22.40	1	20	18	0.01244	127.90	26	3	20	4.24307
19.90	4	19	4	0.31726	91.65	22	15	3	3.29506
129.90	27	13	20	6.38635	126.55	27	14	17	0.05641
55.20	6	19	30	2.39306	75.60	14	14	18	1.62687
98.60	19	20	22	0.57244	63.10	10	7	20	0.25775
60.05	12	15	11	0.49801	115.85	27	15	7	0.59819
45.40	6	5	18	1.04896	72.00	12	13	22	2.74890
85.40	16	5	18	0.03327	18.25	2	3	7	2.12152
113.05	21	14	27	0.21923	54.30	11	8	8	0.82975
90.85	22	19	3	1.19865	78.45	16	3	11	0.95000

Serial number 161-200

Y	X1	X2	X3	Ei	Y	X1	X2	X3	Ei
83.75	13	11	29	0.14973	69.55	13	19	17	1.57796
54.15	11	14	9	0.41445	98.00	23	19	6	0.71196
114.40	21	2	26	2.54748	126.80	25	20	26	0.04712
98.15	21	15	13	1.46984	64.10	9	3	24	1.77117
99.45	24	16	3	8.13330	99.20	20	16	18	0.74909
67.75	11	9	21	1.58066	94.60	22	16	6	2.02920
82.60	18	17	10	1.62188	37.30	3	17	24	0.03099
117.40	25	4	14	2.01538	81.10	18	14	8	1.40296
131.60	25	17	30	0.19482	50.15	9	15	13	0.79900
91.90	22	19	4	0.23131	123.20	26	16	18	0.00287
115.60	25	13	14	2.51283	81.60	15	4	18	0.81333
130.15	27	17	21	0.70115	94.10	20	10	12	0.62821
94.55	21	12	9	0.10858	36.60	6	7	10	7.96025
74.25	14	5	15	2.85285	47.75	7	8	17	0.74890
60.35	13	2	5	4.46655	74.00	13	2	18	2.39371
82.30	13	13	28	0.35661	26.15	2	16	17	0.33121
46.35	8	14	13	2.95833	120.65	24	15	23	0.12636
95.35	22	7	5	0.05484	28.70	1	20	24	0.82037
26.10	1	12	20	0.58960	90.57	15	17	29	1.47171
103.65	21	19	19	0.73882	48.30	11	17	4	2.35948

Serial number 201-240

Y	X1	X2	X3	Ei	Y	X1	X2	X3	Ei
20.05	3	14	7	0.8112	82.75	14	15	25	0.40359
104.15	24	3	5	0.2344	41.05	8	9	7	0.63857
39.75	5	8	17	2.4618	98.05	18	8	23	0.87972
29.30	2	16	20	1.3502	62.45	10	5	19	0.04340
79.95	18	4	5	0.9602	109.20	22	6	18	0.74020
73.85	15	6	11	0.6289	49.85	10	5	7	0.09739
58.70	10	8	16	0.0664	83.30	19	2	4	1.94311
65.65	14	6	7	0.7237	39.45	7	18	11	3.96616
116.95	24	2	17	2.1489	44.25	8	14	11	0.18499
118.25	27	3	7	1.4307	53.25	6	13	27	4.31482
48.80	5	10	26	0.0036	87.20	19	14	10	0.21664
61.55	12	18	13	2.7092	98.60	22	17	10	1.09423
110.55	20	17	29	0.5958	127.75	26	9	21	1.98270
52.35	6	7	25	0.4093	51.00	11	14	6	4.21174
93.85	17	9	23	12.4333	63.90	8	5	28	3.14475
93.50	19	14	16	0.6099	115.75	22	10	25	0.14960
95.35	20	9	13	0.7145	49.20	10	3	6	6.31578
25.35	3	19	13	0.2802	129.40	27	5	18	0.15171
101.40	21	4	14	3.1068	92.65	17	15	23	1.28279
71.95	15	5	9	6.2428	104.45	21	15	19	3.13627

Serial number 241-275

Y	X1	X2	X3	Ei	Y	X1	X2	X3	Ei
112.35	23	5	17	1.57536	126.80	27	18	18	1.03382
20.25	3	13	7	0.55512	64.50	8	2	28	0.29750
99.65	23	16	7	0.50076	59.10	11	5	12	0.83986
38.25	8	2	3	0.33582	38.80	8	15	6	6.41522
27.50	4	2	8	1.06751	40.95	3	4	25	1.99674
59.65	14	15	3	7.54888	30.10	4	10	12	0.47237
95.75	21	6	9	0.00915	61.80	14	20	6	0.23017
81.20	14	7	22	0.11606	72.70	15	17	12	0.27169
88.75	15	5	25	1.51271	24.65	2	13	15	1.63617
35.90	3	3	20	0.09267	58.50	12	7	8	2.23573
47.05	10	19	7	1.50903	81.10	15	17	20	0.62603
82.65	14	5	23	0.50950	33.55	4	19	17	0.22751
24.80	4	5	6	0.91800	58.05	11	5	11	0.00134
37.85	7	5	7	0.01279	40.75	3	5	25	3.32490
43.70	3	6	28	0.61706	91.20	17	17	22	6.21695
54.75	8	14	21	1.62353					
24.90	5	14	4	0.24255					
87.40	15	17	26	2.87319					
43.20	9	13	6	0.03004					
92.95	18	2	17	0.04229					

Figure 1:

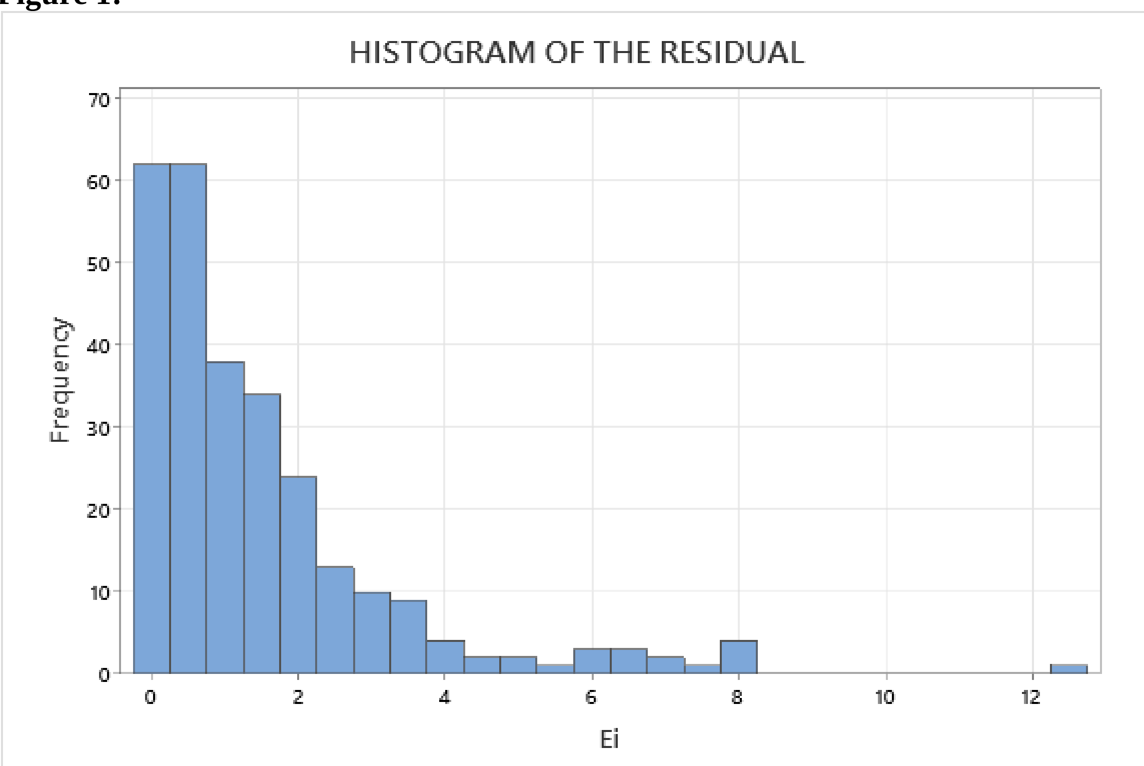
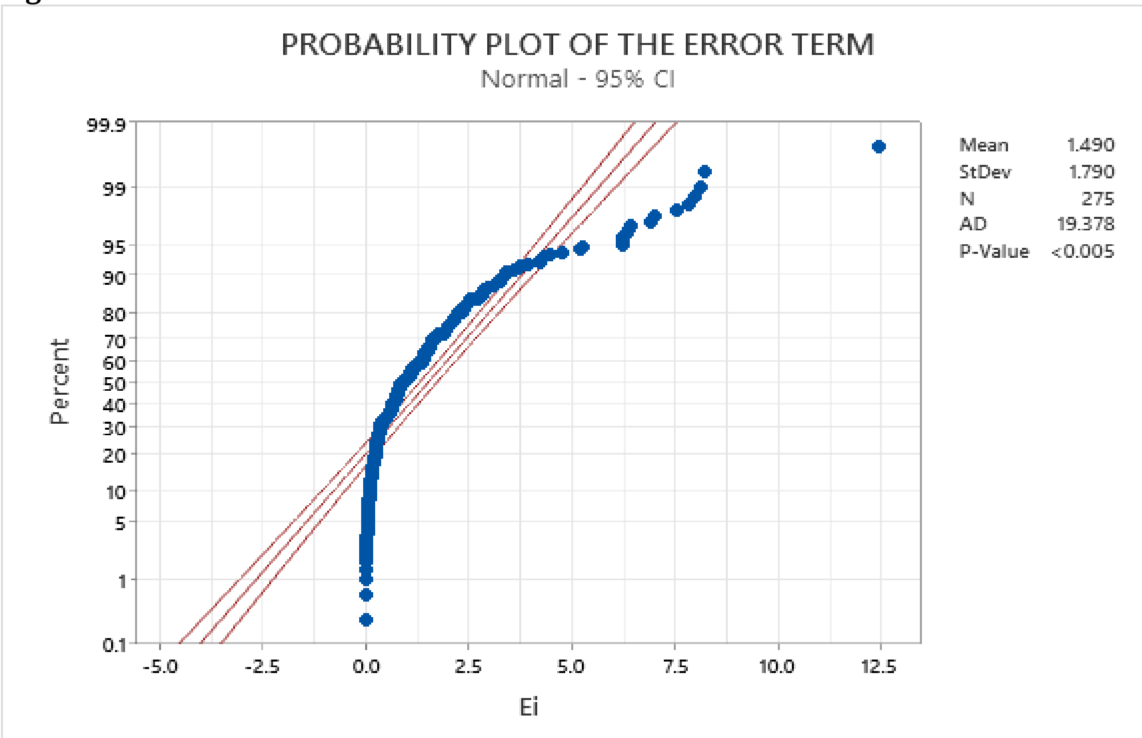


Figure 2:



The above plot revealed that the given did not obey normality assumption. The fitted model parameters are as follows: $\boldsymbol{\beta_0}$ **= 3.5,** $\boldsymbol{\beta_1}$ **= 4.0,** $\boldsymbol{\beta_2}$ **= -0.2 and** $\boldsymbol{\beta_3}$ **= 1.05.**

The summary of the results obtained from the estimators suggests that the Ordinary Least Squares and Weighted Ridge Regression are appropriate in the presence of a non-normal error. However, the Weighted Least Squares is least affected by the distribution of the error term as it pose to be the most efficient estimator (estimator with the least error). This is confirmed by the residual standard error, residual mean absolute deviation and the root mean square error as can be seen below:

Table 2: Summary of Efficiency Test among various Estimators

Estimator	RSE	RMAD	RMSE
Ordinary Least Squares	1.684	2.7978	0.5403
Weighted Least Squares	0.4788	2.7519	0.1167
Weighted Ridge Regression	1.3424	2.8960	0.9853

Having obtained this result, the residual of the weighted least squares estimate was subjected to further error diagnostics. The parameter estimate for the weighted least squares, with the inverse of the response variable as the weight is shown in table 4 above.

The normal distribution test, auto-correlation test, homoscedasticity test and the multi-colinearity test were adopted to identify if there are further violations of the classical linear normal regression model which may have also affected the result.

Jarque Bera test for normality, adopted to evaluate if the error term follows a normal distribution strongly rejects the null hypothesis of normality. This confirms that the error term is non-normal and is suitable for the argument of this paper.

Durbin-Watson test for the presence of auto-correlation do not reject the null hypothesis of no serial correlation, thus indicating strong evidence that there is no serial correlation between the error terms.

Breusch-Pagan test of homoscedasticity also rejected the null hypothesis of homoscedasticity.

Lastly, the test of multicollinearity among the predictor variables using 1% two tail significance level showed that there is no multicollinearity (no indication of perfect correlation between the independent variables).

Table 4. Summary of Diagnostic Test of the Residuals for Weighted Least Squares

Test	Null Hypothesis	Method	Test Statistic	DF	P-Value
Normality	Data is normal	Jarque-Bera	347.95	2	0.000
Auto-correlation	No serial correlation	Durbin Watson	1.9509		0.3402
Homoscedasticity	Constant variance	Breusch-Pagan	3,1301	3	0.372
Multi-collinearity	No perfect correlation	VIF	0.0441		0.000

Summary: Based on the analytical result, the weighted least squares estimator gave the most efficient estimate. All estimators however were unbiased and had their estimates very close to the actual estimate. The residual of weighted least square estimator was further diagnosed and the data was checked for other factors that may have affected the efficiency of other estimators. The diagnostic test further showed that apart from the non-normality assumption (which is our focal point of interest), no other assumption of the classical linear normal regression model was violated.

Conclusion: Based on the analytical result, the following conclusions were reached:

1. In the presence of both heteroscedasticity and non-normal error, the Ordinary Least Squares estimators produces relatively large mean square errors, hence it gives unstable parameter estimates.

2. This study reveal that the consequencies of heteroscedasticity in the presence of non-normal error is more severe compared to when the error distribution is normal.
3. The results of comparisons show that for the condition of heteroscedasticity, weighted least squares estimates are more efficient and precise than the other estimators considered. This also applies when both heteroscedasticity and non-normal error are considered.

Recommendation: We therefore recommend that Weighted Least Squares method of estimation having produced estimates that are more efficient and precise than the other estimators considered in this study (has the best linear and unbiased estimates) should be considered first while dealing with heteroscedasticity and non-normal errors being the common problems in multiple regression models.

REFERENCES:

Asukwo, K. (2019). Econometric Analysis, 5th ed., Pearson Education, Inc.

Basil, G. (2017). Introduction to Linear Regression Analysis. Wiley, New York.

Boneau, D. (2019). Regression Diagnostics: ''Identifying Infuencial Data and Sources of Collinearity''. John Wiley and Sons., New York.

Chikezie, A. (2014). Econometrics Theory and Problems, 2nd Edition, Tony Prints Services Aba.

Favour, K. (2018). ''*Family Likeness in Stature*''. Proceedings of Royal Society, London, Vol. 4, pp, 32-44.

Gabriel, B.(2019). ''*A comparison of Several Regression Models for Analysing Cost of CABG Surgery*, STATISTICS IN MEDICINE; 12:81- 93.

Gujarati, D (2004). ''Econometric Analysis, 4th Edition'' Tata McGraw-Hill, New Delhi.

Kasu, B. (2019). ''*Weighted Ridge Regression: Combining Ridge and Robust Regression Methods''*. NBER Working Paper Series, Volume 9, 11-23.

Okeke, N. (2016). ''*Ridge Regression: Application to Non-orthogonal Probems''*. Technometrics, 12, 72-84.

Okenna, A. (2020). "*On the Laws of Inheritance,''*. Biometrika, Vol. 2, Nov, 2020, Pp. 357-462.

Richard, W. (2018). ''Heteroscedasticity''. University of Notre Dame, New York. https//www.nd.edu/rwilliam.

Stephen, P. (2001). Statistics and Econometrics. 2nd Edition. Enugu: John-J Press Inc.

Welsch, D. (2018). Econometric Theory and Problems, 2nd ed. USA: McGraw-Hill Inc.