Performances of Ordinary and Generalized Least Squares Estimators on Multiple Linear Regression Models with Heteroscedasticity

Adebowale Olusola Adejumo[1*]; Obalowu Job[1]; Titilayo Dorcas Isaac[2]; and Benjamin Agboola Oyejola[1]

[1]Department of Statistics, University of Ilorin, Ilorin, Nigeria.
[2]Department of Mathematics, Federal University Lafia, Nassarawa, Nigeria.

. E-mail: aodejumo@unilorin.edu.ng*

ABSTRACT

This paper focuses on the impact of heteroscedascity on the estimate and variance of model parameters by studying the performances of Ordinary Least Square (OLS) and General Least Square (GLS) estimators in multiple linear regression models with two independent variables only and error term characterized with different magnitudes of heteroscedasticity related to predictors at different sample sizes. This research explored the patterns of the variance estimates offered by the two estimators when data is front with heteroscedascity. Also the studies explored the situations where concepts of substitute, complementary or joint demands/supply may reflect in the stochastic characterization of the error term. From Monte Carlo simulation studies using R-package the GLS estimator maintains its superiority over the OLS in multiple linear regression models.

(Keywords: heteroscedasticity, ordinary least squares, generalized least squares, stochastic error term, magnitude, characterization, Monte Carlo, simulation)

INTRODUCTION

Data may be available for economic research though it is often plagued with some problems. The validity of any analysis ultimately depends on the availability of appropriate data. Just as time series data do encounter non-stationary problems, cross-sectional data too may portray especially the problems of heteroscedasticity, autocorrelation and/or, multicollinearity. Thus in practice, regression analysis depends on the availability of appropriate data. Therefore it is very germane that researchers state clearly the sources of the data used in the analysis, definitions, methods of collection, and any gaps or omissions in the data as well as any transformation in the data, bearing in mind that the macroeconomic data published by the government are often aggregates.

Regression analysis is a statistical tool used for testing hypothesis about the linear relationship between a dependent variable Y and explanatory variable (X) and for valid prediction. Regression analysis is concerned with the study of the dependence of one variable (dependent variable) on one or more of the explanatory variables, with a view to estimating and/or predicting the population mean(average value)of the former in terms of the known or fixed values of the latter.

In linear regression models, the disturbance term ε_i is a surrogate for all those variables that are omitted from the model occasioned by misspecification but collectively affect values of Y. Error term is often included in catchall models as a substitute for all the excluded variables from the model. The disturbance term ε_i is assumed normal with mean zero,

$$E(\varepsilon_i) = 0 \text{ and } Var(\varepsilon_i) = \sigma^2,$$

which describes the homoscedasticity assumption.

When the assumption of homoscedasticity is being violated, i.e.. $Var(\varepsilon_i) \neq \sigma^2$ for every i, heteroscedasticity sets in. Heteroscedasticity is likely to be more predominant in cross sectional data than in time series data.

Cross-sectional data can be collected on family expenditure, income size, demand, production, supply, etc. by randomly drawing sample size (n) from the population (N).

According to Berry and Feldman (1985), slight heteroscedasticity has little effect on significances test, however when heteroscedasticity is marked, it can lead to serious distortion of findings and seriously weaken the analysis, thus increasing the probability of type I error.

Heteroscedasticity does not destroy the unbiasedness and consistency properties of the OLS estimator except that they are no longer efficient. This lack of efficiency makes the usual hypothesis testing procedure dubious, since the estimated standard error is either under or over-estimated in either case resulting in incorrect inferences. And if we persist in using the usual testing procedures despite heteroscedasticity, whatever conclusions we draw or inferences we make, may be misleading therefore remedial measures become necessary.

We refer to a Monte Carlo study conducted by Davidson and Mackinnon (1993). They considered the simple linear model that follows:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \qquad \textbf{(1)}$$

They assume that $\beta_1 = \beta_2 = 1$ and $\varepsilon \sim N(0, X_i^h)$. As the last expression shows, they assume that the error variance is heteroscedastic and is related to the value of the regressor X with power h.

If for example h=1, error variance is proportional to the value of X: if h=2, error variance is proportional to the square of the values of X, and so on. Based on 20,000 replications and allowing for various values of h, they obtain the standard errors of the two regression coefficients using OLS, OLS allowing for heteroscedasticity and GLS estimators.

The most striking feature of these results is that OLS with or without heteroscedasticity, consistently overestimates the true standard error obtained by GLS procedure, especially for large values of h, thus establishing the superiority of GLS estimator. These results also showed that if we do not use GLS and rely on OLS-allowing for or not allowing for heteroscedasticity- the picture is mixed. The usual OLS standard errors are either too large for the intercept or too small for the slope coefficient in relation to those obtained by OLS allowing for heteroscedasticity.

In furtherance of the Monte Carlo experiment conducted by Davidson and Mackinnon (1993) this research attempts to study multiple linear regression models with two independent variables only where the error term is assumed to be related to each independent variable separately and jointly up to a multiplicative constant.

This research examined the performances of OLS and GLS estimators using a multiple linear regression model front with heteroscedasticity of different magnitudes inherent in the error term in relation to the predictor(s).

METHODOLOGY

Heteroscedasticity is potentially a serious problem and researchers need to ascertain its presence in regression models. If its presence is detected, then corrective action becomes imperative. Methods of detecting heteroscedasticity in bivariate data set are many, but for this study we employed Goldfield and Quandt test to detect the likely presence of heteroscedasticity in the regression model. It is a formal method of detecting heteroscedasticity.

Monte Carlo Simulation Study

Consider the general form of heteroscedasticity with the error structure:

$Var(\varepsilon_i) = \sigma^2 X_{ij}^h$,where $h \neq 0$

Generally speaking homoscedasticity situation can be made heteroscedastic by changing the error structure in relation to one of the regressors.

Step 1: Generate X_{ij} with mean μ_x and variance σ_x^2 using sample size (n) of observations.

Step 2: Formulate the error structure by specifying the magnitude of heteroscedasticity, such that h=0.5, 1, 1.5, 2, 2.5, 3

Step 3: Fix Values for β_j, $\beta_0 = \beta_1 = \beta_2 = 1$

Step4: Generate the values of the dependent variable (Y).

$$Y_i = 1 + X_{1i} + X_{2i} + \varepsilon_i \qquad \textbf{(2)}$$

Step 5: Carry out the OLS regression of the dependent variable(Y) on the predictor(X).

Step 6: Test for the presence of heteroscedasticity using Goldfield –Quandt test.

Step 7: If heteroscedasticity is likely in Step 6, go to Step 8 otherwise examine Steps 1 to 5

Step 8: X-ray the performance of OLS and GLS estimators in the presence of heteroscedasticity.

Ordinary Least Square Estimators (OLS)

The classical ordinary least square method (OLS) is widely used in regression analysis. The estimator minimizes the sum of squares error .It becomes unsuitable to apply OLS on regression models when predictor variables are highly correlated or heteroscedasticity is being detected.

For simple linear regression model (SLRM):

$$y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \,,$$

OLS estimator remains unbiased. Heteroscedasticity does not destroy the unbiased-ness property of the OLS estimator, but it becomes less efficient. Hence the variance of the model parameters is no longer the minimum variance in the class of linear estimators.

Recall that for homoscedasticity situation, when h=0:

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}, \qquad x_i = X_i - \bar{X} \qquad \textbf{(3)}$$

And for heteroscedasticity, when h=1

$$Var(\hat{\beta}_2^*) = \frac{\sum X_i^2 \sigma^2}{\sum x_i^2}, \qquad x_i = X_i - \bar{X} \qquad \textbf{(4)}$$

Hence $Var(\hat{\beta}_2^*) < \mathrm{var}(\hat{\beta}_2)$

Generalized Least Square Estimator (GLS)

The generalized least square estimator (GLS) make use of information available in the data and assign different weight to variables according to their sizes such that observation coming from population with greater variability are given less weight than population with smaller variability. To illustrate this supposed:

$$y_i = \beta_0 x_{0i} + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \qquad \textbf{(5)}$$

Where $x_{0i} = 1$ for all i=1,2,.., n

Assuming that the heteroscedasticity variance term σ_i^2 is known up to a multiplicative constant. We shall transform the model equation by dividing through by σ_i to make it homoscedastic:

$$\frac{y_i}{\sigma_i} = \beta_0\left(\frac{x_{0i}}{\sigma_i}\right) + \beta_1\left(\frac{x_{1i}}{\sigma_i}\right) + \beta_2\left(\frac{X_{2i}}{\sigma_i}\right) + \left(\frac{\varepsilon_i}{\sigma_i}\right) \qquad \textbf{(6)}$$

Which for ease of exposition we write as:

$$y_i^* = \beta^*_0 \, x_{0i}^* + \beta^*_1 \, X^*_{1i} + \beta_2^* X_{2i}^* + \varepsilon^*_i$$

It is remarkable to note that the transformed error term in the model becomes homoscedastic.

$$\varepsilon_i^* = \frac{\varepsilon_i}{\sigma_i}$$

$$\text{Var}(\varepsilon^*_i) = \text{E}(\varepsilon_i^{*2}) = \frac{1}{\sigma_i} E(\varepsilon_i^2) \qquad \textbf{(7)}$$

$$= \frac{1}{\sigma_i^2} E(\sigma_i^2) = 1$$

This showed that the variance of the transformed disturbance term ε^*_i is now homoscedastic. Since we are still retaining all other assumption of the classical model, the finding that ε^*_i is homoscedastic suggests that we can now apply OLS to the transformed data which will produce estimates with minimum variance. This affirms that GLS is OLS transformed (see McCullach and Nelder, 1983).

For SLRM, recall that when homoscedasticity assumption holds the OLS minimizes:

$$\sum \varepsilon_i^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \quad \textbf{(8)}$$

But GLS minimizes the unequally weighted residuals, the expression written as:

$$\sum (\varepsilon^{*2}_i) = \sum w_i \varepsilon_i^2 = \sum w_i (y_i - \hat{\beta}*_0 \, x_{0i} - \hat{\beta}*_1 \, x_i)^2 \quad \textbf{(9)}$$

Where $w_i = {1}/{\sigma_i^2}$

Thus GLS minimizes a weighted residual sum of square with $w_i = {1}/{\sigma_i^2}$ acting as the weight.

But OLS minimizes an un-weighted or equally weighted residual sum of square (RSS). This shows that the GLS technique assigns weights to each observation which is inversely proportional to it variance σ_i^2 (Damodar, 2011).

Goldfield and Quandt Test

This test is applicable to a large sample since the number of observations must be at least twice as the number of parameters to be estimated. The test assumes normality and serially independent stochastic error terms.

If we assume that the heteroscedasticity variance σ_i^2 is positively related to one of the explanatory variables in the regression model. Then fit model like in Equation (1), $Y_i = \beta_1 + \beta_2 X_i + \varepsilon$. Goldfield and Quandt procedures follow:

(i) Rank the observation according to the magnitude of explanatory variables X_i, suppose two or more variables are included in the model, a variable X_i which is closely related to Y is chosen for ranking the data. If apriori we are not sure which X variable is appropriate, we can conduct the park test on each of those X variables.

(ii) Select a certain number of central observation, call it "c" which we shall omit from the analysis according to Mont Carlo experiment conducted by Goldfield and Quandt, they suggested the number of omitted observation "c" when n=30, c=8and when n=60, c=16. According to Judge et al the number of omitted observations when n=30, c=4 and when n=60, c=10, here c is specified a priori and the remaining (n-c) observations are divided into two sub-samples of equal number of observations $\left(\frac{n-c}{2}\right)$, the rationale behind this is to "sharpen the difference between the small variance group and the large variance group".

(iii) Fit separate OLS regression to each sub-sample value of x_i. The first $\left(\frac{n-c}{2}\right)$ and the last $\left(\frac{n-c}{2}\right)$ observations.

The idea is to obtain the respective residual sum of squares RSS_1 and RSS_2 small and large variance groups, respectively. Each RSS has $\left(\frac{1}{2}(n-c)-k\right)$ degree of freedom, where k is the number of parameters to be estimated including the intercept.

(iv) Compute the ratio ,

$$F^* = \frac{(\sum \varepsilon_1^2)}{(\sum \varepsilon_1)^2} = \frac{RSS_2}{Rss_1} \sim F_{r_1, r_2} \quad \textbf{(10)}$$

Where $v_1 = v_2 = \left(\frac{1}{2}(n-c)-k\right)$

If $F^* \to 1 or \langle 1$ this implies homoscedasticity otherwise if $F^* \rangle 1 or \rangle F_{v1,v2,\alpha}$

(v) Reject $H_0 i.e.$ heteroscedasticity is likely present (Damodar, 2011).

ANALYSIS

In this study we fit a multiple linear regression model with heteroscedastic error term related to the regressor(s) in order to examine the impact of heteroscedasticity on model parameters. For instance in demand and supply studies demand for certain commodities are competitive, complementary, derived or joint in nature.

This study explores the demand and supply scenario above. It is feasible that heteroscedasticity is being related to one of the regressors or joint regressors in multiple linear regression model (MLRM) as it may be typical of joint demand/supply in economic situations.

We conducted a test for homoscedasticity assumption employing Goldfield and Quandt test to affirm that the simulated data is plagued with heteroscedasticity. Based on 1000 replication the we run the OLS and GLS regression to obtain the standard error estimates of model parameters for different value of h =0.5,1,1.5,2, 2.5, 3, and sample sizes(n)= 20, 25,30,35,45 and 50. The results for test of heteroscedasticity are summarized in Table 1. From Table 1 we observed that heteroscedasticity is likely present as depicted by p-values. The performances of OLS and GLS estimators *vis-a-vis* the variance estimates of model parameters when $var(\varepsilon_i) = (X_{1i})^h$ are summarized in Table 2.

Table 1: Goldfield and Quandt Test When $\varepsilon_i \sim (0, X_{1i}{}^h)$.

N	G	df1	df2	GQ	p-value
25	**0.5**	10	9	6.7499	0.004172
	1	12	12	30.5114	4.11E-07
	1.5	15	14	30.5209	4.18E-08
	2	17	17	3.9128	0.003732
	2.5	20	19	7.0835	3.81E-05
	3	22	22	2.8159	0.009373
30	**0.5**	10	9	10.5647	0.0007756
	1	12	12	15.4367	1.79E-05
	1.5	15	14	5.0454	0.002145
	2	17	17	4.4055	0.00192
	2.5	20	19	3.492	0.0043
	3	22	22	2.9063	0.007776
35	**0.5**	10	9	16.2331	0.0001391
	1	12	12	10.9238	0.0001108
	1.5	15	14	3.2371	0.017
	2	17	17	2.9202	0.01664
	2.5	20	19	3.532	0.004024
	3	22	22	2.2334	0.03294
40	**0.5**	10	9	6.3109	0.005306
	1	12	12	3.4742	0.02013
	1.5	15	14	4.4504	0.003993
	2	17	17	6.0936	0.0002678
	2.5	20	19	6.4637	7.46E-05
	3	22	22	2.7414	0.01095
45	**0.5**	10	9	14.5836	0.0002152
	1	12	12	5.7849	0.002421
	1.5	15	14	2.8091	0.03037
	2	17	17	10.2089	8.04E-06
	2.5	20	19	3.0641	0.008971
	3	22	22	2.5227	0.01745
50	**0.5**	10	9	3.8591	0.02715
	1	12	12	4.4113	0.007829
	1.5	15	14	3.8566	0.007846
	2	17	17	2.4905	0.03412
	2.5	20	19	9.4728	4.08E-06
	3	22	22	2.2215	0.03383

Table 2: Standard Error of Estimates of Model Parameter When $\varepsilon_i \sim (0, X_{1i}{}^h)$.

Sample sizes	G	$s.e(\beta_0)_{OLS}$	$s.e(\beta_0)_{GLS}$	$s.e(\beta_1)_{OLS}$	$s.e(\beta_1)_{GLS}$	$s.e(\beta_2)_{OLS}$	$s.e(\beta_2)_{GLS}$
25	0.5	0.030406	0.0148699	0.001484	0.0007496	0.001772	0.0008492
	1	0.10979	0.066337	0.003481	0.001798	0.006474	0.00385
	1.5	0.086988	0.042256	0.003065	0.001498	0.005085	0.0022
	2	0.07273	0.045202	0.002151	0.001444	0.004076	0.002142
	2.5	0.05903	0.0278775	0.00313	0.0009316	0.00454	0.0020133
	3	0.04627	0.0271126	0.001488	0.0008421	0.002937	0.0014895
30	0.5	0.077036	0.046213	0.003527	0.002269	0.004254	0.002231
	1	0.19475	0.092266	0.00725	0.004262	0.01251	0.007668
	1.5	0.214511	0.062338	0.008126	0.003121	0.012979	0.003881
	2	0.083442	0.03302	0.003855	0.001242	0.004571	0.001788
	2.5	0.084158	0.033341	0.004046	0.001858	0.005479	0.001954
	3	0.11112	0.053401	0.004515	0.002334	0.006279	0.002726
35	0.5	0.133246	0.054214	0.006803	0.003002	0.008296	0.003661
	1	2.2321	1.13918	0.0953	0.05392	0.1318	0.05436
	1.5	1.5	0.06152	0.099313	0.003003	0.004924	0.003686
	2	0.155653	0.103455	0.005486	0.002696	0.011637	0.006535
	2.5	0.221403	0.083344	0.007656	0.002754	0.012566	0.003964
	3	0.172448	0.073908	0.006044	0.002955	0.009359	0.004164
40	0.5	17.4122	8.725	0.5624	0.2595	0.9858	0.5029
	1	1.29E-14	1.29E-14	4.90E-16	2.23E-16	7.72E-16	7.71E-16
	1.5	1.5	1.55E-14	4.40E-15	5.04E-16	1.63E-16	9.15E-16
	2	1.85E-15	6.31E-15	7.68E-17	3.02E-16	1.20E-16	3.85E-16
	2.5	2.18E-15	4.42E-15	9.35E-17	1.35E-16	1.32E-16	2.77E-16
	3	2.09E-14	4.37E-15	6.65E-16	7.78E-17	1.35E-15	3.09E-16
45	0.5	0.93672	0.69899	0.03588	0.02951	0.0623	0.03713
	1	0.40086	0.191453	0.01908	0.008759	0.02378	0.010586
	1.5	0.99617	0.40581	0.03165	0.01808	0.06515	0.02293
	2	0.55086	0.15975	0.03021	0.01393	0.04135	0.01436
	2.5	0.29076	0.170527	0.01563	0.00986	0.0146	0.008552
	3	0.68474	0.38341	0.02571	0.01293	0.0417	0.02218
50	0.5	3.2294	2.33964	0.1318	0.09597	0.1773	0.11928
	1	5.0376	3.43155	0.1334	0.08453	0.3117	0.20617
	1.5	3.3037	1.93188	0.1133	0.05323	0.2054	0.09755
	2	2.8752	1.59767	0.114	0.08293	0.1584	0.07835
	2.5	1.72675	1.11184	0.07961	0.05065	0.10172	0.05539
	3	2.5526	1.63834	0.1028	0.06144	0.144	0.08144

When $var(\varepsilon_i) = (X_{1i})^h$, the following observations are feasible from Table 2: the $s.e(\beta_1)_{GLS}$ is the least, followed by $s.e(\beta_2)_{GLS}$. Next is the $s.e(\beta_1)_{OLS}$ then $s.e(\beta_2)_{OLS}$. The $s.e(\beta_0)_{GLS}$ is smaller than $s.e(\beta_0)_{OLS}$.

Also from the simulation studies we observed that:

$s.e(\beta_0)_{GLS} < s.e(\beta_0)_{OLS}$ 34 out of 36 times (94.4%)

$s.e(\beta_1)_{OLS} < s.e(\beta_2)_{OLS}$ 33 out of 36 times (91.7%)

$s.e(\beta_1)_{GLS} < s.e(\beta_2)_{GLS}$ 33 out of 36 times (91.7%)

$s.e(\beta_2)_{GLS} < s.e(\beta_1)_{OLS}$ 29 out of 36 times (80.6%)

$s.e(\beta_1)_{GLS} < s.e(\beta_0)_{OLS}$ 36 out of 36 times (100%)

For the case of when error term is contaminated with heteroscedasticity of the form $\varepsilon_i \sim (0, (X_1 * X_2)^h)$ which is up to a joint-multiplicative constant in the model $y_i = 1 + x_{1i} + x_{2i} + \varepsilon_i$. Data simulated was examined for heteroscedasticity using Goldfield-Quandt test. The results are summarized in the Table 3. Since the p-values are strictly less than α=0.05. It is likely that heteroscedasticity is present. Again the performances of the OLS and GLS estimators when the error term is related to the product of values of the two regressors up to a multiplicative constant in the model, $y_i = 1 + x_{1i} + x_{2i} + \varepsilon_i$ are summarized in Table 4.

When $\text{var}(\varepsilon_i) = (X_{1i} * X_{2i})^h$ the following observations are feasible from Table 4: The $s.e(\beta_1)_{GLS}$ is the least, followed by $s.e(\beta_2)_{GLS}$. Next is the $s.e(\beta_1)_{OLS}$ then $s.e(\beta_2)_{OLS}$. The $s.e(\beta_0)_{GLS}$ is smaller than $s.e(\beta_0)_{OLS}$.

Also from the simulation studies we observed that:

$s.e(\beta_0)_{GLS} < s.e(\beta_0)_{OLS}$ 36 out of 36 times (100%)

$s.e(\beta_1)_{OLS} < s.e(\beta_2)_{OLS}$ 36 out of 36 times (100%)

$s.e(\beta_1)_{GLS} < s.e(\beta_1)_{OLS}$ 36out of 36 times (100%)

$s.e(\beta_1)_{GLS} < s.e(\beta_2)_{GLS}$ 30out of 36 times (83.3%)

$s.e(\beta_2)_{GLS} < s.e(\beta_1)_{OLS}$ 29 out of 36 times (80.6%)

Table 3: Goldfield and Quandt Test when $\varepsilon_i \sim (0, (X_1 * X_2)^h)$.

n	G	df1	df2	GQ	p-value
25	**0.5**	10	9	7.1784	3.34E-03
	1	12	12	3.8993	0.0129
	1.5	15	14	3.2273	0.01722
	2	17	17	3.4835	0.006933
	2.5	20	19	2.4613	0.0274
	3	22	22	2.4142	0.0221
30	**0.5**	10	9	6.8687	0.003918
	1	12	12	3.1508	0.02884
	1.5	15	14	6.5593	0.0005405
	2	17	17	2.417	0.03875
	2.5	20	19	4.8707	0.0005369
	3	22	22	4.5353	0.0003911
35	**0.5**	10	9	12.5004	0.0003997
	1	12	12	4.842	0.00529
	1.5	15	14	3.075	0.02108
	2	17	17	2.4396	0.03726
	2.5	20	19	2.7817	0.01496
	3	22	22	3.1919	0.004375
40	**0.5**	10	9	3.7481	0.02968
	1	12	12	4.2245	0.00935
	1.5	15	14	4.4585	0.003958
	2	17	17	3.008	0.01444
	2.5	20	19	2.8728	0.01266
	3	22	22	2.53	0.01718
45	**0.5**	10	19	4.7823	0.01372
	1	12	12	4.105	0.0105
	1.5	15	14	3.545	0.01147
	2	17	17	7.2299	8.77E-05
	2.5	20	19	2.5872	0.02153
	3	22	22	4.5315	0.0003936
50	**0.5**	10	9	5.554	0.008288
	1	12	12	3.3262	0.02368
	1.5	15	14	3.1812	0.0183
	2	17	17	7.8807	4.91E-05
	2.5	20	19	3.2054	0.007
	3	22	22	5.207	0.0001368

Table 4: Standard Error of Estimates of Estimators When $\varepsilon_i \sim (0, (X_1 * X_2)^h)$.

Sample sizes	g	$s.e(\beta_0)_{OLS}$	$s.e(\beta_0)_{GLS}$	$s.e(\beta_1)_{OLS}$	$s.e(\beta_1)_{GLS}$	$s.e(\beta_2)_{OLS}$	$s.e(\beta_2)_{GLS}$
	0.5	0.169357	0.060325	0.005599	0.003965	0.01121	0.005214
	1	0.27949	0.114798	0.01035	0.005661	0.01535	0.006124
	1.5	0.127798	0.060348	0.004445	0.002174	0.00718	0.002862
	2	0.131334	0.053012	0.005073	0.002421	0.007507	0.002968
	2.5	0.24022	0.09793	0.006959	0.003681	0.016203	0.007495
25	3	0.174926	0.064282	0.007312	0.004124	0.011564	0.004329
	0.5	0.73419	0.32565	0.03202	0.01293	0.04926	0.01857
	1	0.855	0.4095	0.03197	0.01775	0.04499	0.02366
	1.5	0.50093	0.26555	0.02601	0.01511	0.02982	0.01774
	2	0.85896	0.41373	0.02457	0.01434	0.05393	0.02623
	2.5	0.48543	0.19446	0.01979	0.01133	0.03024	0.0105
30	3	0.49977	0.206673	0.02018	0.009732	0.03045	0.013421
	0.5	4.8764	2.951	0.1749	0.144	0.278	0.1333
	1	1.66292	0.88422	0.0418	0.02644	0.09917	0.05687
	1.5	1.86692	1.06186	0.07447	0.04626	0.11042	0.06176
	2	2.88962	1.79147	0.09247	0.04349	0.18958	0.10411
	2.5	3.1073	1.69663	0.1294	0.07016	0.17	0.08696
35	3	1.73882	1.16339	0.07505	0.05294	0.10095	0.05912
	0.5	16.6515	10.3453	0.6025	0.3946	1.0077	0.5424
	1	12.328	7.0122	0.4587	0.3479	0.8019	0.4919
	1.5	17.7222	10.0505	0.5344	0.2089	1.1602	0.6841
	2	14.5314	6.8789	0.3855	0.1342	0.8852	0.4556
	2.5	10.8174	5.8727	0.4292	0.1717	0.6603	0.3816
40	3	9.5573	5.3948	0.4793	0.2635	0.6329	0.37
	0.5	63.764	21.1292	2.747	1.9429	3.811	0.8241
	1	67.761	24.292	2.275	1.193	4.338	1.129
	1.5	85.208	45.354	3.261	1.88	4.329	2.525
	2	68.734	21.523	2.779	1.216	3.958	1.315
	2.5	70.954	37.192	2.875	1.49	3.855	2.014
45	3	67.745	35.649	2.925	1.48	4.069	2.304
	0.5	311.14	214.86	15.36	13.81	18.72	10.97
	1	458.2	345.087	13.04	7.089	25.3	20.286
	1.5	458.96	230.99	19.92	12.28	23.52	11.37
	2	412.89	222.554	16.18	6.764	23.78	13.423
	2.5	375.73	156.648	14.43	6.909	22.14	8.952
50	3	465.7	239.604	14.7	9.673	27.51	13.066

DISCUSSION OF RESULTS

Empirically, Goldfeld-Quandt test showed that the simulated data via Monte Carlo simulation method is heteroscedastic in nature judging by P -values that are strictly less than the pre-specified level of significance, α=0.05. Also based on 1000 replication, varying magnitude of heteroscedasticity, h=0.5, 1, 1.5, 2, 2.5, and 3 respectively and for different sample sizes (n), 25, 30, 35, 40 and 50. The following observations were deduced from the analysis:

When error variance is related to the values of first explanatory variable (X_{1i}) at different sample sizes and different magnitude of heteroscedasticity, $s.e(\beta_j)_{OLS}$ is greater than $s.e(\beta_j)_{GLS}$ for $j = 0,1,2$. It was also observed from the table that, $s.e(\beta_1)_{GLS}$ gives the least standard error of estimate when compared to the $s.e(\beta_0)_{OLS}$ and $s.e(\beta_1)_{OLS}$, this may be as a result of the error term being related to values of the explanatory variable X_{1i}, in the multiple linear regression model with two independent variables only.

Also it was observed that sample size alone does not affect standard error of estimate but the magnitude of heteroscedasticity. These results are in tandem with that obtained by Davidson and Mackinnon (1993) using SLRM.

When the variance of error term relates to the product of values of variables X1 and X2, in regression models having only two predictors, we observed that the corresponding standard error of estimates of $\beta_j, j = 0,1,2$ offered by the GLS estimator is less than that of OLS estimator for all sample sizes at different magnitude of heteroscedasticity. Specifically $s.e(\beta_1)_{GLS}$ is the least. Next is $s.e(\beta_2)_{GLS}$ .Also $s.e(\beta_0)_{GLS}$ is less than $s.e(\beta_0)_{OLS}$.

The $s.e(\beta_1)_{OLS}$ is strictly less than $s.e(\beta_2)_{OLS}$ most of the times. This may be due to the error term relating to the product of the values of two explanatory variables up to a joint-multiplicative constant in the multiple linear regression models with two independent variables only (see Figures 1 and 2).

Graphical Display of $s.e(\beta_j)$ when $\varepsilon_i \sim (0, X_1^h)$.

Figure 1 (a): For n=25 and Varying Magnitude of Heteroscedasticity.

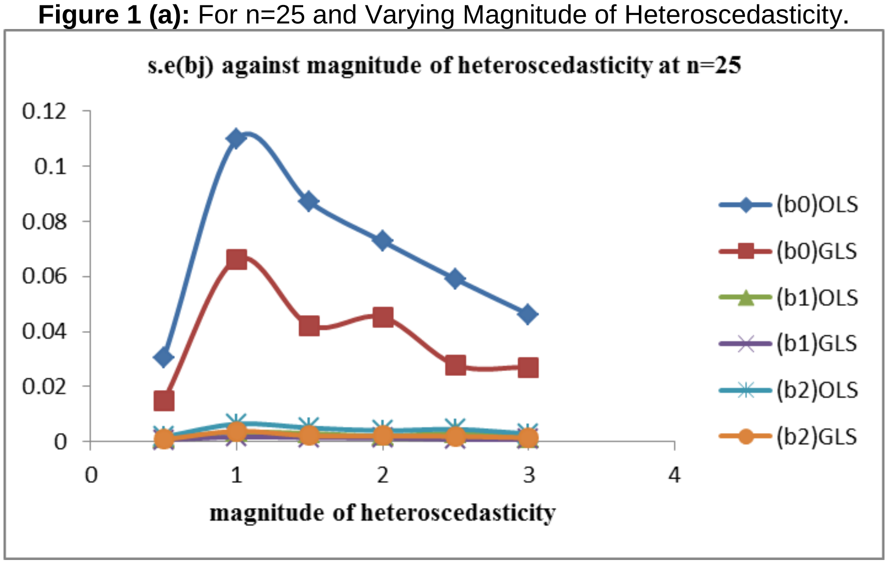
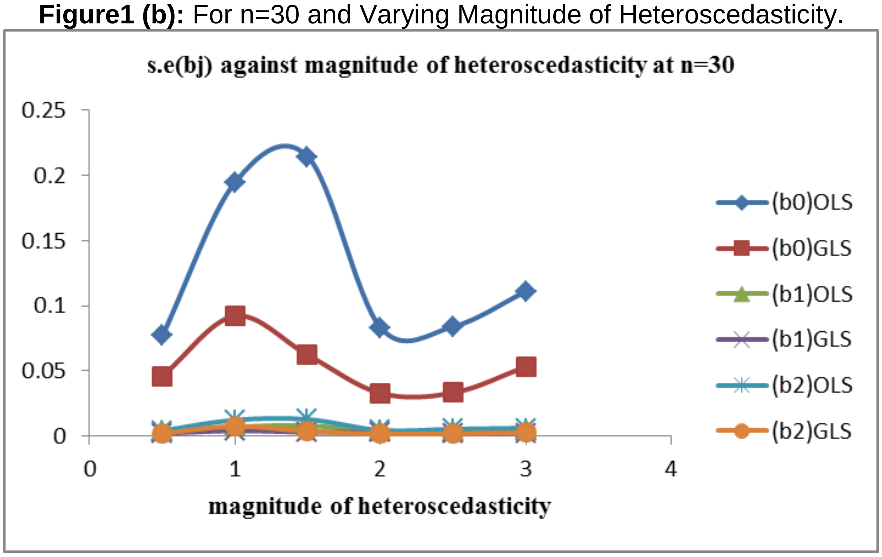


Figure1 (b): For n=30 and Varying Magnitude of Heteroscedasticity.



Graphical Display of $s.e(\beta_j)$ When $\varepsilon_i \sim (0, (X_1 * X_2)^h)$.

Figure 2(a): For n=25 and Varying Magnitude of Heteroscedasticity.

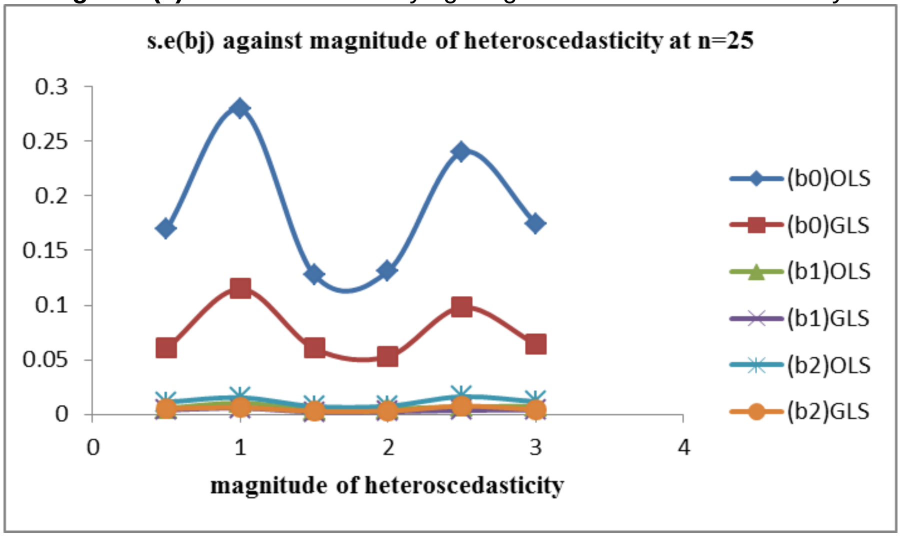
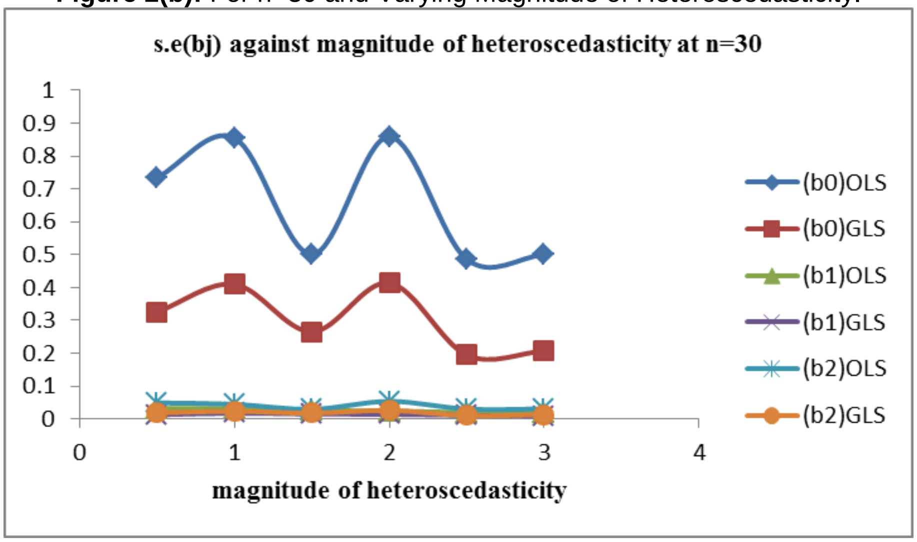


Figure 2(b): For n=30 and Varying Magnitude of Heteroscedasticity.



CONCLUSION

Based on the analysis carried out it was evident that OLS with heteroscedasticity over-estimate the true value of standard error of estimate obtained by the GLS estimator when error term is related to one or two explanatory variables up to multiplicative constant only in a multiple linear regression models hence establishing the superiority of GLS estimator.

In our findings, the effect of the standard error of estimate when error term is related to two explanatory variables up to a multiplicative constant in the multiple linear regression models is similar to when error term is related to the first regressor, except that the picture is not clear for that of $(\beta_1)_{OLS}$ and $(\beta_2)_{OLS}$ as we have in the results when error term is related to the product of the two explanatory variable. Hence we can conclude that heteroscedasticity affects the efficiency property of OLS estimator. There is upward bias scenario in the estimates of variances of the model parameters. Therefore, the magnitude of heteroscedasticity on a particular explanatory variable reduces its standard error of estimate. Again these results are in tandem with the result obtained by Davidson and Mackinnon (1993).

RECOMMENDATIONS

When heteroscedasticity is marked in a research data, it can lead to serious distortion of findings and seriously weaken the analysis, which may lead to misleading conclusion. Hence it is advisable to correct for heteroscedasticity so as to improve precision and make reliable conclusion when modeling a multiple linear regression fraught with heteroscedasticity of any amount of magnitude.

REFERENCES

1. Berry, W.D. and S. Feldman. 1985. “Multiple Regression in Practice”. Sage University Paper Series on Quantitative Application in the Social Science, Series no.07-050. Sage: Newbury Park, CA.

2. Damodar, N.G. 2011. *Basic Econometrics. Fourth Edition*. McGraw-Hill: New York, NY.

3. Davidson, R. and J.U.G. Mackinnon. 1993. *Estimation and Inference in Econometrics*. Oxford University Press: New York, NY.

4. Fomby, T.B., C.R. Hill, and S.R. Johnson. 1984. *Advanced Econometric Methods*. Springer-Verlag: New York, NY.

5. Goldfeld, S.M., and R.E. Quandt. 1972. *Nonlinear Methods of Econometrics*. North-Holland: Amsterdam: The Netherlands.

6. Gujarati, D. 1995. *Basic Econometrics, Third Edition*. McGraw-Hill: New York, NY.

7. Judge, G.G., C.R. Hill, W.E. Griffiths, H. Lütkepohl, and T.C. Lee. 1980. *Theory and Practice of Econometrics*. John Wiley & Sons: New York, NY.

8. Monte Carlo Experimentation using PC-NAÏVE. 1987. *Advances in Econometrics*. 91–125 (with A. J. Neale).

9. McCullach, P. and J.A. Nelder.1983. “Generalized Linear Models”. Number 37 in Monographs on *Statistics and Applied Probability*. Chapman & Hall: New York, NY.

SUGGESTED CITATION

Adejumo, A.O., O. Job, T.D. Isaac, and B.A. Oyejola. 2016. “Performances of Ordinary and Generalized Least Squares Estimators on Multiple Linear Regression Models with Heteroscedasticity”. *Pacific Journal of Science and Technology*. 17(1):68-78.

Pacific Journal of Science and Technology