International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 4 Issue 6, September-October 2020 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Comparative Analysis of Heteroscedastic and Homoscedastic OLS Models

AkpensuenShiaondo Henry[1], Joel Simon[1], Alhaji Abdullahi Gwani[2], Joshua Hassan Jemna[1]

[1]Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi, Nigeria
[2]Department of Mathematical Sciences, Bauchi State University, Gadau, Nigeria

ABSTRACT

This study considered foreign direct investment (FDI) as response variable while, gross domestic product (GDP), inflation and exchange rate were the predictor variables. The data were obtained from the Central Bank of Nigeria Statistical Bulletin spanning from 1970 to 2019. The study aimed at comparing heteroscedatic and homoscedastic OLS modes. Our findings revealed that the predictor variables in the heteroscedastic OLS model were not significant and were able to account for about 44% of the variation in the response variable. The diagnosis of the fitted regression model using BreuschPagan test showed that the assumption of homoscedasticity was violated. To address the problem of heteroscedasticity, all the variables were converted to log form to stabilise the variance. Our results from the now homoscedstic model revealed that all the predictor variables were significant and able to account for about 82% of the variation in the response variable. Therefore, our study established that when the assumption of homoscedasticity is violated, the model parameters become inefficient, the standard errors biased; and the t-statistics and the p-values no more valid. On the other hand, this study evidently proved that homoscedastic OLS model provide better estimates than heteroscedastic OLS model.

KEYWORDS: *Heteroscedasticity, Homoscedasticity, OLS, FDI, GDP*

How to cite this paper: AkpensuenShiaondo Henry | Joel Simon | Alhaji Abdullahi Gwani | Joshua Hassan Jemna "Comparative Analysis of Heteroscedastic and Homoscedastic OLS Models" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-6, October 2020, pp.879-882, URL: www.ijtsrd.com/papers/ijtsrd33553.pdf

IJTSRD33553

1. INTRODUCTION

Conventional regression model seeks to define the relationship between the dependent variable and the independent variables. This regression model could be simple (consisting of one dependent and one independent variable) or multiple (consisting of one dependent and two or more independent variables) [1].

However, linear regression models are tied to certain assumptions about the distribution of the error terms, some of the assumptions include linearity, homoscedasticity, normality and no autocorrelation between the error terms. Moreover, regression model describes the value of the dependent variable as the sum of two parts, the explanatory variables and the error term. The error term is primarily a disturbance to an already stable relationship and is able to capture the remaining information in the dependent variable which could not be explained by the independent variables.

Relating to the assumption of homoscedasticity, if the assumptionis violated, there are serious concerns for the OLS estimation. Although the estimators remain unbiased, the estimated standard error is wrong. Because of this the confidence interval and hypothesis test cannot be relied on. The underlying model would be rendered invalid with the standard errors of the parameters becoming biased. Moreover, if the errors are correlated, the least squares estimators are inefficient and the estimated variances are not appropriate [2-6].

By definition heteroscedasticity is a result of a data generating process that draws disturbances, for each value of the independent variable, from distributions that have different variances. It also implies that dispersion of the dependent variable around the regression line is not constant. Heteroscedasticity usually arises in cross sectional data where the scale of the dependent variable tends to vary across observations, and in highly volatile time series data. It is less common in other time series data where values of explanatory and dependent variables are of similar order of magnitude at all points of time. Thus, when applying regression models in the presence of heteroscedasticity, the ordinary least squares estimation method ceases to provide efficient estimators and appropriate variances. In an attempt to tackle heteroscedasticity, the study seek to profile and manage heteroscedasticity from OLS model and come up with more reliable OLS model devoid of heteroscedasticity.

Various methods were proposed in the literature to detect the presence of heteroscedasticity. Among the formal tests are: white test [7], Breusch-Pagan [8], Glejser test [9], Goldfeld- Quandt Test. [10] and Koenker-Bassett (KB) test [11].

Researchers have continued to admirably investigate and compared different tests of heteroscedasticity. According to [12] white test has low power for small sample. A comparison between Szroeter's asymptotic test and Goldfeld-Quandt (GQ) test. (Goldfeld and Quandt, 1980), Breusch-Pagan test (Breusch and Pagan, 1979) and BAMSET (Ramsey, 1969) was conducted by [14]. Goldfeld-Quandt test being the most popular and performed satisfactorily. Breusch-Pagan (BPG) test is also popular and powerful. The BAMSET is less sensitive. For the purpose of this paper we shall apply Breausch Godfrey test in detecting heteroscedastricity because of its popularity.

The remaining part of this work is organized as follows; materials and methods are presented in section 2, section 3 takes care of results and the discussion while conclusion of the study is handled in section 4.

2. Materials and Method

2.1. Method of Ordinary Least Squares Linear Regression

The least squares estimation procedure uses the criterion that the solution must give the smallest possible sum of squared deviations of the observed $Y_t$from the estimates of their true means provided by the solution. Let $\hat{\beta}_0$and $\hat{\beta}_1$be numerical estimates of the parameters $\beta_o$and β_1 respectively, and

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_t\hat{X}_t. \qquad (1)$$

Be the estimated mean of Y_t for each X_t t = 1, ..., n.

The least squares principle chooses $\hat{\beta}_0$ and $\hat{\beta}_t$ that minimize the sum of squares of residuals, (SSE)

$$SSE = \sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2 = \sum_{t=1}^{n}\varepsilon_t^2 \qquad (2)$$

Where,$\varepsilon_t = (Y_t - \hat{Y}_t)$ is the observe residuals for the ith observation

Also we can express ε_i in terms of Y_t, X_t, $\beta_o$and β_1. Hence, we have

$$\varepsilon_t = Y_t - \beta_0 - \beta_1 X_t \qquad (3)$$

Equation (3) becomes

$$SSE = \sum_{t=1}^{n}(Y_t - \beta_0 - \beta_1 X_t)^2 \qquad (4)$$

The partial derivative of SSE with respect to the regression constant $\hat{\beta}_0$, th

$$\frac{\delta SSE}{\delta\beta_0} = \frac{\delta}{\delta\beta_0}\left[\sum_{t=1}^{n}(Y_t - \beta_0 - \beta_1 X_t)^2\right] \qquad (5)$$

With some subsequent rearrangement, the estimate of $\hat{\beta}_0$ is obtained as

$$\hat{\beta}_0 = \left[\frac{\sum_{t=1}^{n} Y_t}{n}\right] - \beta_1\left[\frac{\sum_{t=1}^{n} X_t}{n}\right] \qquad (6)$$

The partial derivative of SSE with respect to the regression coefficient β_1. That is

$$\frac{\delta SSE}{\delta\beta_1} = \frac{\delta}{\delta\beta_1}\left[\sum_{t=1}^{n}(Y_t - \beta_0 - \beta_1 X_t)^2\right] \qquad (7)$$

Rearranging equation (7), we obtained the estimate of β_1.

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} Y_t X_t - \frac{\sum_{t=1}^{n} Y_t \sum_{t=1}^{n} X_t}{n}}{\sum_{t=1}^{n} X_t^2 - \frac{(\sum_{t=1}^{n} X_t)^2}{n}} \qquad (8)$$

2.2. Breusch Pagan Test

To illustrate this test, consider the P- variable linear regression model

$$y_i = \beta_1 + \beta_2 X_{2i} \dots \beta_p X_{pi} + \varepsilon_i \qquad (9)$$

Assume that the error variance σ_I^2 described as

$$\sigma_I^2 = f(\gamma_1 + \gamma_2 k_{2i} + \dots + \gamma_m y_{mi}) \qquad (10)$$

That is σ_I^2 is some function of the non-stochastic variables y's (it is assumed that the predictor variable is stochastic in nature and the regressor variables are non-stochastic in nature); some or all of the X's can serve as y's [14]

Specifically assume that $\sigma_I^2 = f(\gamma_1 + \gamma_2 k_{2i} + \dots + \gamma_m y_{mi})$, that is σ_I^2 is a linear function of z's. If

$\gamma_2 = \gamma_3 = \dots = \gamma_m = 0, \sigma_I^2 = \gamma_1$, then the variance is constant. Therefore, to test whether σ_I^2 is homoscedastic, one can test the hypothesis that $\gamma_2 = \gamma_3 = \dots = \gamma_m = 0$. This is the basics of Breusch Pagan test.

3. Results and discussion

This paper uses a data set on foreign direct investment (FDI) as response variable while gross domestic product (GDP), exchange rate and inflation as predictor variables spanning from 1970 to 2019. The data was obtained from CBN statistical bulletin.

Since the aim of our study is to compare heteroscedastic and homoscedastic OLS models, we begin by modelling the relationship between the response and predictor variables via linear regression. The fitted regression model is shown in equation 11while the estimates of the model are shown in table I below.

$$FDI = 9.97 \times 10^8 + 1.34 \times 10^{-5} GDP - 3539539 INFLATION + 13907293 EX \ (11)$$

Table 1: Estimates of OLS Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDP	1.34E+05	1.96E+05	0.682763	0.4983
INFLATION	-3539539.	17980202	-0.196858	0.8448
EX	13907293	8182957.	1.699544	0.0961
C	9.97E+08	5.30E+08	1.878977	0.0667
R-squared	0.444524			

From the estimates of the linear regression model in table 1 we observed that all the predictor variables are not significant since the p-values corresponding to GDP (0.4983), inflation (0.8448) and ex (0.0961) are more than 5% significance level and were able to only explain about 44% ($R^2 = 0.444524$) of the variation in FDI. A smallvalue of R^2(0.444524) is a good suggestion that the model does not fits the data very well. However, it is not the only measure of a good model when the model is to be used to make inferences [3]. Linear regression models are tied to certain assumptions about the distribution of the error terms. If these are seriously violated, then the model is not useful for making inferences. Therefore, it is importantto consider the appropriateness of the model for the data before further analysis based on that model is undertaken. To diagnosed the fitted model for heteroscedasticity, we apply the Breausch Pagan test, the result is shown in table 2

Table 2: Breusch-PaganHeteroskedasticity Test

F-statistic	8.056068	Prob. F(3,45)	0.0002
Obs*R-squared	17.12119	Prob. Chi-Square(3)	0.0007
Scaled explained SS	29.98588	Prob. Chi-Square(3)	0.0000

From the result in table 2, it is apparent that heteroscedasticity exist in the model since the p-value (0.0007) is less than 5% level of significance. To address this problem, we convert the variables to log form to stabilise the variance and run the regression model again. The result is shown in table 3.

Table 3: Estimates of OLS in Log Form

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(GDP)	0.289110	0.094049	3.074033	0.0036
LOG(INFLATION)	0.025728	0.107760	0.238753	0.0018
LOG(EX)	0.071811	0.122708	0.585219	0.0001
C	12.50668	2.385016	5.243856	0.0000
R-squared	0.823952			

From table 3 all the predictor variables are significance since their corresponding p-values are less than 5% level of significance level and jointly explain about 82% of the variation of the response variable. A large value of R^2(0.823952) is a good indication of how well the model fits the data. However, it is not the only the yardstick for measuringa good model when the model is to be used to make conclusions [15]. Linear regression models are tied to certain assumptions about the distribution of the error terms. For instance if the assumption of homoscedasticity which is our interest in this paper is violated, we have the problem of heteroscedasticity. Some of the consequences of heteroscedasticity are that, the ordinary least squares estimates will be inefficient i.e. they will no longer have the minimum variance in a class of unbiased estimators and hence are not BLUE, the conventional estimator of the variance of the error term is biased, the conventional formula for the OLS estimators of the variance of regression coefficients is wrong, the OLS estimator of the variances and covariances of the regression coefficients are biased, the conventionally constructed confidence intervals can no longer be valid, the t and F statistics based on the OLS regression do not follow the t and F distribution respectively and hence standard hypotheses tests are invalid. Therefore, to test for heteroscedasticicity, we again apply Breusch-Pagan test shown in table 4. Observing results from table 4, the p-value (0.7415) is more than 5% level of significance which indicates that the model is homoscedastic.

Table 4: Breusch-PaganHeteroskedasticity Test

F-statistic	0.680278	Prob. F(3,45)	0.5687
Obs*R-squared	2.125832	Prob. Chi-Square(3)	0.5467
Scaled explained SS	1.248181	Prob. Chi-Square(3)	0.7415

Comparing the estimates of the heteroscedastic model with the estimates of the homoscedastic model.

Table 5: heteroscedatic OLS model versus Homoscedastic OLS model

Model Heteroscedastic model					Homoscedastic model			
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
Parameter	9.9×10^{8}	1.34×10^{15}	-3539539	13907293	0.289110	0.025728	0.025728	0.071811
Std error	5.3×10^{8}	1.96×10^{5}	17980202	8182957	0.094049	0.094049	0.107760	0.122708
t-value	1.878977	0.682763	-0.196858	5.3×10^{5}	3.07403	3.074033	0.235753	0.585219
p-value	0.0667	0.4983	0.8448	0.0961	0.0000	0.0036	0.0018	0.0001

From Table 5, the core difference is the coefficients, standard errors and the p-values when calculations based on the estimated variance of the coefficient probability distribution, that is, the coefficient of standard error, t-statistic and probability value (p-value). The standard errors are smaller when accounting for heteroscedasticity; that is to say, in homoscedastic regression model, the standard error, t-statistic and p-value are significantly different from those of the heteroscedastic regression model. The implication is that homoscedastic regression model gives better estimates than the heteroscedastic regression model.

4. Conclusion

The study modelled the effect of violating the assumption of constant variance or homoscedasticity in a linear regression model. First of all, the relationship between the response variable, FDI , and the predictor variables, GDP, inflation and exchange rate , was determined using the ordinary least squares estimation method. The results of the ordinary least squares estimated regression revealed that GDP, inflation and exchange ratewere not able contributed significantly to FDI and were able to explain about 44.45% of the variance in FDI. Furthermore, evidence from Breusch -Pagan test, revealed that heteroscedasticity exist in the model. To address the effect of heteroscedasticity on the model, the variables were converted to log form to stabilise variance and a regression model was run again. The results of our analysis revealed that the predictor variables (GDP, inflation and exchange rate) became significant to the response variable (FDI) and were able to explain about 82% of the variation in the response variable (FDI). Therefore, our study established that when the assumption of homoscedasticity is violated, the model parameters become inefficient, the standard errors biased; and the t-statistics and the p-values no more valid. On the other hand, this study obviously proved that homoscedastic OLS models provide better estimates than heteroscedastic OLS models.

References

[1] Emanuel A. A &Imoh U. M (2018). Modelling the auto correlated errors in time series: A Generalised least squares Approach. Journal of advances in Mathematics and Computer Science. 26(4), 1-15

[2] Granger CWJ, Newbold P (1974). Purious regressions in econometrics. Journal of Econometrics.2:111–120.

[3] Rawlings JO, Pantula SG, Dickey DA (1998). Applied regression analysis: A research tool. 2nd ed.Springer;.

[4] Gujarati DN (2004). Basic econometrics. 4th ed. McGraw-Hill;.

[5] Akpan EA, Moffat IU (2016). Dynamic time series regression: A panacea for spurious correlations. International Journal of Scientific and Research Publications. 2016;6(10):337-342.

[6] Akpan EA, Moffat IU, Ekpo NB (2016).Modeling regression with time series errors of gross domestic product on government expenditure. International Journal of Innovation and Applied Studies.18 (4): 990-996.

[7] White H (1980). A heteroskedasticity consistence covarianc matrix estimator and direct test for heteroskedasticity.Econometrica.48(2).

[8] Breusch T. S and Pagan A. R (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrical; 47(5).

[9] Glejser H (1969). A new test for heteroscedasticity. Journal of the American statistical association; 64(325), 316-324

[10] Goldfel S. M and Quandt R. E (1965). Some test for heteroscedasticity. Journal of American statistical association; 60(310), 539-547

[11] Koenker R and Bassett G.Jr (1982). Robust test for heteroscedasticity base on quantiles. Journal of econometric society ; 43(61)

[12] Ramsey, J. B. (1969). Tests for speci_cation error in classical linear least-squares regression analysis. Journal of the Royal Statistical Society, B31, 252.

[13] Griffiths W. E, Surukha K (1985). A monte Carlo evaluation of the power of some test for heteroscedasticity. University of New England, department of Econometrics

[14] Usman et.al. (2019). The use of weighted least squares method when the error variance is heteroscedastic. Benin Journal of statistics; 2, 85-93

[15] Jack E. M and Brian R. H. (2007) Evaluating Aptness of a Regression Model, Journal of Statistics Education, 15:2, DOI:10.1080/10691898.2007.11889469