International Journal of Mathematical Analysis and Modelling

(Formerly Journal of the Nigerian Society for Mathematical Biology)

Volume 7 Issue 2, 2024

ISSN (Print): 2682 - 5694
ISSN (Online): 2682 – 5708
www.tnsmb.org

Detection and correction of heteroscedasticity and its effect on modelling of Nigerian economic data

E.O. Idowu*†, E.M. Ikegwu‡, A.A. Fadiji†§ and M.U. Evro†**

Abstract

The study centred on detection and correction of heteroscedasticity and its effect on modelling of economic data. Data were collected from CBN Statistical bulletin on five economic variables namely; Gross Domestic Product (GDP), Inflation Rate, Exchange Rate (Exch_rate), Balance of Payment and Government Debt from 1987 - 2017. The data was analysed using the Ordinary Least Square method and variance stability techniques was used to correct heteroscedasticity using log transformation of the dependent variable while Breusch-Pagan, White and Harvey-Godfrey tests were used to detect the presence of heteroscedasticity in the variables at 5% level of significance with the aid of E-views 9. The regression model fitted to the data set was **GDP = 37242.63 - 256.7745 Inflr – 40.196 Exchr + 252.364 BoP + 3.147 GovD** and it showed that only inflation rate and Government debt were discovered to be significant predictors ($p < 0.05$) while and heteroscedasticity was detected. Result showed that after log transforming the variables, heteroscedasticity was eliminated as shown by Breusch-Pagan, White and Harvey-Godfrey test respectively. It is concluded therefore that transforming economic data helps correct and eliminate detected heteroscedasticity in modelled economic variables.

Keywords and phrases: Heteroscedasticity, GDP, Inflation Rate, Exchange rate, OLS

1 Introduction

Most linear model has as a basic assumption homoscedasticity which implies that the variance of the residual must be constant across the time interval or the explanatory variables. However, in most studies, this assumption is violated, and heteroscedasticity is observed. Heteroscedasticity represents a situation where the variance of the residuals varies across observations within a regression model, i.e. the disturbances (error) term does not have a constant or equal variance independent of time (t). Heteroscedasticity was viewed as the violation of constant variance assumption of ordinary least squares estimator [1, 2]. The violation of the assumption of homoscedasticity could increase Type I

*Corresponding Author; E-mail: godsbabe2002@gmail.com
†Department of Statistics, Yaba College of Technology, Yaba Lagos, Nigeria.
‡Department of Statistics, Yaba College of Technology, Yaba Lagos, Nigeria.
E-mail: emmanuel.ikegwu@yabatech.edu.ng
§E-mail: ademola.fadiji@yabatech.edu.ng
**E-mail: merit.evro@gmail.com

error, reduce statistical power, lead to inefficient estimation or incorrect inference [2 - 6] but does not affect the unbiasness of the model estimates.

Heteroscedasticity according to [7] is a common feature of cross - sectional data and correcting for it very vital for appropriate inference with the estimated model. This study explored consistent lack of fit tests to assess the presence of heteroscedasticity in linear regression models and hence correct for same as expounded by other studies [4, 8].

Economic growth is usually exponential in nature according to [9] and so modelling such always requires transformation using appropriate procedures like the Box – Cox transformation. This study attempted the detection and correction of heteroscedasticity in economic data and also evaluated its effect in economic models. The outcome will not only contribute to literatures on the concept of heteroscedasticity but also enable governmental agencies and policy makers as it deals with correcting its effects on models of economic growth.

1.1 Detection

It is always necessary to assess the assumption of the linear model and one of such is the assumption of homoscedasticity (equal variance) of the residual over the period. Graphical diagnostic procedures, such as plotting the residuals versus the time period or the fitted value, often provide useful visual aid [10, 11, 12]. Other formal tests are also employed to evaluate whether the patterns observed are due to random fluctuations [13]. Some of the formal tests include white's test, Gold Quandt (GQ) test, Harvey test, Breusch – Pagan – Godfrey test [14, 1, 15]. However, the White test procedure has become popular for detecting heteroscedasticity and is included in most software.

1.2 Correction

[7] used WLS to correct the presence of heteroscedasticity which led to an efficient and significant result. [4] managed the problem of heteroscedasticity by transforming the response variable. However, [16] dealt with the problem of non – constant variance by transforming both the predictor and response variables.

2 Methodology

This study used five (5) economic variables namely: Gross Domestic Product (GDP), Inflation Rate, Exchange Rate, Balance of Payment and Government Debt for a 31 years period spanning 1987 - 2017 obtained from the Central Bank of Nigeria statistical Bulletin to model economic growth in Nigeria.

With the major objectives of detection and correction of heteroscedasticity in economic data and also evaluating its effect on models of economic growth, the OLS model was fitted and the presence of heteroscedasticity examined. The log transformation was employed in the OLS model to obtain the estimates and examine the resultant effects on heteroscedasticity. However, because it has been shown that heteroscedasticity in longitudinal time series data was not as impactful as in cross – sectional data

2.1 The OLS Model Specification

The functional model in this study is given as:

$$GDP = \beta_0 + \beta_1 * Inflr + \beta_2 * Exchr + \beta_3 * BoP + \beta_4 * GovD + \varepsilon_i \qquad (1)$$

where

GDP = Gross Domestic Product (a proxy for measuring economic growth)

Inflr = Inflation Rate; Exchr = Exchange Rate; BoP = Balance of Payment;

GovD = Government Debt; ε_i = Stochastic error term

β_i's = Regression parameters (i = 0, 1, …, 4)

2.2 Detecting Heteroskedasticity

2.2.1. White's General Heteroscedasticity Test

White's test estimates the parameters of an auxiliary regression model with OLS residual U_i as the dependent variable in the form below:

$$U_i^2 = \beta_0 + \beta_1 X_1 + \cdots + v_i \quad (2)$$

The new coefficient of determination (R^2) is used to test the null hypothesis that there is no heteroscedasticity and the test statistics is

$$n.R^2 \sim x_{df}^2 \quad (3)$$

where df is the number of parameters in the auxiliary regression model without intercept [17].

2.2.2. The Breusch – Pagan - Godfrey LM Test

In the Breusch – Pagan – Godfrey test, the OLS model yields the residuals and the auxiliary regression is obtained as follows:

$$P_i^2 = \beta_0 + \beta_1 X_1 + \cdots + v_i \quad (4)$$

$$P_i^2 = \sum \left({u_i^2}/{\sigma^2} \right) \quad (5)$$

With u_i^2 and σ^2 were estimated from the initial OLS model. Then, $\Theta = {1}/{2}\,(ESS)$ where ESS is the explained sum of square from the auxiliary regression and $\Theta \sim \chi_{m-1}^2$ [8].

2.3 Correcting/ Resolving Heteroscedasticity

The variance stabilising technique of correcting heteroscedasticity was adopted using log transformation of the dependent variable (GDP) in line with literature [4].

The transformed model using log transformation is:

$$\text{Log\_GDP} = \beta 0 + \beta_1 \text{INFLR} + \beta_2 \text{EXCHR} + \beta_3 \text{BOP} + \beta_4 \text{GovD} + \varepsilon_t \quad (6)$$

The data was analysed using Statistical package for Social Sciences (SPSS 25.0).

3 Results

The result of the analysis of the economic variables are presented and interpreted and includes the test for the presence or otherwise of heteroscedasticity.

Table 1. Descriptive statistics for individual variables

Statistic	**GDP**	**Inflr**	**Exchr**	**BoP**	**GovD**	**Residual**
Mean	38796.5	20.63	111.847	-0.489	3644.99	1.13e-11
Median	25647.30	12.880	91.500	-1.247	1166.001	-545.230
Maximum	135277.2	72.840	305.800	10.000	40640.40	13468.230
Minimum	15242.60	5.380	49.780	-9.700	36.789	-15555.54
Std. dev	25548.47	18.646	62.249	4.775	7490.089	9067.091
Skewness	1.864	1.526	1.863	0.654	4.092	-0.0834
Kurtosis	7.435	3.916	5.648	3.359	20.568	1.823
Probability	0.000	0.001	0.000	0.304	0.000	0.402
Observations	31	31	31	31	31	31

Table 1 reveals that the mean (std deviation) of GDP is 38796.5 (25548.47), Inflation rate 20.63 (18.65), Exchange rate 111.85, (62.249), Balance of payment -0.489, (4.775) and Government debt 3644.99 (7490.089). It also showed that the residual is normally distributed ($p > 0.05$) is almost symmetrical conforming with OLS normality assumption.

Table 2. Estimation coefficients and model significance

Regressors	Actual data			Log transformed data		
	B	t	p-value	B	t	p-value
Constant	37242.63	8.478826	0.0000	10.548	68.629	0.0000
Inflation Rate	-256.7748	-2.605731	0.0150*	-0.0097	-2.8066	0.0094*
Exchange Rate	-40.19632	-1.144763	0.2627	-0.0013	-1.0814	0.2894
Balance of payment	252.3637	0.647752	0.5228	0.0108	0.7937	0.4345
Government Debt	3.146929	10.62749	0.0000*	0.0000564	5.4397	0.0000*
$R^2 = 0.874$; Adj. $R^2 = 0.855$; $p(F_{4,26}) < 0.001$; DW = 0.452				$R^2 = 0.683$; Adj. $R^2 = 0.635$; $p(F_{4,26}) < 0.001$; DW = 0.498		

The OLS model fitted is given as:

$$\boldsymbol{GDP = 37242.63 - 256.775 * Inflr - 40.196 * Exchr + 252.364 * BoP + 3.147 * GovD} \tag{7}$$

From the relative statistics, we discover that two (2) variables Inflation Rate and Government debt are statistically significant ($p < 0.05$) while Exchange Rate and Balance of payment are not statistically significant ($p > 0.05$)

The inflation rate ($\beta_{Inflr} = -256.775$) has the greatest effect on GDP as a unit change in inflation rate decreases GDP by 256.775 units. Also, BOP ($\beta_{BoP} = 252.36$) which mean a unit change in BOP increases GDP by 252.3. in addition, a unit increase in exchange rate ($\beta_{Exchr} = -40.196$) decreases GDP by 40.196 and Government Debt ($\beta_{GovD} = 3.15$) imply that a unit change in Govt Debt increases GDP by 3.15.

The $p(F_{4,26}) < 0.001$ implies that the dependent variables combined significantly contributed to GDP while the Adjusted R^2 of 0.855 shows that Inflation rate, Exchange rate, Balance of Payment and Government Debt collectively explained 85.5% of the total variability in GDP over the years modelled. Lastly, the Durbin Watson obtained was 0.452 which shows that there is negative serial correlation among the dependent variable.

The OLS model of the fitted log transformed data is given as:

$$\boldsymbol{Log\_GDP = 10.548 - 0.0097 * Inflr - 0.0013 * Exchr + 0.0108 * BoP + 0.0000564 * GovD} \tag{8}$$

From the relative statistics, two (2) variables also, Inflation Rate and Government debt are statistically significant ($p < 0.05$) while Exchange Rate and Balance of payment are not statistically significant ($p \geq 0.05$) (see table 2)

The result shows that as a unit change in inflation rate ($\beta_{Inflr} = -.0097$) decreases GDP by about 0.1%, a unit change in exchange rate ($\beta_{Exchr} = -.0013$) reduces GDP by 0.13%, a unit increase BOP in ($\beta_{BoP} = 0.0108$) increases GDP by 1.1% and a unit change in Govt Debt ($\beta_{GovD} = 5.64E\text{-}05$) increases GDP by 0.006%.

The $p(F_{4,26}) < 0.001$ implies that the dependent variables combined significantly contributed to Log_GDP while the Adjusted R^2 of 0.635 shows that Inflation rate, Exchange rate, Balance of Payment and Government Debt collectively explained 63.5% of the total variability in Log_GDP over the years modelled. Lastly, the Durbin Watson obtained was 0.498 which shows that there is negative serial correlation among the dependent variable.

3.1 Heteroscedasticity Test

H_0: There is no heteroscedasticity vs H_1: There is heteroscedasticity

Table 3: Test of heteroscedasticity

Test	a. Actual data		b. Log transformed data	
	p -value	**Decision**	**p - value**	**Decision**
Breusch – Pagan – Godfrey	0.0089*	heteroscedasticity	0.0997	No Heteroscedasticity
White	0.0305*	heteroscedasticity	0.1097	No Heteroscedasticity
Harvey	0.0093*	heteroscedasticity	0.0492	Heteroscedasticity

The three tests in table 3(a) show that heteroscedasticity was a problem as they are all significant ($p < 0.05$) and it is concluded that there is presence of heteroscedasticity (i.e. unequal variance).

3.2 Correction of Heteroscedasticity

The adopted correction for this study using log transformation to transform the dependent variable (GDP) is presented also in table 2.

The Breusch - Pagan – Godfrey and White tests above show that heteroscedasticity was no longer problem as they are not significant ($p > 0.05$) and it is concluded that there is absence of heteroscedasticity (i.e. unequal variance) but the Harvey was slightly different (see table 3 (b)).

5 Discussions and Conclusion

The result of the study showed that heteroscedasticity was present in the time series economic data because they are exponentially distributed agreeing other studies [14, 9]. The observed problem was detected using the traditional tests and corrected for using the Cox transformation by log transforming the dependent variable and the residuals were homoscedastic agreeing with other researchers [18, 4]. From our analysis, we have been able to detect heteroscedasticity in the residual of the modelled economic data and correction was made using log transformation. The log transformed data show no presence of heteroscedasticity. Unequal variance (heteroscedasticity) affects efficiency therefore making the OLS estimate no longer BLUE, even though the unbiasedness property holds. To eliminate the chances of heteroscedasticity, more observations and more explanatory variables should be used for analysis for efficient estimate (result). Other methods for detection and correction of heteroscedasticity should be explored

References

[1] Rehal, V. (2023, February 10). White Test for Heteroscedasticity. Retrieved June 24, 2024, from https://spureconomics.com/white-test-for-heteroscedasticity/#google_vignette
[2] Obomeghie, M. A. (2008). Applied Econometrics. An Introductory Text. Benin City: Isa - Mini Printers.
[3] Rehal, V. (2023, February 08). Heteroscedasticity: Causes and Consequences. Retrieved June 24, 2024, from https://spureconomics.com: https://spureconomics.com/heteroscedasticity-causes-and-consequences/
[4] Rosopa, P. J., Schaffer, M. M., & Schroeder, A. N. (2013). Managing Heteroscedasticity in General Linear Models. Psychological Methods, 18(3), 335 - 351.
[5] Ruppert, D., Wand, M. P., & Carroll, R. J. (2009). Semiparametric regression during 2003 - 2007. Electronic Journal of Statistics, 3, 1193 – 1256.
[6] Greens, W. H. (2000). Econometric Analysis, 4th ed. (4th ed.). Upper Saddle River, New Jersey: Prentice Hall.
[7] Agunbiade, D. A., & Adeboye, N. O. (2012). Estimation of Heteroscedasticity Effects in a Classical Linear Regression Model of a Cross - Sectional Data. Progress in Applied Mathematics, 4(2), 18 – 28.
[8] Gujarati, D. N. (2013). Basic Econometrics (5th ed.). Glasgow: Bell & Bain Ltd.
[9] Jones, C. I. (2016). The Facts of Economic Growth. In C. I. Jones, Handbook of Macroeconomics 2A: 3 - 69. online: Elsevier B.V. doi:10.1016/bs.hesmac.2016.03.002.
[10] Tan, F., Jiang, X., Guo, X., & Zhu, L. (2021). Testing Heteroscedasticity for Regression Models based on Projections on JSTOR. Statistica Sinica, 31(2), 625-646. doi:27008056.
[11] Feng, S., Li, G., Tong, T., & Luo, S. (2019). Testing for heteroskedasticity in two-way fixed effects panel data models. Journal of Applied Statistics, 1634682. doi:10.1080/02664763.2019.1634682.
[12] Cohen, P., West, S. G., & Aiken, L. S. (2007). Applied multiple regression/correlation analysis for the behavioral sciences. Erlbaum: Mahwah.
[13] Qian, L., & Wang, S. (2001). Bias-corrected heteroscedasticity robust covariance matrix (sandwich) estimators. Journal of Statistical Computation and Simulation, 70, 161 – 174.
[14] Romeo, G., Buonaccorsi, J. P., & Thoresen, M. (2023). Detecting and correcting for heteroscedasticity in the presence of measurement error. Communications in Statistics - Simulation and Computation, 2190061. doi:10.1080/03610918.2023.2190061.
[15] Berenguer-Rico, V., & Wilms, I. (2021). Heteroscedasticity testing after outlier removal. Econometric Reviews, 40(1), 51 - 58. doi:10.1080/07474938.2020.1735749.
[16] Van Keilegom, I., & Wang, L. (2010). Semiparametric modeling and estimation of heteroscedasticity in regression analysis of cross-sectional data. Electronic Journal of Statistics, 4, 133 – 160.
[17] White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity. Econometrica, 48, 817 – 838.
[18] Astivia, O. L., & Zumbo, B. D. (2019). Heteroskedasticity in Multiple Regression Analysis: What it is, How to Detect it and How to Solve it with Applications in R and SPSS. Practical Assessment, Research & Evaluation, 24(1), 1 - 16.