NIPES

Journal of Science and Technology Research

Journal homepage: www.nipesjournals.org.ng

Application of Transformation of Variables in Remedying Heteroscedasticity in Nigeria GDP, Conditioning and Some Fiscal Variables

Aideyan Donald Osaro
Dept of Mathematical sciences, Taraba State University, Jalingo

Article Info

***Keywords*:**
Heteroscedasticity, Estimates, Variance, Explanatory, Transformation.

Received 29 January 2023
Revised 10 February 2023
Accepted 10 February 2023
Available online 15 March 2023

https://doi.org/10.5281/zenodo.7737935
ISSN-2682-5821/

Abstract

Heteroscedasticity is a significant problem in regression analysis and occur in any situation where the error variance is not constant. The presence of heteroscedasticity is problematic, as changes in the dependent variable cannot be accurately attributed to individual explanatory variables. It can cause estimated coefficients to be unstable and have high variances and thus be potentially inaccurate to guide management policy. The main objective of this research work is to detect the violation of the constancy of variance in the data set and provide remedial measures to remove it in the classical linear regression model. The two estimators were compared using standard error of regression and coefficient of determination. A number of key findings are identified such as B-P and Koenker test, white test with the following statistics: p-value of 0.00, 50.031 before transformation and 0.718,0.405 and 22.338 after transformation respectively. Based on these figures, we conclude that transformed least squares (TLS) outperformed ordinary least squares (OLS) in the presence of heteroscedasticity. It should be used as method of analysis when the error variance is not constant. However, government should spend much especially on expenditure on economic service and expenditure on administration so as to increase the GDP.

1.0 Introduction

Regression analysis is the most important tools and methods used in the process of statistical analysis [1]. It is concerned with describing and evaluating the relationship between a variable called a dependent variable and one or more other known variable(s) called independent variable(s). The regression model has good predictive ability by estimating the coefficient using the least squares method and the investigation of the assumptions such as linearity, constant variance, normality and the independent of the disturbance term [3,4,5]. When one of these assumptions is violated, the tests of hypothesis such as t-test, f-test is no longer valid or inappropriate. In this research, we are going to study one of the assumptions in the regression model which is the constant variance. We shall also talk about some of the most common and appropriate detection and remedy methods in order to examine and solve the problem of heteroscedasticity to a good model for prediction. Heteroscedasticity is a phenomenon where data set violates a statistical assumption (i.e homoscedasticity) [3]. When this assumption is violated, this can lead to increase in type 1 error rates or decrease statistical power [2].

1.1. Consequences of Heteroscedasticity

The consequences of Heteroscedasticity are:

- ✓ The OLS ESTIMATORS are no longer BLUE because they are no longer efficient, so the regression prediction will be inefficient too [6].
- ✓ Because of the inconsistency of the covariance matrix of the estimated regression coefficients, the tests of hypothesis (t-test, f-test) are no longer valid [9].

This research work is aimed at testing and finding a remedy to a violation of assumption of constancy of variance (homoscedasticity) in the regression model by the use of transformation of variable with the following objectives.
1. If there is Heteroscedasticity existence in the fitted econometric model.
2. To provide remedy and to correct the non-conformity of the data to assumption.
3. To assertion whether government expenditures and the contribution on GDP in Nigeria has declined.

2.0. Methodology

2.1. Research design

The endogenous model often specified for testing the effects of fiscal variables on growth broadly consist of a regression equation with gross domestic product (GDP) as dependent variable and a set of conditioning variable as well as some fiscal variables of interest as the regressors [7]. This same broad approach is adopted in this study. Our model contains Gross Domestic Product as the dependent variable while Expenditure on Economic Services (EES), Social Community Services (ESCS), Economic Transfer (ETRAN) and Economic Administration(EAD) are the independent variables.
Specifically, the model can be specified in this form as:
GDP= f (EES, ESCS, ETRAN, EAD) (1)
Equation (1) above can be transformed into an econometric model as follows:

$$GDP = \beta_0 + \beta_1 EES + \beta_2 ESCS + \beta_3 ETRAN + \beta_4 EAD + \varepsilon_i \quad (2)$$

Where,

Y= Gross domestic product
X_1= Expenditure on economic service
X_2= Expenditure on social and economic service
X_3= Expenditure on transfer
X_4= Expenditure on administration
β_0 : is the value of Y (GDP) when X_1 (EES), X_2 (ESCS), X_3 (ETRAN) and X_4 (EAD) are all zero.
β_1 : is the increase in Y (GDP) with a unit increase in X_1 (EES) when X_2 (ESCS), X_3 (ETRAN) and X_4 (EAD) are all zero
β_2 :Is the unit increase in Y(GDP) with a unit increase in X_2 (ESCS) when X_1 (EES), X_3 (ETRAN) and X_4 (EAD) are all zero.
β_3 : is the unit increase Y (GDP) with a unit increase in X_3 (ETRAN) when X_1 (EES), X_2 (ESCS) and X_4 (EAD) are all zero.

β_4 : is the unit increase in Y (GDP) with a unit increase in X_4 (EAD) when X_1 (EES), X_2 (ESCS) and X_3 (ETRAN) are all zero.

Heteroscedasticity possess potentially severe problems for inferences based least squares. It is useful to be able to test for heteroscedasticity and if necessary, modify our estimation procedures accordingly [2]. Several types of tests have been suggested.

In this research work, the **BREUSCH-PAGAN** and **WHITE TEST** for Heteroscedasticity will be used.

2.2. Transformation of Variable or (Log transformation)

Since the residuals are asymmetrically distributed, we use a log transformation of GDP so as to correct Heteroscedasticity in the data.

i.e $LogGDP = f(EES, ESCS, ETRAN, EAD)$ (3)

More specifically

When the matrix X has full rank of p, the transformed estimator β can be obtained by minimizing the sum of squares residuals [8].

$$\hat{e}'\hat{e} = \left(logY - X\hat{\beta}\right)'\left(logY - X\hat{\beta}\right) \quad (4)$$

Hence,

$$\hat{\beta} = [X'X]^{-1}X'\,logY \quad (5)$$

Where $\hat{\beta}$is p x 1 vector of estimated parameters.$\hat{\beta}$Provide the minimum variance of any linear function of the observations. When there are q dependent variables the Transformed estimator in equation (3.15) can be generalized as follows;

Transformed $\hat{\beta} = [X'X]^{-1}X'\,logY$ (6)

Where the transformed $\hat{\beta}$is the transformed least square estimate of β.

3.0 Results and Discussion

3.1. Breusch-pagan and koenker test before transformation

Original Regression model:

Dependent variable

GDP

R-square

.850

OLS Output1

	b	se	t	sig
constant	-442986.1	616194.13	-.719	.476
EES	8.762	4.947	1.771	.083
ESCS	-36.385	9.966	-3.651	.001
ETRAN	36.978	5.764	6.415	.000
EAD	-.420	.399	-1.053	.298

---- Breusch-Pagan and Koenker test statistics and sig-values --------

	LM	Sig
BP	263.418	.000
Koenker	33.576	.000

Null hypothesis: heteroskedasticity not present (homoskedasticity)

if sig-value less than 0.05, reject the null hypothesis

------- ANOVA TABLE --------

	SS	df	MS	F	Sig
Model	526.836	4.000	131.709	22.160	.000
Residual	273.399	46.000	5.943	-999.000	-999.000

3.2. Breusch-pagan and Koenker test after transformation

Original Regression model:

Dependent variable
LG10GDP
R-square
.637
OLS Output2

	b	se	t	sig
constant	4.654	.142	32.826	.000
EES	5.317	.000	1.622	.112
ESCS	-1.048E-5	.000	-2.033	.048
ETRAN	8.527E-6	.000	3.080	.003
EAD	1.933E-7	.000	.708	.483

------- Breusch-Pagan and Koenker test statistics and sig-values-

	LM	Sig
BP	2.098	.718
Koenker	4.006	.405

Null hypothesis: heteroskedasticity not present (homoskedasticity)

If sig-value less than 0.05, reject the null hypothesis

------- ANOVA TABLE --------

	SS	df	MS	F	Sig
Model	59.612	4.000	14.903	20.157	.000
Residual	34.010	46.000	.739	-999.000	-999.000

3.3. White-Test before transformation

Table 1:
Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.990[a]	.981	.975	7999349350 214.07200

a.Predictors: (Constant), EES,ESCS, ETRAN,EAD,EES2,ESCS2,EAD2,ETRAN2,ESCS.EES, EES.ETRAN,ESCS.EAD,ESCS.ETRAN,ESCS.EAD, ETRAN.EAD

3.4. White-Test after transformation

Table 2:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.662[a]	.438	.280	3.09638

a. Predictors: (Constant), EES,ESCS, ETRAN,EAD,EES2,ESCS2,EAD2,ETRAN2,ESCS.EES, EES.ETRAN,ESCS.EAD,ESCS.ETRAN,ESCS.EAD, ETRAN.EAD

3.5. Discussion of Findings and Interpretation of Results

Here, the results of the performances of the estimators considered under heteroscedasticity are discussed. The estimators were assessed based on correctness of signs of the coefficients estimates, standard error of estimates and coefficient of determination.

3.6. Test of Heteroscedasticity

Breuch-pagan test

Ordinary least squares (Before transformation)

$\boldsymbol{H_0}$: Error variance constant (Homoscedasticity)

$\boldsymbol{H_1}$: Error variance is not constant (Heteroscedasticity)

$\propto = 0.05$

Decision rule: reject the null hypothesis if p-value $\leq \propto$

Interpretation: From the results in OLS output1 for breusch-pagan and koenker test before transformation, we therefore reject the null hypothesis and conclude that the error variance is not constant (Heteroscedastic) since the p-value $< \propto$

Transformed least squares (after transformation)

$\boldsymbol{H_0}$: Error variance constant (Homoscedasticity)

$\boldsymbol{H_1}$: Error variance is not constant (Heteroscedasticity)

$\propto = 0.05$

Decision rule: reject the null hypothesis if p-value $\leq \propto$

Interpretation: From the results in the OLS output 2 for breusch-pagan and koenker test after transformation, we therefore fail to reject the null hypothesis and conclude that the error variance is constant (Homoscedastic) since the p-value $> \propto$

3.7. White Test

Ordinary least squares (before transformation)

Hypothesis

$\boldsymbol{H_0}$: Error variance is constant (Homoscedasticity)

$\boldsymbol{H_1}$: Error variance is not constant (Heteroscedasticity)

n = 51

$R^2 = 0.981$

Test statistic $nR^2 \sim X_p^2$

$nR^2 = 51 \times 0.981 = 50.031$

$X_{14}^2 = 23.685$

Decision rule: Since $X_{cal}^2 > X_{tab}^2$, we reject H_0 and conclude that error variance is not constant (Heteroscedasticity)

.**Transformed least squares (after transformation)**

$\boldsymbol{H_0}$: Error variance constant (Homoscedasticity)

$\boldsymbol{H_1}$: Error variance is not constant (Heteroscedasticity)
n = 51
$R^2 = 0.438$
Test statistic $nR^2 \sim X_p^2$
$nR^2 = 51 \times 0.438 = 22.338$
$X_{14}^2 = 23.685$
Decision rule: Since $X_{cal}^2 < X_{tab}^2$, we fail reject H_0 and conclude that error variance is constant.

3.8 Results of OLS and TLS
The results are assessed in terms of
(a) Correctness of signs
(b) Comparisons of standard error of regression(se)
(c) Coefficient of determination

Model
OLS
$GDP = -442986 + 8.762EES - 36.385ESCS + 36.978ETRAN - 0.420EAD$
TLS
$LOGGDP = 4.654 + 5.317EES - 1.048E - 5ESCS + 8.527E - 6ETRAN + 1.933E - 7EAD$
(a) In terms of correctness of signs

Going by OLS, it can be seen that ESCS with value of (-36.385) and EAD have unexpected negative sign which contradicts econometric theory.
Going by the TLS, it could be deduced that all the variables have positive signs except ESCS which is by far closer to zero than that of OLS and it shows a high negative impact on GDP.
(b) Using standard error of regression (se)

OLS produced have large standard errors which yield large confident interval that makes an insignificant coefficient seems to be significant and are the variances are not constant while TLS have smaller values and constant variances. Based on the this comparison, TLS offer best improvement over the OLS.
(c) In terms of having high coefficient of determination

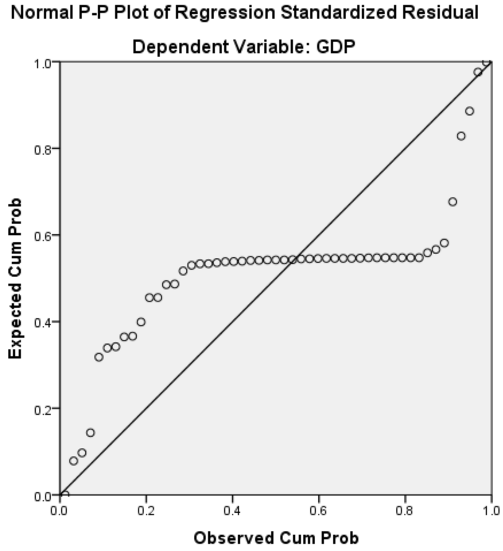
OLS have the high level of coefficient of determination at 85% which indicates a high relationship between the dependent and independent variables. This shows that 85% changes in GDP are explained by the independent variables.
TLS which happened to be the better model has a coefficient of determination of 65% , because OLS has over estimated the coefficient of determination

4.0. Conclusion and Recommendation
4.1 Conclusion
Based on this research, the following conclusion are drawn
1. The error variance of the data used in this study is heteroscedastic.
2. Log transformation of the dependent variable is good alternative to OLS when the data violate the assumption of homoscedasticity.
3. Government expenditure in Nigeria has a great impact on GDP with 64% of the changes in GDP being explained by change in government expenditure.

4.2 Recommendation
1. When the assumption of homoscedasticity is violated, log transformation of the dependent variable should be used as method of analysis.

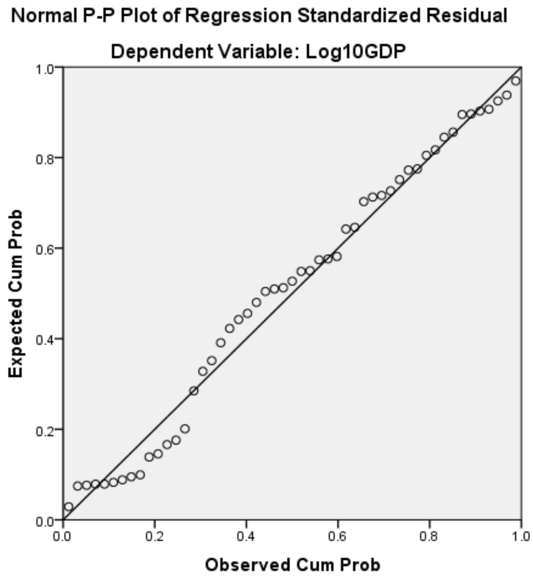
2. Nigeria government should spend much, especially on expenditure on social and economic service and expenditure on administration so as to increase the GDP.
3. Always check for the violation of the homoscedasticity in a data set, especially in cross-sectional data.

Appendix 1
Ordinary least squares (before transformation)



Appendix 2

Transformed least squares (after transformation)



Remark: In the Normal P-P plot above, the OLS error are not normally distributed and the residual did not fall approximately on the diagonal straight line. But the TLS are normally distributed and the residuals fall approximately on the diagonal straight line, an indication that TLS is better than OLS.

References

[1] Adenomon M.O (Phd) and Micheal V.A (2017) International journal of innovative finance and economics research 6(3):74-89

[2] Andrew F Hayes (2009)"using Heteroscedasticity consistent standard error estimator in OLS regression An Introduction Software. Journal of statistical computation and simulation. Vol. 76 No. 6, pp: 787-803

[3] Baser O. (2007) "modelling transformed Health Care Cost with Unknown Heteroscedasticity," Value of health. Vol. 13. No. 3. Online ISSN: 1524 – 4733. Print ISSN: 1098 - 3015

[4] CBN Annual Reports of Accounts (various Issues) 2000 – 2012.

[5] Central Bank of Nigeria (2005-2013), Annual Reports and Statement of Accounts A CBN publications. Retrieved on 20/09/2015

[6] Koenker. V v R (1981) "A Note on studentizing a test for Heteroscedasticity". Journal of applied Econometrics 3,139-43.

[7] Nnanna O.J.(2013) Financial Markets in Nigeria, Internationl Journal of Econometrics and Financial Issues. Vol. 3, No. 2, pp: 395 – 3108.

[8] Obadan M. (2006) Overview of Exchange rate Management in Nigeria from 1986 to 2006. Central Bank of Nigeria: Understanding monetary policy Series No. 8, Vol. 30, No. 3

[9] White H. (1980) "A Heteroscedasticity consist covariance matrix estimator and a Direct test of Heteroscedasticity" Open Access Library Journal, vol 4. No. 8 pp 817-818.

.