International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2021; 6(5): 01-07

www.mathsjournal.com
Received: 01-06-2021
Accepted: 03-08-2021

Biu O Emmanuel
Department of Mathematics and Statistics, University of Port Harcourt, P.M.B 5323, Rivers State, Nigeria

Nwakuya T Maureen
Department of Mathematics and Statistics, University of Port Harcourt, P.M.B 5323, Rivers State, Nigeria

Corresponding Author:
Biu O Emmanuel
Department of Mathematics and Statistics, University of Port Harcourt, P.M.B 5323, Rivers State, Nigeria

Investigating heteroscedastic disturbances in some transformed economic models while eliminating multicollinearity

Biu O Emmanuel and Nwakuya T Maureen

Abstract
When economic variables move together over time, it may cause increase/decrease in economic development. Trend factors in time series are the strongest sources of multicollinearity. The recent recession in Nigeria suggest multicollinearity among some Nigeria economic variables. This research examined the presence of multicollinearity among some economic variables using Revised Frisch's Confluence (or Bunch –map) Test and Variance Inflation Factor in some transformed models namely; Semi-logarithmic (model B), Inverse Semi-logarithmic (model C), Double logarithmic (model D) and basic linear model (model A). These models were compared to identify the best fitted using AIC as a comparison criterion. Furthermore re-parameterization of the models was done to eliminate multi-collinear variables and also show the coefficients effect of each model. In addition, the presence heteroscedastic disturbance was seen to be present in model A and model B using Golfeld-Quandt test. A parsimonious model $Y_i = f(X_3, X_4)$ of the inverse semi logarithm model C was found to be the best fitted model. We also observed that heteroscedastic models (model A and B) had very high AIC's in comparison to the homoscedastic models (model C and D).

Keywords: multicollinearity, variance inflation factor, heteroscedastic disturbance, Golfeld-Quandt test and Akaike information criterion

1. Introduction
Ordinary least square regression is arguably the most widely used method for fitting economic models. Two of the basic assumptions of linear statistical model are that the explanatory variables are not perfectly linearly correlated and that the variance of each disturbance term ε_i conditional on the chosen values of the explanatory variables is a constant. When these explanatory variables are correlated it is termed multicollinearity. It is typical (or habitual) to check for the presence of multicollinearity among the explanatory variables and heteroscedasticity of the residuals once the economic model is built. Multicollinearity and heteroscedasticity have potentially been a serious problem in the theory of econometrics. Most economic models are prone to facing these problems. In this research work we faced the problem of separating the effect of some Nigeria economic variables on an outcome variable (GDP), due to the presence of multicollinearity. This, we tried to overcome using four different models. The research applied two multicollinearity tests to confirm if there is existence of multicollinearity among the economic variables and also check for heteroscedasticity in the four models considered using Gold-Quandt test.

2. Heteroscedasticity
Heteroscedastic disturbance arises when the variance of error of the fitted model varies as opposed to linear regression assumption, where

$$Var(\varepsilon) = E(\varepsilon^2) = \sigma^2 \tag{1}$$

But in heteroscedastic situation

$$Var(\varepsilon_i) = E(\varepsilon_i{}^2) = \sigma_i{}^2 \quad (2)$$

One of the ills of this situation is that it makes the determination of the variance of the coefficients become inapplicable given that,

$$Var(\widehat{b}_i) = \frac{\sigma^2 \sum x_i}{\sum (x_i - \bar{x})^2} \quad (3)$$

but if σ^2 varies, (that is σ_i^2), it implies that σ_i^2 cannot be taken out of the summation because it will not be a finite constant but would rather tend to change with an increasing range of x values and hence rendering equation (1.3) inapplicable. The inference drawn despite presence of heteroscedastic may be misleading and disturbs the optimal properties of ordinary least squares (OLS) estimators. In this work, we tested for the presence heteroscedastic disturbance on different models using the Golfeld-Quandt test.

3. Multicollinearity

Another crucial condition for the application of least squares is that the explanatory variables are not perfectly linearly correlated that is $r_{x_i x_j} \neq 1$, when this assumption is violated it is termed multicollinearity that is the existence of a weakly perfect or not exact linearly correlated relationship among some or all the explanatory variables of a regression model. When this happen the usual least-squares analysis of the linear regression model can be made dramatically inadequate. In most cases, despite the significant of the regression, the individual coefficients are not significant which means neither contribute significantly to the model after the other one is included. But individually they contribute an awful lot. If both variables are removed from the model, the fit would be inadequate (or worse). Hence, multicollinearity is a situation of not being able to separate the effects of two or more variables on an outcome variable. If two or more variables are significantly alike, it becomes impossible to identify which of the variables accounts for variance in the response variable Ginker and Lieberman, (2017) [1].

The nature of the problem may also be illustrated geometrically as shown,

$$\boldsymbol{Y} = \hat{\boldsymbol{Y}} + \boldsymbol{\varepsilon} \quad (4)$$

Where, Y is the observation on the response, $\widehat{Y}$ is the fitted response and e_i represents the experimental error. For instance given that x_2, x_3 vectors are not perfectly collinear and they span a two-dimensional subspace in $\Re^n$ dropping a perpendicular from y to that subspace slits y into

$$\hat{Y} = X\beta = \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 \quad (5)$$

The regression vector $\widehat{Y}$ is a unique linear combination of the column vectors x_2, x_3. If it only span a one-dimensional subspace (line) in $\Re^n$ the $\widehat{Y}$ vector is still unambiguously determined by dropping a perpendicular from y to the line, but $\hat{y}$ cannot be expressed uniquely in terms of X_2 and X_3. Several techniques are frequently employed in diagnosing multicollinearity. In econometrics there are several measures of multicollinearity diagnostics that are commonly employed.

4. Methodology

The data set used in this work is an economic data collected from the Central Bank of Nigeria (CBN) Bulletin with Gross domestic product (GDP) as the dependent variable while the Administrative sector, Social sector, Economic sector and Transportation sector as the explanatory variables. We adopted four models to determine the model that best fits the data while eliminating collinear variables and heteroscedastic disturbances. Goldfeld-Quandt test was used to check for heteroscedasticity while multicollinearity was disgonised using Revised Frisch's Confluence (or Bunch –map) Test and Variance Inflation factors (VIF). The models were compared using Akaike Information Criterion (AIC).

4.1 Model Specification

The economic models considered are expressed are follows

Model A (Linear model): $GDP_i = \beta_0 + \beta_1 Admin. + \beta_2 Socio + \beta_3 Econ. + \beta_4 Trans. + e_i$ (6)

Model B (Semi log): $GDP = \beta_0 + \beta_1 \ln(Admin) + \beta_2 ln(Socio) + \beta_3 \ln(Econ.) + \beta_4 \ln(Trans.) + e_i$ (7)

Model C (Inverse semi-log). $Ln(GDP) = \beta_0 + \beta_1 Admin. + \beta_2 Socio + \beta_3 Econ. + \beta_1 Trans. + e_i$ (8)

ModelD(Double-log): $\ln(GDP) = \beta_0 + \beta_1 \ln(Admin) + \beta_2 ln(Socio) + \beta_3 \ln(Econ.) + \beta_4 \ln(Trans.) + e_i$ (9)

4.2 Diagnostics Measures of Heteroscedasticity
4.2.1 Goldfeld-Quandt Test
To use this method we assumed that the heteroscedastic variance σ_i^2 is positively related to one of the explanatory variables in the regression model. Consider the regression model with two variables:

$$Y_i = \beta_1 + \beta_2 X_i + \mu_i \tag{10}$$

Since the heteroscedastic variance σ_i^2 is positively related to X_i we have;

$$\sigma_i^2 = \sigma^2 X_i^2 \tag{11}$$

where σ^2 is a constant. The assumption postulates that the σ_i^2 is proportional to the square of the X variables this means that the larger the size of the X variables, the larger the σ_i^2. To efficiently test this, the following steps of the Goldfeld-Quandt was implemented, Gujarati and Porter (2009) [2].
Given that the Null hypothesis is Homoscedasticity.
Step 1: Order or rank the observations according to the values of X_i in order of increasing magnitude.
Step 2: where k is specified based on hypothesis rather than experiment, Omit c central observations and divide the remaining $(n-c)$ observations into two groups each of $\frac{n-c}{2}$ observations.
Step 3: Fit the OLS regressions to the first $\frac{n-c}{2}$ observations and the last $\frac{n-k}{2}$ observations and obtain the respective residual sums of squares RSS_1 and RSS_2, where RSS_1 representing the RSS from the regression corresponding to the smaller X_i values(the smaller variance group) and RSS_2 representing that from the larger X_i values(the large variance group). These

$$\frac{n-c}{2} - k \text{ or } \left(\frac{n-c-2k}{2}\right) df \tag{12}$$

where k is the number of parameter to be estimated and this includes the intercept.
Step 4: Compute the ratio

$$\lambda = \frac{RSS_2 / df}{RSS_1 / df} \tag{13}$$

4.3 Diagnostics Measures of Multicollinearity
4.3.1 Revised Frisch's Confluence (Bunch-Map) Test
The seriousness of the effect of multicollinearity seems to depend on the degree of inter-correlation of the independent variables as well as the correlation coefficient in the model. We might therefore suggest that the standard errors, partial correlation coefficients and coefficient of determination R^2, may be used for detecting multicollinearity. However, a combination of all these criteria may help in the detection of multicollinearity. These procedures are as follows:

- We regress the dependent variable on each one of the independent variables separately (i.e all the elementary regression)
- We choose the elementary regression which appears to give the most plausible result.
- We sequentially insert additional variables and examine their effect on the individual coefficients, the associated standard errors and the overall coefficient of multiple determination R^2.
- A new variable is classified as useful and retained in the model if it improves R^2 without rendering the individual coefficients unacceptable on a prior consideration. It is classified as superfluous and not included in the model if it does not improve R^2 and does not affect to any considerable extent the values of the individual coefficients; otherwise it is classified as detrimental if it affects considerably the sign of value of the coefficients.

4.3.2 Variance Inflation Factor (VIF)
Wonsuk *et al.* (2014) [5] defined variance inflation factor (VIF) as a measure of how much the variance of the estimated regression coefficient b_i is "inflated" by the existence of correlation among the predictor variables in the model. According to the author, a VIF of 1 means that there is no correlation among the i^{th} predictor and the remaining predictor variables, and hence the variance of b_i is not inflated at all. The general rule of thumb is that VIFs exceeding 10 are signs of serious multicollinearity requiring

correction. The test of multicollinearity is necessary to ascertain the independence of regression variables. The variance inflation factor for the ith suspected mediator variable is given by:

$$VIF_i = \frac{1}{1-R_{ij}^2} \quad (14)$$

where R^2 is the coefficient of determination.

$$R^2 = \frac{\text{explained variation}}{\text{total variation}} = \frac{SSR}{SST} \qquad 0 \le R^2 \le 1 \quad (15)$$

where Total sum of squares (Total Variation): $SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$ (3.32)

Regression sum of squares (Explained Variation): $SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$ (16)

4.4 Criteria for Model Selection

The Model selection criteria considered in this research is Akaike Information Criterions (AIC)

The general equation for calculating AIC given by Henry (2010) is AIC=-2* ln (Likelihood) +2*p (17)

where ln is the natural logarithm, (Likelihood) is the value of the likelihood, p is the number of parameter in the model.

AIC can also be calculated using residual sum of squares from regression: AIC= n* ln (RSS/n) +2*k (18)

where n is the number of data points, RSS is the residual sum of squares and k is the number of parameters

5. Results

Table 1: Heteroscedastic disturbance test for Model (A to D)

Model	Golfeld-Quandt test	Critical :F (n-c-2k)/2=$F_{(8,8,0.05)}$ =3.44
Model A:	625.780944	Heteroscedasticity
Model B:	664.402925	Heteroscedasticity
Model C:	0.58964135	Homoscedasticity
Model D:	5.8809E-05	Homoscedasticity

Table 1.0 show that only model A and model B were heteroscedastic.

5.1 Multicollinearity test using Revised Frisch's Confluence (or Bunch –map) Test and Variance Inflation factors

Table 2: Parameter Estimates of the Regression of GDP on Actual Economic variables Using Model A

Model	Parameter Estimates with p-values in parenthesis					R^2(%)	Durbin-Watson Statistic	VIF
	β_0	β_1	β_2	β_3	β_4			
Y_i =f(X_1)	-2795(0.52)	25.713(0.000*)				65.3	0.5739	-
Y_i=f(X_2)	3385(0.4300)	24.597 (0.000*)				57.2	0.7709	-
Y_i =f(X_3)	6470(0.201)	9.362(0.001*)				39.6	0.2966	-
Y_i =f(X_4)	-100 (0.965)	9.1681(0.000*)				87.6	1.3757	-
Y_i =f (X_1, X_2)	-2552(0..590)	24.08(0.032*)	1.81(0.860)			65.5	0.5895	7.0 7.0
Y_i =f (X_1, X_3)	-3504(0.436)	31.561(0.000*)	-3.217(0.367)			66.6	0.6044	3.6 3.6
Y_i =f (X_1, X_4)	-1489(0.573)	4.358(0.292)	8.074(0.000*)			88.1	1.1515	3.0 3.0
Y_i =f (X_2, X_3)	3711(0.382)	37.81(0.003*)	-6,572(0.214)			60.1	1.0536	6.5 6.5
Yi=f (X_2, X_4)	-847(0.717)	4.597(0.204)	8.143(0.000*)			88.4	1.3282	2.2 2.2
Y_i =f (X_3, X_4)	-1386(0.546)	2.359(0.073)	8.3069(0.00*)			89.2	1.3856	1.4 1.4
Y_i =f (X_1, X_2, X_3)	-2213(0.633)	24.02(0.030*)	14.99(0.302)	-6.526(0.18)		68.3	0.7972	7.0 12.6 6.5
Y_i =f (X_1, X_2, X_4)	-874(0756)	0.135(0.985)	4.499(0.490)	8.13(0.00*)		88.4	1.322	9.4 7.0 3.0
Y_i =f (X_1, X_3 ,X_4)	-497(0.849)	-5.178(0.463)	3.757(0.111)	9.097(0.00*)		89.5	1.7162	9.5 4.6

								3.8
Y_i =f (X_2, X_3, X_4)	-1538(0.508)	-7.137(0.398)	4.767(0.135)	9.019(0.00*)		89.6	1.5305	13.0 8.5 2.9
Y_i =f(X_1, X_2, X_3,X_4)	-878(0.746)	-3.671(0.625)	-5.751(0.524)	5.291(0.124)	9.441(0.0*)	89.7	1.753	10.5 14.5 9.5 4.3

Footnote: p-values in parenthesis where *significant at 5%.

In Table 2.0, three models have high VIF value (> 10) in one or two independent variables. However, the model with the smallest VIF is Y_i =f(X_3, X_4) after dropping the two collinear variables (X_1 and X_2) with only one significant coefficient and R^2 of 89.2%.

Table 3: Parameter Estimates of the Regression of GDP on SEMI-LOG Economic variables using Model B

Model	**Parameter Estimates with p-values in parenthesis**					**R^2 (%)**	**Durbin-Watson Statistic**	**VIF**
	β_0	β_1	β_2	β_3	β_4			
Y_i =f(X_1)	-5234(0.002*)	11837(0.000*)				52	0.1876	-
Y_i=f(X_2)	-28411(0.004*)	9211 (0.000*)				57.5	0.2933	-
Y_i =f(X_3)	-45697(0.002*)	10567(0.000*)				51.8	0.21528	-
Y_i =f(X_4)	-15531 (0.068)	5855(0.000*)				48.5	0.2254	-
Y_i =f (X_1, X_2)	-37710(0.028*)	3560(0.487)	6854(0.080*)			58.4	0.2406	5.0 5.0
Y_i =f (X_1, X_3)	-53935(0.001*)	6424(0.187)	5571(0.200)			55.6	0.2103	4.1 4.1
Y_i =f (X_1, X_4)	-43569(0.055)	8649(0.175)	1774(0.580)			52.7	0.1644	6.6 6.6
Y_i =f (X_2, X_3)	-31369(0.054)	8155(0.098)	1366(0.813)			57.6	0.26115	7.8 7.8
Yi=f (X_2, X_4)	-27713(0.006*)	7277(0.032*)	1574(0.481)			58.5	0.2630	3.6 3.6
Y_i =f (X_3, X_4)	-39880(0.006*)	6444(0.036*)	3070(0.085)			58.0	0.1968	2.2 2.2
Y_i =f (X_1, X_2, X_3)	- 38609(0.056**)	3457(0.0.520)	6505(0.242)	540(0.928)		58.4	0.2354	5.2 10.0 8.2
Y_i =f (X_1, X_2, X_4)	-33334(0.135)	2048(0.774)	6668(0.098)	966(0.756)		58.6	0.2403	9.3 5.1 6.7
Y_i =f (X_1, X_3 ,X_4)	-36259(0.102)	-1943(0.827)	7375(0.115)	3681(0.271)		58.1	0.2088	14.3 4.7 7.5
Y_i =f (X_2, X_3, X_4)	-34082(0.043*)	4468(0.500)	3021(0.624)	1960(0.416)		58.9	0.2169	14.7 8.8 4.0
Y_i =f (X_1, X_2, X_3, X_4)	-33658(0.140)	-267(0.977)	4411(0.533)	3167(0.698)	2058(0.628)	58.9	0.2178	15.6 15.9 14.6 12.0

Footnote: p-values in parenthesis where *significant at 5%.

In Table 3.0, four models have high VIF value (> 10) in one or two independent variables. Model **Y_i =f(X_3, X_4)** also has the smallest VIF after dropping two variables (X_1 and X_2) as the collinear variables with R^2 of 58.0% which is low with only one significant coefficient.

Table 4: Parameter Estimates of the Regression of GDP on LOG-SEMI Economic variables Model C

Model	**Parameter Estimates with p-values in parenthesis**					**R^2 (%)**	**Durbin-Watson Statistic**	**VIF**
	β_0	β_1	β_2	β_3	β_4			
Y_i =f (X_1)	6.7110(0.000*)	0.00225(0.000*)				84.2	1.2461	-
Y_i=f (X_2)	7.3286(0.000*)	0.00203(0.000*)				65.4	1.0637	-
Y_i =f (X_3)	7.4914(0.000*)	0.00084(0.000*)				53.98	0.4262	-
Y_i =f (X_4)	7.2484(0.000*)	0.00065(0.000*)				74.9	0.5807	-
Y_i =f (X_1, X_2)	6.6152(0.000*)	0.00289(0.000*)	-0.00071(0.201)			85.4	1.1755	7.0 7.0
Y_i =f (X_1, X_3)	6.6679(0.000*)	0.00261(0.000*)	-0.00019(0.290)			85.0	1.2956	3.6

								3.6
$Y_i = f(X_1, X_4)$	6.7537(0.000*)	0.00155(0.000*)	0.000264(0.011*)			88.4	0.8841	3.0 3.0
$Y_i = f(X_2, X_3)$	7.3324(0.000*)	0.00218(0.013*)	-0.00008(0.840)			65.4	1.1046	6.5 6.5
$Yi=f(X_2, X_4)$	7.0969(0.000*)	0.000933(0.013*)	0.000446(0.000*)			81.1	0.8543	2.2 2.2
$Y_i = f(X_3, X_4)$	7.0187(0.000*)	0.00042(0.0002*)	0.0005(0.000*)			84.2	0.8032	1.4 1.4
$Y_i = f(X_1, X_2, X_3)$	6.6188(0.000*)	0.00289(0.000*)	-0.00057(0.450)	-0.000(0.78)		85.5	1.2049	7.0 12.6 6.5
$Y_i = f(X_1, X_2, X_4)$	6.6680(0.000*)	0.00214(0.001*)	-0.00063(0.201)	0.000(0.012*)		89.2	0.7523	9.4 7.0 3.0
$Y_i = f(X_1, X_3, X_4)$	6.7559(0.000*)	0.00153(0.013*)	0.000086(0.963)	0.000(0.024*)		88.3	0.8833	9.5 4.6 3.8
$Y_i = f(X_2, X_3, X_4)$	7.0054(0.000*)	-0.000631(0.431)	0.000631(0.039*)	0.000(0.00*)		84.7	0.7736	13.0 8.5 2.9
$Y_i = f(X_1, X_2, X_3, X_4)$	6.6677(0.000*)	0.00188(0.003*)	-0.0013(0.059*)	0.0004(0.156)	0.00(0.0*)	90.3	0.6367	10.5 14.5 9.5 4.3

Footnote: p-values in parenthesis where *significant at 5%.

In Table 4.0, three models have high VIF value (> 10) in one or two independent variables. However, the best fit model is $Y_i = f(X_3, X_4)$ after dropping two variables (X_1 and X_2) as collinear variables with R^2 of 84.2% with the both variables significant at 5%.

Table 5: Parameter Estimates of the Regression of GDP on LOG-LOG Economic variables Model D

Model	**Parameter Estimates with p-values in parenthesis**					**R^2 (%)**	**Durbin-Watson Statistic**	**VIF**
	β_0	β_1	β_2	β_3	β_4			
$Y_i = f(X_1)$	1.339(0.010*)	1.209(0.00*)				91.4	0.6318	-
$Y_i = f(X_2)$	4.337(0.000*)	0.832(0.00*)				79.0	0.5379	-
$Y_i = f(X_3)$	2.679(0.002*)	0.966(0.00*)				73.1	0.2768	-
$Y_i = f(X_4)$	5.151(0.000*)	0.589(0.00*)				82.7	0.6721	-
$Y_i = f(X_1, X_2)$	1.675(0.005*)	1.087(0.00*)	0.158(0.224)			92.1	0.5916	5.0 5.0
$Y_i = f(X_1, X_3)$	1.308(0.013*)	1.105(0.00*)	0.107(0.456)			91.7	0.6463	4.1 4.1
$Y_i = f(X_1, X_4)$	1.939(0.010*)	0.991 (0.000*)	0.121 (0.240)			92.0	0.4978	6.6 6.6
$Y_i = f(X_2, X_3)$	3.871(0.000*)	0.666(0.015*)	0.215 (0.489)			79.5	0.4260	7.8 7.8
$Yi=f(X_2, X_4)$	4.497(0.000)	0.392 (0.008)	0.359 (0.001)			87.5	0.6995	3.6 3.6
$Y_i = f(X_3, X_4)$	3.464(0.000*)	0.460(0.00*)	0.396 0.000*)			90.3	0.6129	2.2 2.2
$Y_i = f(X_1, X_2, X_3)$	1.725(0.014*)	1.024(0.00*)	0.177 (0.343)			92.0	0.5888	5.2 10.0 5.2
$Y_i = f(X_1, X_2, X_4)$	2.150(0.006*)	0.855 (0.001*)	0.138 (0.293)	0.105(0.313)		92.4	0.5042	9.3 5.1 6.7
$Y_i = f(X_1, X_3, X_4)$	2.127(0.006*)	0.717 (0.020*)	0.191(0.205)	0.171 (0.122)		92.6	0.5063	14.3 4.7 7.5
$Y_i = f(X_2, X_3, X_4)$	3.272(0.000*)	-0.148(0.547)	0.580 0.018*)	0.433(0.000*)		92..4	0.5042	9.3 5.1 6.7
$Y_i = f(X_1, X_2, X_3, _4)$	2.131(0.008*)	0.720 (0.028*)	0.006 (0.981)	0.185 (0.491)	0.169(0.233)	92.6	0.5052	15.6 15.9 14.6 12.0

Footnote: p-values in parenthesis where *significant at 5%.

Table (5.0) shows that two models have high VIF value (> 10) in one or two independent of variables. However, the best fit for model is ($Y_i = f(X_3, X_4)$) after dropping the two collinear variables (X_1 and X_2) with R^2 equal to 90.3% which is slightly high and the two variable coefficients significant at 5%.

Table 6: Comparison of the best fit models [Model A to D], Using AIC Equation (17)

Model	Parameter Estimates with p-values in parenthesis β_0	β_3	β_4	R^2	RSS	AIC	VIF	Models Identified
Model A: $Y_i = f(X_3, X_4)$	-1386(0.546)	2.359(0.073**)	8.307(0.000*)	89.2	6156346432	567.5187	1.4 1.4	3
Model B: $Y_i = f(X_3, X_4)$	-39880(0.006*)	6444(0.036*)	3070(0.085**)	58.0	4033990668	556.9505	2.2 2.2	4
Model C: $Y_i = f(X_3, X_4)$	7.0187(0.000*)	0.00042(0.000*)	0.0005(0.000*)	84.2	34.519	92.53775	1.4 1.4	3
Model D: $Y_i = f(X_3, X_4)$	3.4639(0.000*)	0.4604(0.000*)	0.3960 (0.000*)	90.3	37.020	94.28646	2.2 2.2	2

Table 6.0 shows model C (inverse semi-logarithmic) with the lowest AIC as the best fitted model with $\mathbf{Y_i = f(X_3, X_4)}$ as the parsimonious model. Note should be taken that the two heteroscedastic models A and B have very high AIC in comparison to the homoscedastic models (C and D).

6. Conclusion

This research examined multicollinearity among economic variables in the presence of heteroscedasticity disturbance using linear model and some transformed models namely; Semi-logarithmic, Inverse Semi-logarithmic and Double logarithmic. Multicollinearity was detected using Variance Inflation Factor (VIF) and Bunch-map and Farrah-Glauber test while heteroscedasticity was confirmed using Gold-Quandt test. The results confirmed existence of multicollinearity among the explanatory variables and heteroscedasticity was found only in the linear model (model A) and the semi-log model (model B). The models were compared to determine the best fitted model. Furthermore re-parameterization of the models was done to show the effect of coefficients of each model and determine the most parsimonious model. The best fitted model based on AIC was seen to be the Inverse Semi-logarithmic (model C) within which after eliminating the multi-collinear variables the model $y = F(x_3+x_4)$ was found to be the most parsimonious with the two coefficients showing significant effect. It was also observed that the models with heteroscedasticity had only one significant variable in their best fitted model while the homoscedastic models showed that both variables were significant, based on that we can conclude that the presence of heteroscedasticity hinders the true effect of regression coefficients. It was also observed that the heteroscedastic models had very high AIC's in comparison to the homoscedastic models, so we can therefore say that the relative information lost by heteroscedastic models is quite high thereby increasing the prediction error of the estimators. This analysis was done using Microsoft Excel, Minitab 17 and SPSS 21 software.

7. Reference

1. Ginker T, Lieberman O. "Robustness of binary choice models to conditional heteroscedasticity". Economics Letters 2017;150:130-134.
2. Gujarati DN, Porter DC. Basic Econometrics (Fifth ed.). Boston: McGraw-Hill Irwin 2009, 400.
3. Henry GA. Comparison of Akaike Information Criterion (AIC) and Bayesian Criterion (BIC) in Selection of an Asymmetric Price Relationship. Journal of Development and Agricultural Economics 2010;2(1):001-006.
4. Holgersson HET, Shukur G. "Testing for multivariate heteroscedasticity". Journal of Statistical Computation and Simulation 2004;74(12):879.
5. Wonsuk Yoo, Robert Maryberry, Sejong Bae, Karan Singh, Qinghua (Peter) He, James W Lilard Jr. A Study of Effects of Multicollinearity in the Multivariate Analysis. International Journal of Applied Science and Technology 2014;4(5):9-19.