International Journal of Mathematics Trends and Technology
ISSN: 2231-5373/ https://doi.org/10.14445/22315373/IJMTT-V69I2P508
Volume 69 Issue 2, 61-69, February 2023

Original Article

Modelling and Detecting Multicollinearity of Some Economic Variables on Gross Domestic Product using Variance Inflation Factor

Naluba, Richard Zornata, Orumie Ukamaka Cynthia, Biu, Emmanuel Oyinebifun

[1]*Department of Mathematics/Statistics, Ignatius Ajuru University of Education, Rivers State, Nigeria*
[2,3]*Department of Mathematics and Statistics, University of Port Harcourt, Rivers State, Nigeria*

Received: 28 December 2022 Revised: 01 February 2023 Accepted: 12 February 2023 Published: 20 February 2023

Abstract *- This study finds multicollinearity between various economic variables and the gross domestic product (GDP) (Exchange Rate, Labour Force, Market Capitalization and All Shared Index). For the actual and logarithm transformation data sets, simple and multiple linear regression models are fitted between dependent and independent variables. Then, using the correlation matrix, variance inflation factor (VIF), and Eigenvalues, the presence of multicollinearity is evaluated. The outcome demonstrates that multicollinearity exists in some of the regression coefficients of the models, making the predicted coefficients on GDP irrelevant. A significant multicollinearity is indicated by VIF values greater than 5. Three independent variables (Exchange Rate, Market Capitalization, and All Shared Index) are significantly correlated with GDP, but only Labour Force is not, according to a comparison of the actual and logarithm models. However, when it came to R-square, VIF, and Akaike Information Criteria, the logarithm model outperforms the actual model (AIC). In light of this, this study suggests using the logarithm model to estimate GDP using various economic independent variables (Exchange Rate, Market Capitalization and All Shared Index).*

Keywords *- Exchange Rate, Multicollinearity, Logarithm transformation, Variance Inflation Factor.*

1. Introduction

The value of the products and services delivered by the national economy less the value of the goods and services used up in production is known as the gross domestic product (GDP), according to [1]. It is equivalent to the total of gross domestic investment, net exports of goods and services, gross private domestic investment, and personal consumption and expenditure. Economic growth, which determines a nation's degree of wealth, comprises the application of successful development tactics. It has drawn the attention and concern of economists, academics, and those who design economic policies from all over the world to the problem of the existence of multicollinearity as a result of the violation of one of the assumptions of the ordinary least square.Market capitalization promotes development regardless of whether a nation is developed, in transition, or underdeveloped.[2]

Multicollinearity is a statistical phenomenon in which there is strong correlation between two or more predictor variables in a multiple regression model[3]. In this situation, simple adjustments to the model or data can cause the regression coefficient estimates to fluctuate unexpectedly. It only affects calculations relating to dependent and independent variables employed, not the model's overall predictive capacity or dependability, at least not within the sample data themselves. In essence, multicollinearity may result in significant issues[4]

Researchers have noted repeatedly that GDP has played a significant effect in the economic development of nations and the state of the humanities from recent history to the present. The theory of econometrics may be seriously affected by the ongoing issue of the error term's variance fluctuating or by the violation of one of the multicollinearity assumptions.
There are various methods for calculating GDP [1]

The most widely used method of measuring GDP is the expenditure approach, which adds together government spending, private investment, consumption, and net exports. The income technique, which adds up all of the national accounts income that goes into GDP, is another way to measure GDP.

This study's objective is to model and identify multicollinearity utilizing the variance inflation factor of some economic variables on the GDP. The goals of this study are to fit simple and multiple linear regression models of economic variables on GDP and obtain estimates of their parameters, identify multicollinearity of economic variables on GDP using variance inflation factor (VIF), compare the coefficient of determination (R2), MSE, and AIC of the actual and logarithm transformation data model, and finally choose the most appropriate (best) model for GDP forecasting.

2. Materials and Methods

The data utilized for this study was taken from the Nigerian Central Bank's statistical bulletin and includes information on the labor force, the currency rate, market capitalization, and an aggregated measure of GDP from 2000 through 2019. The existence of multicollinearity may have an impact on the analysis of ordinary least squares (OLS) estimators from both the numerical and statistical points of view. Multicollinearity is significant to many different disciplines where linear regression models are utilized. Regression and the variance inflation factor are the methods that this study considers for multicollinearity detection (V.I.F). Here, the approach of the multicollinearity test and the standard least-squares analysis of the regression models are discussed.

Gross Domestic Product (GDP) model shall be carefully defined as;

$$GDP = f(LF, ER, MC, ASI) + \varepsilon \tag{1}$$

Such that

$$GDP_i = \beta_o + \beta_1 LF_i + \beta_2 ER_i + \beta_3 MC_i + \beta_4 ASI_i + \varepsilon_i \tag{2}$$

where $\beta_o,\ \beta_{1,}\ \beta_2,\ \beta_3,\ and\ \beta_4$ are estimated parameters.

LF = Labour Force, ER= Exchange Rate, MC= Market Capitalization, ASI= All Shared Index and $\varepsilon = Error\ term$.
The classical linear regression model for the economic variables can be written thus

$$Y_i = \beta_o + \beta_1 x_{1\mathrm{i}} + \beta_2 x_{2\mathrm{i}} + \beta_3 x_{3\mathrm{i}} + \cdots + \beta_{\mathrm{k}} x_{\mathrm{ki}} + \mu_i \tag{3}$$

$i = 1, 2, 3, \dots,$ n where $\beta_o = the\ intercept,\ \beta_1\ to\ \beta_k$=regression coefficients, μ= stochastic disturbance term, $i = i^{th}$ observation and n^{th} the size of the population.

The assumptions of OLS, which is the most popular technique for fitting or estimating the regression coefficient of a multiple linear regression model, are.

Linear in parameters for example $y = f(X, \theta) + \varepsilon$:

$$f(X, \theta) = X\beta \tag{4}$$

Random sampling
Zero conditional mean

$$\mathrm{E}(y/x) = \mathrm{E}(\beta_o + \beta_1 x + \cdots \mu/x) \tag{5}$$

$$\mathrm{E}(\mu/x) = 0,\ \mathrm{E}(\beta_o/x) + \mathrm{E}(\beta_1 x/x) + \mathrm{E}(\mu/x).$$

$$\mathrm{E}(y/x) = \hat{\beta}_o + \hat{\beta}_1 x + 0$$

Sample variance
The errors are uncorrelated
Homoscedasticityvar$(\mu/x) = \delta^2$
The errors are normally distributed
When the residuals' roughly constant variance requirement is broken or not met, multicollinearity is said to be an issue. The OLS is therefore no longer the most accurate and unbiased linear estimator of the regression coefficient.

2.1. Variance Inflation Factor (VIF)

The VIF is employed as a multicollinearity indicator in multiple regressions. Because larger amounts of VIF are known to negatively affect the results of multiple regression analyses, researchers prefer lower levels of VIF.The VIF is also defined as the ration between the variance of the OLS estimator of the original model and the variance of the model in which the variables are orthogonal [6]and [7].

In particular, if adding or removing a regressor producing a large changes in the estimates of the regression coefficients, mutlicollinearity will be indicated [8]

Tolerance and Variance Inflation Factor are two metrics that might help a researcher find multicollinearity[11].When the climatic variables exhibit mutlicollinearity, estimation of the variables coefficients using OLS may result in regression coefficients much larger than the practical situation would deem necessary (or reasonable) [10].

It should be mentioned that, according to [11], the variance of the OLS estimator for a typical regression coefficient (let's say _j) may be shown as follows;

$$\text{Var}(\hat{\beta}_i) = \frac{\delta^2}{S_{ii}(1-\mathcal{R}_i^2)} \tag{6}$$

Assume that X_i and the other explanatory variables in the model do not have a linear relationship. The variance of $\hat{\beta}_i$ will then be equal to $\frac{\delta^2}{S_{ii}}$. and $\mathcal{R}_i^2$ will be zero. By dividing this by the expression for Var($\hat{\beta}_i$), shown before, we arrive at the variance inflation factor as

$$\text{VIF}(\hat{\beta}_i) = \frac{1}{(1-\mathcal{R}_i^2)} \tag{7}$$

$$\text{Tolerance } (\hat{\beta}_i) = \frac{1}{VIF} = 1 - \mathcal{R}_i^2) \tag{8}$$

The jth diagonal element of Column matrix can be written as

$$VIF_i = \frac{1}{1-R_i^2} \tag{9}$$

where R_i^2 is the coefficient of determination obtained, when it is regressed on the remaining $p-1$ regressors.

2.2. Multicollinearity

The linear association between two or more explanatory factors through a strong linear relationship in which the effects of the dependent variables and the explanatory variables cannot be distinguished is known as multicollinearity. The only extant statistical test for the diagnosis of multicolinearity was presented by [12] and has received strong criticism [13], [14],[15] and [16].

The idea of orthogonality is frequently used to describe the issue of linear multicollinearity. If an eigenvalue is less than one, particularly when it is equal to or close to zero, then the explanatory variables are not orthogonal and the eigenvalues are equal to one. This causes the issue of linear multicollinearity.

Several statistical literature offers quantify collinearity with the most common pairwise correlation coefficient (r), [17] and [18]. [19] stated that the most useful class of indices depends on the complexity of the data set. Sometimes the per-variable-indices may indicate collinearity[18],although the variable-set indices miss it [6]and [19].
Two types of linear multicollinearity are thought to exist [20];

2.2.1. The Perfect Multicollinearity

Since it is impossible to discover the inverse of the matrix in this sort of multicollinearity, the specific information matrix (X^1X) has a determinant equal to zero ($|X^1X|) = 0$ [21], making it impossible to obtain the estimators of the general linear regression model.The analysis of the Eigen value can identify the approximate nature of the linear dependency existing between the variables [18].

2.2.2. Semi Multicollinearity

This is the research topic where the value of a certain determinant of the information matrix (X^1X) is extremely small and very near to zero, $|XX| = 0$ [23]. As a result, estimators for the parameters can be found; however, due to this particular type of multicollinearity, estimations are erroneous and estimated parameter differences are very significant.
The VIF calculates the parameter estimates for all the explanatory variables in the model and measures the inflation of those estimates.

$$\text{The VIF with a formula VIF} = \frac{1}{1-R_j^2}, \quad \text{j= 0,1,2…m} \tag{10}$$

2.3. Test for Detecting Multicollinearity

[22] also proposed a procedure for detecting multi collinearity which comprised of three tests (i.e Chi-square test, F-test and T-test). [18] measured the severity of multicollinearity by the condition number defined as $\|n\|:\|m\|$, where $\|n\|$ and $\|m\|$ are the largest and the smallest eigenvalues (respectively) of $\|X'X\|$. However, multicollinearity' is a fuzzy concept, dialectical in nature [9].

The parameter estimates are unbiased even in the presence of multicollinearity; $E(\beta_1) = \beta \; for \; all = 1,2 \ldots K.$ It is not always true to say that the presence of multicollinearity in a model causes the variance of parameter estimates to increase,

as both the numerator and the denominator of variances are typically affected by terms involving variables, meaning that the final size of these variances may not be large.
To confirm the independence of regression variables, the multicollinearity test is required. The variance inflation factor is thus[5]:

$$\text{VIF} = \frac{I}{I-R_i^2} \quad (11)$$

when R^2 is the coefficient of determination

$$R^2 = \frac{Explainedvariation}{TotalVariation} = \frac{SSR}{SST} \quad 0 \leq R^2 \leq 1 \quad (12)$$

and

$$SST = n\sum(y_1 - y)^2 \quad (13)$$

2.3.1. Akaike information Criterions (AIC)

The model selection criteria considered here is Akaike information Criterions (AIC);

$$AIC = [-2K\ In\ \pi + 2Kp] \quad (14)$$

AIC can also be calculated using residual sum of squares from regression

$$AIC = n\ Kn\ (RSS/n) + 2k \quad (15)$$

Large standard errors may make the regression coefficients appear insignificant, which may cause important variables to be omitted.

For example to test
H: β = 0, we use t-ratio as

$$t_1 = \frac{b}{\sqrt{var(b)}} \quad (16)$$

Since 't' is tiny because var (b) is large, it is more frequently accepted. Therefore, detrimental multicollinearity seeks to eliminate crucial factors.

2.4. Remedies of Multicollinearity

Numerous methods have been suggested to address the issues brought on by the presence of multicollinearity in the data. If the data is presented as a time series, then longer time series may cause one to disregard information from the past that is too recent. Increasing the size won't assist if the multicollinearity is caused by any identity or precise relationship. Drop a few collinear variables to comply with the X-full matrix's rank requirement. The variables with low t-ratio values can be removed first.

2.4.1. Use some Relevant Prior Information

One could conduct a search using some pertinent historical data regarding the regression coefficients. This could result in the specification of some coefficient estimations. The specification of some precise linear restrictions and stochastic linear restrictions are part of the larger general situation. For this, techniques like mixed regression and restricted regression can be applied. This function is served by the information's accuracy and relevancy. Such analyses depend heavily on the accuracy and usefulness of the data, but it might be difficult to guarantee this.

2.4.2. Employ Generalized Inverse

The generalized inverse can be used to get the inverse of $X^1X\ if\ rank(X^1X) = K.\ B = (X^1X)^{-1}X^1y$ can therefore be used to estimate B. Except when using inverse X^1X the estimates won't be unique in this situation. Different approaches to determining the generalized inverse may have various outcomes.

2.4.3. Use of Principal Component Regression

Principal component analysis is the methodology on which principal component regression is based. The K-explanatory variables are converted into a new set of orthogonal variables known as principal components. This technique is typically used to reduce the dimensionality of data while retaining some levels of explanatory variable variability, which is expressed by the variability in the study variable. Then, using the ordinary least square method, study variables are regressed on the set of chosen principal components.

OLS is used without any issues because all principal components are orthogonal and thus mutually independent. Once the estimates of regression coefficients for the smaller set of orthogonal variables (principal components) have been obtained, they are mathematically transformed into a new set of estimated regression coefficients that correspond to the original correlated set of variables. These updated estimated coefficients are major component estimators for the regression coefficient.

2.5. Model Selection Criteria

The information criteria are used to choose the most appropriate fitted model after the actual and log models were estimated using the ordinary least square approach.

2.5.1. Akaike Information Criteria

It often describes how bias and variance are exchanged when a model is specified for a collection of data. The AIC value of each model could be used to rank them. Thus

$$AIC = -2InL(\theta) + 2K \quad (18)$$

In this study, variance inflation factor and correlation are two techniques for multicollinearity detection (VIF).

2.6. Model Specification

The economic models considered can be written in the following ways

A. $$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon_i \quad (20)$$

B. $$In(Y)_i = \beta_0 + \beta_1 In(X_1) + \beta_2 In(X_2) + \cdots + \beta_k In(X_k) + \varepsilon_i \quad (21)$$

where

Y_i= is the observation on the response (i^th dependent variable; GDP)

β_0 = Intercept (constant) on Y axis

$\beta_1, \beta_2, \cdots, \beta_k$= regression coefficients

X_i is the controllable variable (or independent variable). $X_{1i}, X_{2i}, \cdots, X_{ki}$ = independent variables (i^{th} predictor value)

ε_i=i^{th} random error term, Note: K = 4.

X_1= Exchange Rate

X_2= Labour Force

X_3= Market Capitalization

X_4= All Share Index.

The total model to build is $2^K - 1 = 2^4 - 1 = 15$ models

3. Results

3.1. Data

The economic data used were obtained from the CBN Bulletin of the Centre Bank of Nigeria. The information acquired consists of monthly economic statistics on the exchange rate, the labor force, market capitalization, all-share index, and Nigeria's real GDP. (Shown in Appendix A.) We used the two models to decide which one would work best for VIF without multicollinearity. The two models were then put side by side.

3.2. Detecting Multicollearity on the Two Models

The section detects multicollearity, using Correlation analysis and Variance Inflation factors (VIF).

3.2.1. Model A: Relationship between the dependent (Y) and independent Variables($X_1$to X_4) (Actual values) Model A

By calculating the straightforward correlation coefficient between the variables, we first assess how bad the multicollinearity is.

Table 1. Correlation coefficient between dependent (Y) andindependent Variables (X_1-X_4)

Correlation Matrix	**Real Gross Domestic Product.Y**	**Monthly Exchange Rate.X1**	**Labour Force.X2**	**Market Capitalization.X3**	**All Share Index.X4**
Real Gross Domestic Product.Y	1				
Monthly Exchange Rate.X_1	0.553	1			
Labour Force.X_2	0.676	0.867	1		
Market Capitalization.X3	0.561	0.591	0.849	1	
All Share Index.X4	-0.060	-0.124	-0.020	0.438	1

Hence, there exist moderate multicollinaerity from the correlation result in Table 4.1

Detecting Multicollinaerity by Variance Inflation Factors (VIF): Model A

Next, we acquire the elementary regression (or each variable regression) and combine regression, presenting R2, the coefficients p-values, and VIF values to analyze the impact of multicollinaerity (using Equation 20).

Table 2. Parameter Estimates of the Regression of GDP on Actual Economic variables

Model	Parameter Estimates with p-values in parenthesis					R^2 (%)
	β_0	β_1	β_2	β_3	β_4	
$Y_i = f(X_1)$ VIF	7960 (0.000**)	26.52 (0.000**) 1.00				30.56
$Y_i = f(X_2)$ VIF	-1107 (0.000**)	4.85×10^{-4} (0.000**) 1.00				45.66
$Y_i = f(X_3)$ VIF	7771 (0.000**)	0.6881 (0.000**) 1.00				31.45
$Y_i = f(X_3)$ VIF	14722 (0.000**)	-0.0316 (0.391) 1.00				0.36
$Y_i = f(X_1, X_2)$ VIF	-13958 (0..000**)	-6.36 (0.194) 4.04	5.69×10^{-4} (0.000**) 4.04			46.11
$Y_i = f(X_1, X_3)$ VIF	6351 (0.000**)	16.32 (0.000**) 1.54	0.4415 (0.000**) 1.54			38.99
$Y_i = f(X_1, X_4)$ VIF	7818 (0.000**)	26.56 (0.000**) 1.02	0.00442 (0.891) 1.02			30.57
$Y_i = f(X_2, X_3)$ VIF	-1204 (0.000**)	5.14×10^{-4} (0.000**) 3.59	-0.058 (0.634) 3.59			45.72
$Y_i = f(X_2, X_4)$ VIF	-10277 (0.000**)	4.85×10^{-4} (0.000**) 1.00	-0.0247 (0.364) 1.0			45.88
$Y_i = f(X_3, X_4)$ VIF	-12135 (0.000**)	0.8912 (0.000**) 1.24	-0.1978 (0.000**) 1.24			43.02
$Y_i = f(X_1, X_2, X_3)$ VIF	-19613 (0.000**)	-11.29 (0.061) **5.85**	7.36×10^{-4} (0.000**) **13.68**	-0.209 (0.178) **5.20**		46.66
$Y_i = f(X_1, X_2, X_4)$ VIF	-13462 (0.000**)	-7.79 (0.128) 4.24	5.85×10^{-4} (0.000**) 4.17	-0.0328 (0.225) 1.05		46.50
$Y_i = f(X_1, X_3, X_4)$ VIF	**10.559** **(0.000**)**	**7.85** **(0.035**)** **2.13**	**0.731** **(0.000**)** **2.59**	**-0.1594** **(0.000**)** **1.71**		**44.28**
$Y_i = f(X_2, X_3, X_4)$ VIF	-7016 (0.232)	4.13×10^{-4} (0.001**) **11.20**	0.142 (0.551) **13.85**	-0.0521 (0.330) 3.85		45.98
$Y_i = f(X_1, X_2, X_3, X_4)$ VIF	-22059 (0.045**)	-12.49 (0.106) **9.61**	7.92×10^{-4} (0.003**) **50.65**	-0.290 (0.416) **31.43**	0.0170 (0.803) **6.34**	46.68

Here, p-values in parenthesis where **significant at 5%. The bold model has lowest VIF (best model). Note when VIF is greater than 5 it implies that multicollinearity is presence.

Three models in Table 2 have VIF values that are above 5 in one or more independent of the regression variables, which is a sign of significant multicollinearity. However, after removing one variable (X2), the bold model ($Y_i = f(X_1, X_3, X_4)$) with a moderate R2 of 44.28% is the best fit for the model. Moreover, the significance level for the three variable coefficients is 5%. The three variables' respective VIF values are 2.13, 2.59, and 1.71, which are all less than 5.

3.2.2. Model B: Relationship between the independent Variables (log-log) Model B

Similarity, we calculated each step in section 3.2.1 above, and by calculating the simple correlation coefficient between the variables, we assess the severity of multicollinearity (using Equation 21).

After using the logarithm to modify the data sets, we set the VIF value to 5. The fact that two models in Table (4) have excessive VIF values (>5) for one or more independent regression variables indicates substantial multicollinearity. However, after removing one variable (X2), the bold model ($Y_i = f(X_1, X_3, X_4)$) with a moderate R2 of 75.48% is the best match for the model. Moreover, the significance level for the three variable coefficients is 5%. The three variables' VIF values, which are each less than 5, are 1.90, 2.45, and 1.72 for $Y_i = f(X_1, X_3, X_4)$.

The identical model, which has three variables (Exchange Rate, Market Capitalization, and All Share Index) and four parameters, was found in Tables 2 and 4. This finding indicates that, while the labor force has a little impact on the Nigerian GDP, the exchange rate, market capitalization, and all-share index do.
We fit the identified model below with its fitted graphs for both data sets using Gretl statistical software (actual and logarithm transformation data sets).

Model A: OLS, using observations 2004:01-2020:12 (T = 204)

Table 3. Dependent variable: REAL_GROSS_DOME

	Coefficient	***Std. Error***	***t-ratio***	***p-value***	
const	10559	1179.06	8.9554	<0.00001	***
MonthlyExchang	7.85317	3.69106	2.1276	0.03459	**
Market_Capitali	0.731018	0.10421	7.0149	<0.00001	***
All_Share_Index	-0.157368	0.0361223	-4.3565	0.00002	***

Mean dependent var	13743.87		S.D. dependent var	4905.486
Sum squared resid	2.72e+09		S.E. of regression	3689.152
R-squared	0.442785		Adjusted R-squared	0.434427
F(3, 200)	52.97591		P-value(F)	3.03e-25
Log-likelihood	-1962.927		Akaike criterion	3933.853
Schwarz criterion	3947.126		Hannan-Quinn	3939.222
Rho	0.215130		Durbin-Watson	1.564966

Model 2: OLS, using observations 2004:01-2020:12 (T = 204)

Table 4. Dependent variable: l_REAL_GROSS_DO

	Coefficient	***Std. Error***	***t-ratio***	***p-value***	
const	9.37611	0.500585	18.7303	<0.00001	***
l_Monthly_Excha	0.16532	0.0346237	4.7748	<0.00001	***
l_Market_Capita	0.382698	0.0264043	14.4938	<0.00001	***
l_All_Share_Ind	-0.405839	0.0518694	-7.8243	<0.00001	***

Mean dependent var	9.480143		S.D. dependent var	0.304826
Sum squared resid	4.625413		S.E. of regression	0.152076
R-squared	0.754783		Adjusted R-squared	0.751105
F(3, 200)	205.2018		P-value(F)	8.89e-61
Log-likelihood	96.76509		Akaike criterion	-185.5302
Schwarz criterion	-172.2577		Hannan-Quinn	-180.1612
Rho	0.562689		Durbin-Watson	0.875805

Table 5. Summaries of the identified Model A and B

	Model	Parameter Estimates with p-values in parenthesis					R^2	AIC
		β_0	β_1	β_2	β_3	β_4		
Model A: Actual data sets.	$Y_i = f(X_1, X_3, X_4)$	10.559 (0.000**)	7.85 (0.035**)	0.731 (0.000**)	-0.1594 (0.000**)		44.28	
	VIF		2.13	2.59	1.71			3933.853
Model B: Logarithm Transformation data sets.	$Y_i = f(X_1, X_3, X_4)$	9.376 (0.000**)	0.1659 (0.000**)	0.3827 (0.000**)	-0.4059 (0.000**)		75.48	
	VIF		1.90	2.45	1.72			-185.5302

As a result, using VIF, Models A and B both identified the same model without multicollinaerity. The best model that fit the data set has three variables (Exchange Rate, Market Capitalization, and All Share Index), along with four parameters.

5. Summary Conclusion and Recommendations

5.1. Summary

In this study, the presence of multicollinearity and the logarithm-logarithm transformation of the economic variables were investigated. The Variance Inflation Factor (VIF) test for multicollinearity and the correlation coefficient matrix are the approaches used. The appropriate model without Variance Inflation Factor was determined by computing the two models and comparing them. Minitab 20, Gretl 21, and Microsoft Excel are the statistical programs used.
Finally, the variance inflation factor was used to identify the regression model's parameters, coefficient of determination, and multicollinearity (VIF).

Additionally, we contrast the results of the fitted models and pinpoint the factors that significantly affect GDP. Utilizing the model adequacy criterion technique (i.e. AIC) and R-Square, the model with three variables and four parameters was determined to be the best model. Consequently, the model that best captures the relationship between the dependent variable and the independent variables (Exchange Rate, Market Capitalization, and All Share Index) (Nigeria Real Gross Domestic Product). The work's goals, which included fitting a regression model, determining the coefficient of determination, and identifying multicollinearity using variance inflation factors, were ultimately confirmed.

6. Conclusion

The Variance Inflation Factor, a multicollinearity test, has been used to compare theperformance of the various models and calculated coefficients.

The best model for the observation under consideration according to the research was one with three variables and four parameters but without the influence of VIF.

References

[1] Bureau of Economic Analysis, Measuring the Economy: A Primer on GDP and The National Income and Product Accounts. [Online]. Available: https://www.bea.gov/resources/methodologies/measuring-the-economy

[2] Elena Goldman, and Xiangjin Shen, *Analysis of Asymmetric GARCH Volatility Models with Application to Margin Measurement*, Bank of Canada.

[3] David A. Belsley, "Collinearity and Forecasting," *Journal of Forecasting*, vol. 3, no. 2, pp. 183-196, 1984. *Crossref*, https://doi.org/10.1002/for.3980030206

[4] Alfred DeMaris, *Regression with Social Data: Modeling Continuous and Limited Response Variables*, John Wiley and Sons, Mathematics, 2004.

[5] Stephen D. Simon, and James P. Lesage, "The Impact of Collinearity Involving the Intercept Term on the Numerical Accuracy of Regression," *Computer Science in Economics and Management*, vol. 1, pp. 137–152, 1988. *Crossref*, https://doi.org/10.1007/BF00427160

[6] Johnston, J., M´etodos de Econometr´ıa. Vicens Vives, 1984.

[7] Robert M. O'brien, "A Caution Regarding Rules of Thumb for Variance Inflation Factors," *Quality & Quantity*, vol. 41, pp. 673–690, 2007. *Crossref*, https://doi.org/10.1007/s11135-006-9018-6

[8] G.W. Steward, "Collinearity and Least Squares Regression," *Statistical Science*, vol. 2, no. 1, pp. 68–84, 1987. *Crossref*, https://doi.org/10.1214/ss/1177013439

[9] Dsn Liao, *Collinearity Diagnostics for Complex Survey Data*, Unpublished doctoral dissertation, University of Maryland, 2010.

[10] N.R. Draper, and H. Smith, *Applied Regression Analysis*, 2nd Edition Wiley, New York, 1981.

[11] Wooldridge, J. (2000) Introductory Econometrics: A Modern Approach. South-Western College Publishing, New York, 483-484, 2000.
[12] Donald E. Farrar, and Robert R. Glauber, “Multicollinearity in Regression Analysis: The Problem Revisited,” *The Review of Economic and Statistics*, vol. 49, no. 1, pp. 92–107, 1967. *Crossref,* https://doi.org/10.2307/1937887
[13] Yoel Haitovsky, “Multicollinearity in Regression Analysis: Comment,” *The Review of Economics and Statistics*, vol. 51, no. 4, pp. 486–489, 1969. *Crossref,* https://doi.org/10.2307/1926450
[14] T.K. Kumar, “Multicollinearity in Regression Analysis,” *The Review of Economics and Statistics*, vol. 57, no. 3, pp. 365–366, 1975.
[15] C. Robert Wichers, “The Detection of Multicollinearity: A Comment,” *The Review of Economics and Statistics*, vol. 57, no. 3, pp. 366-368, 1975. *Crossref,* https://doi.org/10.2307/1923926
[16] John O’Hagan, and Brendan McCabe, “Tests for the Severity of Multicolinearity in Regression Analysis: A Comment,” *The Review of Economics and Statistics*, vol. 57, no. 3, pp. 368–370, 1975. *Crossref,* https://doi.org/10.2307/1923927
[17] John Fox, and Georges Monette, “Generalized Collinearity Diagnostics,” *Journal of the American Statistical Association*, vol. 87, no. 417, pp. 178–183, 1992.
[18] David A. Belsley, *Conditioning Diagnostics, Collinearity and Weak Data in Regression*, New York: John Wiley & Sons, Inc., 1991.
[19] Roman Salmeron Gomez, Catalina Garcia Garcia, and Jose Garcia Perez, “Comment on “A Note on Collinearity Diagnostics and Centering” by Velilla (2018),” *The American Statistician*, vol. 74, no. 1, pp. 68-71, 2020. *Crossref,* https://doi.org/10.1080/00031305.2019.1635527
[20] Russell Davidson, and James G. MacKinnon, *Estimation and Inference in Econometrics*, New York, 1993.
[21] Andrew C. Harvey, *The Econometric Analysis of Economic Time Series,* Second Edition Cambridge, 1990.
[22] Harald Cramér, *Mathematical Methods of Statistics,* Princeton, NJ: Princeton University Press, 1946.
[23] Christian Gourieroux, and Alian Monfort, *Simulation-Based Econometric Methods*, Oxford: Oxford University Press, 1996.