AUTOCORRELATION DETERMINATION IN SEVERAL REGRESSION MODELS' RESIDUAL WITH NIGERIA GROSS DOMESTIC PRODUCT (GDP) DATA

[1]NKPORDEE, Lekia, [2]ABOLAJI Emmanuel, Tychicus [3]OGOLO, Ibinabo Magnus

[1]Department of Mathematics and Statistics,
Kampala International University, Kansanga, Kampala, Uganda

[2]Department of Mathematics and Statistics,
Ignatius Ajuru University of Education, Rivers State Nigeria

[3]School of Foundation Studies,
Rivers State College of Health Science and Management Technology, Rumueme, Port Harcourt, Nigeria
Corresponding Author Email: lekia.nkpordee@kiu.ac.ug

Abstract

Six regression models have been utilized in this particular research to check for the existence of correlation in model residual using secondary data on Nigeria Gross Domestic Product at Current Basic Prices. The appropriate model for this investigation is the Sine design that has the very least mean square error, the very least standard deviation, smallest AIC, smallest BIC and HQC, although the particular design contains the greatest R^2 valued while the Cosine design has the greatest Durbin Watson statistical value, but many parameters came up because of transformation effect. It was concluded from this study that the sine transformation technique could be more effective in solving the issue of autocorrelation in real life data. Therefore, it was suggested that much more comprehensive work be performed making use of other transformation methods not mentioned or even used in this particular current study to determine a far more appropriate design to be used in the research or for dealing with autocorrelation issues in the data.

Keywords: *Autocorrelation, Residual, Regression Analysis, Statistical Model, Gross Domestic Product*

INTRODUCTION

Regression plus time series are helpful studies of phenomena which have to understand {their direction} in the long term, which believe a fit model's suitable data after managing autocorrelation. Regression analysis describes a technique of mathematically learning which variables may also impact. The significance of regression analysis is inside the reality it provides a good statistical method which allows in an enterprise to analyze the relationship between two or maybe more variables. Shalabh (2017) thought on the analysis of the populace suggests the usage of ratio technique and analysed {its properties} within the presence of measurement blunders.

The benefits of regression analysis are manifold: The regression method of forecasting can be used for, as the call implies, {forecasting and locating} the causal connection between variables. An important related, nearly the same, the concept contains the blessings of linear regression, and that is the procedure for modeling the worth of just one variable at the fee(s) of a single. Realizing the benefits of regression analysis, the rewards of linear regression, in addition to the advantages of {regression evaluation and the regression and regression evaluation} approach of forecasting might help a small company, and really any company, obtain a much more info of the variables (or maybe elements) that might change {its achievement} within the coming years, months, and weeks into the world.

The significance of regression examination would be that {its miles} exactly about info: info approach numbers and figures which outline {your commercial enterprise}. The advantages of regression analysis are that it might enable you to crunch the numbers to assist you make much better choices for your business enterprise already and into destiny.

Autocorrelation refers to the connection not between 2 or maybe more variables but 2 successive random variables. Autocorrelation describes the level of correlation between the values of the exact same variables across distinct observations in the information. The idea of autocorrelation is usually talked about in the context of time sequence data where observations happen at {different points} in time (e.g., air temperature measured on various days of the month). If the heat values that occurred closer together in time are, in reality, much more similar compared to the temperature values that occurred farther apart in time, the information will be autocorrelated. Nevertheless, autocorrelation also occurs in cross sectional details when the observations are connected in other way. In a survey, for example, one might expect folks from nearby geographic locations to offer much more common responses to one another than individuals that are much more geographically distant. Likewise, pupils from exactly the same category may perform far more similarly to one another than pupils from numerous classes. Thus, autocorrelation is able to occur whether observations are dependent in areas other than time. Successive values of the random variable μ are independent i.e.

$$Cov(U_i, U_j) = E\left[\left(U_i - E(U_i)\right)\left(U_j - E(U_j)\right)\right] \tag{1}$$

$$= E\left[U_i, U_j\right] = 0 \tag{2}$$

If the assumption above is violated, we have autocorrelation of successive values. There are two major approaches use to detect or test for the presence of autocorrelation, they are:

a. Von Newman Test Ratio and
b. Durbin Watson Test.

Resulting factors of autocorrelation in a specified data consist of the following:

1. Even when the residuals happen to be serially correlated, The OLS estimations are statistically impartial.
2. With the autocorrelation coefficient of the disturbance term, the OLS - variance of parameter estimates is likely to be larger when compared with other econometric variances.
3. Again, the variance of the disturbance term in the presence of autocorrelation, particularly in the situation of positive autocorrelation, may be underestimated.
4. If the OLS estimation is efficient.

One goal of this analysis would be to check for the existence of autocorrelation in a number of regression models with transformation variables. The study therefore uses Durbin Watson's test methods in GDP data from Nigeria. Data was gathered from the World Bank Databased and Statista, 2022, annual reports during (1960 - 2021). This study is designed in order to determine the best technique for testing the existence of autocorrelation in a number of regression models and to identify the model with the lowest disturbance error by comparing estimates of these techniques, even though it estimates the error.

Problem Statement

Autocorrelation is able to induce difficulties in typical analyses (such as regular least squares regression) that believe independence of observations. In a regression analysis, autocorrelation of the regression residuals are able to happen if the unit is incorrectly specified. For instance, in case you're trying to model a linear connection although observed relationship is non-linear (i.e., it follows a curved or maybe U-shaped function), subsequently the residuals is autocorrelated.

In ordinary least squares (OLS), the adequacy of a product specification may be examined in part by creating whether there's autocorrelation of the regression residuals. Problematic autocorrelation of the mistakes, which are unobserved, may typically be recognized since it makes autocorrelation within the observable residuals. Autocorrelation of the mistakes violates the typical least squares presumption that the mistake terms are uncorrelated, which means that the Gauss Markov theorem doesn't use, which OLS estimators are not the best Linear Unbiased Estimators (BLUE). While it doesn't bias the OLS coefficient estimates, the regular errors often be underestimated (and the t scores overestimated) when the autocorrelations of the mistakes at minimal lags are positive.

Among the main problems with minimum square method with precious time sequence is good chance that the ε_i, s isn’t impartial. It's pointed out it's correlation of the ε_i, s without of y, s that will be stayed away from. It's feasible that if the x's and y's are both correlations in time the mistakes is fairly uncorrelated, therefore, inconsiderable amount of research been dedicated to this issue. Based upon this particular observed gap, this particular analysis attempts in order to determine the ideal way of evaluating the existence of autocorrelation in many regression models and also to identify the product with the very least disturbance error.

Aim and Objectives of the Study

The aim of this study is to test for autocorrelation in several regression models residual using the Nigeria gross domestic product (GDP) data. Specifically, the objectives of the study are to:

i. Obtain the parameter estimates for each of the regression models under consideration.

ii. Transform the actual data using (Natural logarithms, square roots, Sine, Cosine and Tangent).

iii. Estimate the residuals for each of the transform data.

iv. Obtain the values for the model selection criteria (i.e. AIC, BIC and HQC).

v. Identify a suitable model with the least disturbance error for the Nigeria GDP data.

LITERATURE REVIEW

The study described statistical methods for analyzing the interactions among variables. Believe that the first independent variable X, where's the additional dependent variable Y is governed by a few random variations. This's viewed as regression analysis. The newspaper unveiled the linear design under the assumption that its error terms might be represented by uncorrelated random variables with mean of zero and constant variance. However, the paper should significance tests for hypothesized values of its parameters α, β and also estimates of the anxiety in expected values of the dependent variable for anew value of predictor (Dorland, 2018).

Autocorrelation, likewise known as serial correlation occurs when residuals from adjacent dimensions inside a time series aren't outside of each other. Is the correlation of any signal with a delayed copy of itself as a characteristic of delay? Informally, it's the similarity between observations as a characteristic of the precious time lag between them. The evaluation of autocorrelation is a mathematical tool for locating repeating patterns, like the existence associated with a regular signal obscured by sound, or perhaps determining the missing basic frequency in signal implied by its harmonic frequencies. It's often-used in signal processing for analyzing functions or series of values, like time domain. Unit root processes, trend stationary processes, autoregressive processes, plus moving average processes are particular types of tasks with autocorrelation.

In data, the autocorrelation of a random practice in the Pearson correlation between values of the system at different times, as a characteristic of the 2 situations and of the precious time lag (Bisgaard, & Kulahci, 2015b).

In data, the Durbin Watson statistic is an exam statistic Utilized to identify the presence of autocorrelation (a connection between values separated from one another by a certain time lag) in the residuals (prediction errors) originating from a regression analysis. It's named after James Durbin and Jeoffrey Watson. The tiny sample distribution of this ratio was derived by Neumann (1941).

Durbin and Watson (1950, 1951) put on the statistic on the residuals from minimum square regression, and also created bounds assessments for the null hypothesis which the mistakes are serially uncorrelated contrary to the answer that they stick to a very first order autoregressive process.

Later, John Denis Sargan and Bhargave developed many von Neumann-Durbin-Watson type test statistics for the null hypothesis which the mistakes on a regression model adhere to a procedure and have a device root against the alternative hypothesis that the mistakes stick to stationary very first order auto regression ({Sargan & Bhargava, 2018). Be aware that the distribution of the test statistic doesn't rely on the estimated regression coefficients and the variance of the mistakes, ({Chatterjee & Simonoff, 2013).

Stanley and Edwin (2016) used Box Jenkins modeling procedure for determine an ARIMA model and forecasting. They used information of malaria case from Ministry of Health (Kabwe District) Zambia for the period (2009 2013) for age one to five years, The results showed that a suitable design is merely an ARIMA (one, 0,0) as well as the forecasted malaria for {January and February}, 2014 are 220 and 265, respectively (Stanley & Edwin, 2016).

Sadia AbdEkareem (2012) using Box and Jenkins Method (Identification, Diagnostics Checking of version, forecasting) to discover the ideal forecasting design on the variety of patient with malignant tumors within an bar province with the month information for the time (2006 2010).The result of information analysis demonstrate that the suitable and proper design is integrated Auto regressive type of purchase (two) ARIMA (two, one, zero).

AbuElgasim and Khalda (2018) conducted a study on Using Transformation of Variable for treating Autocorrelation in Simple Regression (First Cochrane-Occutt and Difference methods). The target of the research was examining it's occasionally required to make use of transformation of adjustable for treating autocorrelation of regression or maybe time series model. Different there are numerous methods have been utilized for this purpose. This paper has the original distinction technique and also the Cochrane Orcutt method in autocorrelation of malaria data in Khartoum State, viewed in outpatient clinics. The information collected from the Federal Ministry of Health, National Health Information Centre, and yearly reports during (2001 - 2015). By calculating the estimates for each technique and mistakes, the paper demonstrated that the approach to variations is much better in the healing of auto correlation, and you will find boundaries of negatives error, because minimizing these mistakes from minimum square method.

METHODS AND MATERIALS

Nature and Source of Study Data: The collected data used for this study was a secondary statistical data extracted from survey information performed in Nigeria on Gross Domestic Product at Current Basic Prices matching the data length of 62 from 1960 to 2021. The data were sourced from World Bank Databased and Statista, 2022. Appendix A contains the information.

Data Analysis Tools: The parameters that make up the models are obtained using the Gretel 19 and SPSS 23 statistical software. The researcher used this software to estimate the parameters for different regression models to make record evaluation easier. The Microsoft Excel 16 program was used to transform the data under study.

Model Specification

The hypothesis below was tested:

H_0: There exists the presence of autocorrelation in the regression models obtained from the data.

H_1: There is no presence of autocorrelation in the regression models obtained from the data.

The regression models to be considered are:

1. **The actual model:**

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i \quad (3)$$

However, the models after transformation are

2. **Natural Log model:**

$$\ln Y_t = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + e_i \quad (4)$$

3. **Square root model:**

$$\sqrt{Y_t} = \beta_0 + \beta_1 \sqrt{X_1} + \beta_2 \sqrt{X_2} + e_i \quad (5)$$

4. **Sine Model**

$$\mathrm{Sin}(Y_t) = \beta_0 + \beta_1 \mathrm{Sin}(X_1) + \beta_2 \mathrm{Sin}(X_2) + e_i \quad (6)$$

5. **Cosine Model**

$$\mathrm{Cos}(Y_t) = \beta_0 + \beta_1 \mathrm{Cos}(X_1) + \beta_2 \mathrm{Cos}(X_2) + e_i \quad (7)$$

6. **Tangent Model**

$$\mathrm{Tan}(Y_t) = \beta_0 + \beta_1 \mathrm{Tan}(X_1) + \beta_2 \mathrm{Tan}(X_2) + e_i \quad (8)$$

Where

$t = 1 \ldots n$

Y_t is Nigeria GDP; the dependent variable, endogenous variable.

$X_1 + X_2 + \ldots + X_K$ are (Inflation rate and Oil price per barrel) the independent variables, explanatory variables (k=2)

$\beta_0, \beta_1, \ldots, \beta_K$ are coefficients of the independent variable effect on the dependent variable.

e_i is the Error term.

n is the sample size and k is the number of independent variables (p=2).

Durbin-Watson statistic

Tests for existence of correlation in residuals by finding out if the correlation between two (2) adjacent error terms are 0. The test assumes that errors are caused by an autoregressive process of first order. Minitab assumes that observations are in meaningful order, such as in the time order. The formula is:

$$d = \frac{\sum_{i=1}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i{}^2} \tag{9}$$

Notation

e_i = i^{th} residual

e_{i-1} = residual for the previous observation

n = number of observations

where T is the variety of observations. Notice that if you have a long sample, subsequently this could be linearly mapped to the Pearson correlation of the time-series information because of its lags. Since d is around the same as 2(1' r), where r will be the sample autocorrelation on the residuals. d = two indicates no autocorrelation. The importance of d always lies between zero and four. If the Durbin–Watson statistic is considerably less than two, there's proof of good serial correlation. As an approximate rule of thumb, if Durbin–Watson is under 1.0, there could be cause for alarm. Small values of d indicate successive error terms are, on average, near in value to each other, or even positively correlated. If d > 2, successive error terms are, on average, a lot different in value from each other, i.e., negatively correlated. In regressions, this may indicate an underestimation of the amount of statistical significance.

To test for positive autocorrelation at significance α, the test statistic d is compared to lower and upper critical values (dL, α & dU, α):

- If d < dL, α, there is statistical evidence that the error terms are positively autocorrelated.
- If d > dU, α, there is no statistical evidence that the error terms are positively auto correlated.
- If dL, α < d < dU, α, the test is inconclusive.

A positive serial correlation is a serial correlation where a good error for one observation raises the chance of an optimistic error for another observation. In order to check for negative autocorrelation with significance, the test statistic (4' d) is compared with lower and upper critical values (dL, α and dU, α):

- If (4 − d) < dL, α, there is statistical evidence that the error terms are negatively auto correlated.
- If (4 − d) > dU, α, there is no statistical evidence that the error terms are negatively auto correlated.
- If dL, α < (4 − d) < dU, α, the test is inconclusive.

Negative serial correlation means that a good mistake for one observation enhances the possibility of a bad mistake for one more observation along with a bad error for one observation enhances the risks of an optimistic error for another.

The critical values, dL, α & dU, α, differ by amount of significance (α), the quantity of observations, and also the number of predictors in the regression equation. Their derivation is complicated - statisticians usually get them from the appendices of statistical texts.

The following assumptions need to be satisfied:

1. The regression model includes a constant
2. Autocorrelation is assumed to be of first-order only.

Model Selection Criteria

To find the right function, the model selection Criteria is used. The best model minimizes the criterion. A number of criteria for choosing among models have been created recently; including the residual sum of squares (ESS) multiplied by a penalty factor depending on the complexity of the model. Here are some examples of these criteria:

Akaike Information Criteria (AIC)

Akaike (1974) also developed another procedure known as the Akaike Information Criteria. Below is the form of the statistic.

$$\text{AIC}=\left(\frac{ESS}{T}\right)e\left(\frac{2k}{T}\right) \tag{10}$$

The value of AIC decreases when some variable is dropped.

Hunnan and Quinn (HQ) criterion

Hunnan and Quinn (1979) developed a procedure which is known as HQ criteria. Statistic of the procedure can be represented as

$$HQ=\left(\frac{ESS}{T}\right)(InT)\left(\frac{2K}{T}\right) \tag{11}$$

The value of HQ will decrease provided there are at least 16 observations.

SCHWARZ Criterion

Craven and Wahba (1978) developed a procedure which is known as SCHWARZ (BIC) Criteria. The form of this procedure is represented as

$$\text{SCHWARZ}=\left(\frac{ESS}{T}\right)T^{\mathrm{K/T}} \tag{12}$$

The value of SCHWARZ will also decrease provided there are at least 8 observations.

SCHWARZ Bayesian Information Criterion (BIC)

The SCHWARZ (BIC) criteria method was developed by Craven and Wahba in 1978. Below is a description of this method's form:

$$BIC = n\ln\left[\frac{SSE}{n}\right] + k\ln(n) \tag{13}$$

If there are at least eight observations, the fee for SCHWARZ (BIC) will also decrease.

Coefficient of Determination (R^2)

R^2 is one of the crucial statistical variables used to make decisions and draw conclusions from data. It is a method for calculating the proportion of a variable's results that are greater than those of other variables. The structure of this procedure is shown as:

$$R^2 = \frac{SSR}{SST} \tag{14}$$

Where; SSR= Sum of square Residual

SST= Sum of Square Total

Models Accuracy Measures

The model having the smallest accuracy (MAPE, MAE, and MSE) will be used to find the best fit. An excellent model minimizes the criterion. A number of criteria for picking some models have been suggested recently, which includes Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Mean Squared Error (MSE). The mean square error (MSE) is used for the purposes of this study.

MSE (Mean Square Error): No matter what the model is, the denominator n is utilized to figure out this. This allows comparing MSE values between models. MSE is thus more sensitive to the biggest forecast error than MAE.

$$MSE = \frac{\sum_{t=1}^{n}|(Y_t - \hat{Y}_t)|^2}{n} \tag{15}$$

Where Y_t equals the actual value at time *t*, $\hat{Y}_t$ equals the fitted value, and *n* equals the number of observations.

RESULTS

Based on regression analysis modeling, the Durbin-Watson statistic is used to test for the existence of autocorrelation in residuals by figuring out if the correlation between 2 adjacent error terms is 0. The test assumes that errors are caused by an autoregressive process of first order. This method was used with the help of SPSS Statistical software and Gretl to check for the existence of autocorrelation in a number of regression models. Then the Durbin Watson statistic results along with other relevant model selection criteria have been summarised in Table 1 - 6 and Table 7 displays the six models parameter estimates with their model selection criteria and R - square utilizing the Ordinary Least Square Method.

The Actual Model

Table 1: Estimated parameters for The Actual Model

Variable	*Coefficient*	*Std. Error*	*t-ratio*	*p-value*	
const	7.61108	1.19493	6.369	<0.0001	***
Inflation	−0.0541218	0.0433743	−1.248	0.2170	
Oil Price	−0.0244247	0.0217884	−1.121	0.2668	

Mean dependent var	5.997478	S.D. dependent var	5.046911
Sum squared resid	1489.261	S.E. of regression	5.024114
R-squared	0.041505	Adjusted R-squared	0.009014
F(2, 59)	1.277425	P-value(F)	0.286350
Log-likelihood	−186.5201	Akaike criterion	379.0403
Schwarz criterion	385.4217	Hannan-Quinn	381.5458

Sources: Author's Work (2022)

Table 1 shows that the series were not significant at 5% level of significance.

Therefore, **the required actual model is:**

$$\widehat{Y}_t = 7.61108 - 0.0541218X_1 - 0.0244247X_2$$

Natural Log model

Table 2: Estimated parameters for the Natural Log model

Variable	*Coefficient*	*Std. Error*	*t-ratio*	*p-value*	
const	1.52663	0.470323	3.246	0.0019	***
LNI	−0.0375654	0.190057	−0.1977	0.8440	
LNO	−0.0304148	0.119640	−0.2542	0.8002	

Mean dependent var	1.351910	S.D. dependent var	1.183771
Sum squared resid	85.23021	S.E. of regression	1.201907

R-squared	0.002924	Adjusted R-squared	-0.030875
F(2, 59)	0.086506	P-value(F)	0.917246
Log-likelihood	−97.83906	Akaike criterion	201.6781
Schwarz criterion	208.0595	Hannan-Quinn	204.1836

Sources: Author's Work (2022)
Foot note: SRI = natural log of inflation, LNO = natural log of oil price

Table 2 shows that the series were not significant at 5% level of significance.

Therefore, **the required natural log model is:**

$$\widehat{\ln Y_t} = 1.52663 - 0.0375654 \ln X_1 - 0.0304148 \ln X_2$$

Square root model

Table 3: Estimated parameters for the Square root model

Variable	*Coefficient*	*Std. Error*	*t-ratio*	*p-value*	
const	2.63499	0.379815	6.938	<0.0001	***
SRI	−0.0661932	0.0813873	−0.8133	0.4193	
SRO	−0.0306265	0.0474387	−0.6456	0.5210	
Mean dependent var	2.243106	S.D. dependent var	0.990854		
Sum squared resid	58.69535	S.E. of regression	0.997415		
R-squared	0.019935	Adjusted R-squared	-0.013287		
F(2, 59)	0.600051	P-value(F)	0.552098		
Log-likelihood	−86.27620	Akaike criterion	178.5524		
Schwarz criterion	184.9338	Hannan-Quinn	181.0579		

Sources: Author's Work (2022)
Foot note: SRI = square root of inflation, SRO = square root of oil price

Table 3 shows that the series were not significant at 5% level of significance.

Therefore, **the required square root model is:**

$$\widehat{\sqrt{Y_t}} = 2.63499 - 0.0661932\sqrt{X_1} - 0.0306265\sqrt{X_2}$$

Sine Model

Table 4: Estimated parameters for the Sine model

Variable	*Coefficient*	*Std. Error*	*t-ratio*	*p-value*
const	0.0617025	0.0908768	0.6790	0.4998
SII	0.152755	0.131972	1.157	0.2517
SIO	−0.0670638	0.119687	−0.5603	0.5774
Mean dependent var	0.048243	S.D. dependent var	0.703376	
Sum squared resid	29.26185	S.E. of regression	0.704247	
R-squared	0.030391	Adjusted R-squared	-0.002477	

F(2, 59)	0.924636	P-value(F)	0.402347
Log-likelihood	−64.69785	Akaike criterion	135.3957
Schwarz criterion	141.7771	Hannan-Quinn	137.9012

Sources: Author's Work (2022)
Foot note: SII = sine of inflation, SIO = sine of oil price

Table 4 shows that the series were not significant at 5% level of significance.

Therefore, **the required sine model is:**

$$\widehat{\mathrm{Sin}(\mathrm{Y_t})} = 0.0617025 + 0.152755\mathrm{Sin}(\mathrm{X_1}) - 0.0670638\mathrm{Sin}(\mathrm{X_2})$$

Cosine Model

Table 5: Estimated parameters for the Cosine model

Variable	*Coefficient*	*Std. Error*	*t-ratio*	*p-value*
const	0.120207	0.0909768	1.321	0.1915
COI	0.149532	0.124829	1.198	0.2358
COO	−0.0599031	0.140482	−0.4264	0.6714

Mean dependent var	0.132521	S.D. dependent var	0.708124
Sum squared resid	29.73662	S.E. of regression	0.709937
R-squared	0.027829	Adjusted R-squared	-0.005126
F(2, 59)	0.844456	P-value(F)	0.434919
Log-likelihood	−65.19678	Akaike criterion	136.3936
Schwarz criterion	142.7750	Hannan-Quinn	138.8991

Sources: Author's Work (2022)
Foot note: COI = cosine of inflation, COO = cosine of oil price

Table 5 shows that the series were not significant at 5% level of significance.

Therefore, **the required cosine model is:**

$$\widehat{\mathrm{Cos}(\mathrm{Y_t})} = 0.120207 + 0.149532\mathrm{Cos}(\mathrm{X_1}) - 0.0599031\mathrm{Cos}(\mathrm{X_2})$$

Tangent Model

Table 6: Estimated parameters for the Tangent model

Variable	*Coefficient*	*Std. Error*	*t-ratio*	*p-value*
const	−0.708942	0.571913	−1.240	0.2200
TAI	0.119130	0.259545	0.4590	0.6479
TAO	0.00149076	0.00349323	0.4268	0.6711

Mean dependent var	−0.647944	S.D. dependent var	4.376771
Sum squared resid	1160.918	S.E. of regression	4.435828
R-squared	0.006509	Adjusted R-squared	-0.027169
F(2, 59)	0.193270	P-value(F)	0.824779
Log-likelihood	−178.7990	Akaike criterion	363.5979

Schwarz criterion	369.9793	Hannan-Quinn	366.1034

Sources: Author's Work (2022)

Foot note: TAI = tangent of inflation, TAO = tangent of oil price

Table 6 shows that the series were not significant at 5% level of significance.

Therefore, **the required tangent model is:**

$$\widehat{\mathrm{Tan}(\mathrm{Y_t})} = -0.708942 + 0.119130\mathrm{Tan}(\mathrm{X_1}) + 0.00149076\mathrm{Tan}(\mathrm{X_2})$$

Table 7: The six models parameter estimates with their model selection criteria, R-square and Durbin-Watson statistics using the Ordinary Least Square Method

Variables	parameters	Actual Model	Square Root Model	Nat. Log Model	Sine Model	Cosine Model	Tangent Model
		Coefficient (p-value)	**Coefficient (p-value)**	**Coefficient (p-value)**	**Coefficient (p-value)**	**Coefficient (p-value)**	**Coefficient (p-value)**
	Constant	7.61108(0.0000***)	2.63499(0.0000***)	1.52663(0.0019***)	0.0617025(0.4998)	0.120207(0.1915)	−0.708942(0.2200)
$\mathbf{X_1}$	$\boldsymbol{\beta_1}$	−0.0541218 (0.2170)	−0.0661932(0.4193)	−0.0375654(0.8440)	0.152755(0.2517)	0.149532(0.2358)	0.119130(0.6479)
$\mathbf{X_2}$	$\boldsymbol{\beta_2}$	−0.0244247 (0.2668)	−0.0306265(0.5210)	−0.0304148(0.8002)	−0.067064(0.5774)	−0.05990(0.6714)	0.00149076(0.6711)
	$\mathbf{R^2}$	0.041505	0.019935	0.002924	0.030391	0.027829	0.006509
	Adjusted $\mathbf{R^2}$	0.009014	-0.013287	-0.030875	-0.002477	-0.005126	-0.027169
	St.D	5.046911	1.183771	0.990854	**0.703376**	0.708124	4.376771
	MSE	24.0203	0.9467	1.3747	**0.47197**	0.4796	18.7245
	AIC	379.0403	178.5524	201.6781	**135.3957**	136.3936	363.5979
	BIC	350.9602	184.9338	208.0595	**141.7771**	142.7750	369.9793
	HQC	381.5458	181.0579	204.1836	**137.9012**	138.8991	366.1034
	Log-likelihood	−186.5201	−86.27620	−97.83906	−64.69785	−65.19678	−178.7990
	Durbin-Watson	1.457	1.592	1.798	1.826	2.086	1.914

Sources: Author's Work (2022)

Criteria

5% critical values for Durbin-Watson statistic, n = 62, k = 2
dL = 1.5232
dU = 1.6561

Table 7 shows the Durbin Watson statistic of the six models as well as their model selection criteria and R-square using the Ordinary Least Square Method. The Sine transformed model is identified as the most suitable model based on its (AIC, MSE and BIC).

DISCUSSION OF FINDING

Six multiple linear regression models were utilized to evaluate and deal with the existence of correlation issue in model residual using Nigeria Gross Domestic Product (GDP) data. It's noted from Table 7 by comparison that the actual model had a Durbin Watson value of 1.457 that is lower than the DU and DL, consequently, there's statistical proof that the error terms are favourably autocorrelated. Once again, the square root design had a Durbin Watson value of 1.592 that's higher compared to the DL yet lower than DU; therefore, the test is not conclusive for square root model. Likewise, the majority of the models (natural log, sine, cosine, and tangent) had Durbin Watson - values of 1.789, 1.826, 2.086, and 1.914, which are closer to 2.0 and better compared to DL and Du. Thus, the issue of autocorrelation is solved by these transformations.

The appropriate model for this research is a Sine design that has the very least mean square error, lowest AIC, HQC and smallest BIC that make it the best design, although the particular design contains the greatest R^2 valued while the Cosine design has the greatest Durbin Watson statistical value, but many parameters came up because of transformation effect.

This research's conclusion is in agreement with that of Khalda and AbuElgasim (2018) that carried out a research study on Using Transformation of Variable for treating Autocorrelation in simple regression (first difference & Cochrane - Occutt techniques). Their finding shows that the technique of differences is much better in treating autocorrelation and that there are limits of negative errors because of the reduction of these errors from the least square method.

CONCLUSION AND RECOMMENDATION

Based on the results previously mentioned, it was determined that the best model for this study was Sine, with the very least mean square error, the very least standard deviation, the smallest AIC, HQC, and the smallest BIC. It was also determined that the sine transformation technique may be more effective in getting rid of or treating the issue of autocorrelation in real life data.

Therefore, it is suggested that much more comprehensive work be performed making use of other transformation methods not mentioned or even used in this particular current study in order to determine a far more appropriate design to be used in the research or even to treat autocorrelation issues in the data. It's also suggested that stimulated data must be studied under study making use of these transformations to see if they produce the same results as the current.

References

1. AbuElgasim, A. & Khalda, O. (2018). Using transformation of variable for treating autocorrelation in simple regression (first difference & cochrane-occutt methods). *Elixir Statistics*, 114, 49492-49496.
2. Akaike, AIC. (1974). A new at the statistical model identification. *IEEE Transformation on Automatic control*, 19(0); 716-723.
3. Bisgaard, S. & Kulahci, M. (2015b). The effect of autocorrelation on statistical process control procedures. *Engel wood Cliffs*
4. Chatterjee, S. & Simonoff, J. (2013). Handbook of regression analysis. *John Wiley & Sons*. ISBN 1118532813.
5. Craven, P. & Wahba, G. (1978). Estimating the dimension of a model. Annuals of statistical, 6, 461-4164
6. Dorland, F.M. (2018). Applied linear statistical methods. *Prentice-hall, Inc.*
7. Neumann, J.V. (1941). Distribution of the ration of the mean square successive difference to the variance. *The Annals of Mathematical Statistics*, 12(4), 367-395.
8. Sargan, J.D. & Bhargava, A. (2018). Testing residuals from least squares regression for being generated by the Gaussian Random Walk". *Econometrica*. 51(1), 153 – 174. JSTOR 1912252
9. Shalabh, B. (2017). Ratio method of estimation in the presence of measurement errors. *Journal of Indian Society of Agricultural Statistics* 50(2):150– 155.