Scholars Journal of Economics, Business and Management
OPEN ACCESS
Abbreviated Key Title: Sch J Econ Bus Manag
ISSN 2348-8875 (Print) | ISSN 2348-5302 (Online)
Journal homepage: https://saspublishers.com

Econometric Modeling of Nigeria's GDP; A Variable Selection Approach

Guobadia Emwinloghosa Kenneth[1], [2*]

[1]Department of Administration, Federal Medical Centre, Asaba, Delta State, Nigeria
[2]Department of Statistics, University of Benin, Benin, Nigeria

DOI: 10.36347/sjebm.2021.v08i01.003 | **Received:** 15.01.2021 | **Accepted:** 24.01.2021 | **Published:** 28.01.2021

***Corresponding author:** Guobadia Emwinloghosa Kenneth

Abstract **Original Research Article**

The output of four variable selection techniques in the construction of a model that best estimates a dependent variable is critically evaluated in this analysis. The techniques for variable selection are the direct search on the t equation, the method of forward selection, the method of backward exclusion, and the method of stepwise regression. In contrast, economic data of 32 years were collected on Real Gross Domestic Product each, that was the dependent variable used as a measure of economic development and growth, and seven factors; Growth Market Capitalization, All-Shares Index, Market Turn-Over, Nigerian Trade Economy Transparency, Transaction Value, Nigerian Stock Exchange Total Listing. Using the four variable selection techniques, the residual mean square, modified R^2, and the variance inflation factor obtained from the use of each of these techniques, which are the criteria for determining the best model, the actual gross domestic product was compared with the seven variables by rating them on the basis of the formula that best met the evaluation criteria. The result shows that the backward elimination method performs better in variable selection based on the sample collected with a mean rank of 1.67 taken across the parameters, and supports the use of the all possible combination method as a control.

Keywords: Variable Selection Technic, Regression Analysis, Backward Elimination Method, Stepwise Regression, Growth Market Capitalization, All-Shares Index, Market Turn-Over, Real Gross Domestic Product.

INTRODUCTION

Regression is a very powerful tool which uses other independent variables to explain a variable. But a certain variable is determined by a large number of variables in real life. Others are measurable and some are immeasurable, such as wages, agriculture, the environment, the success of students, etc. Some of these factors are significant, while others are less significant, since it is not possible to account for all the variations in a certain factor (Independent variable). Thus, variables whose effect on the dependent variable is comparatively more important must be chosen. It is important to decide the exact subset of regressors that should be included in the model in a situation where there is a pool of candidate regressors that should contain all the influential variables [1-3]. The variable selection problem is named to find a suitable subset of regressors for the model and it involves the use of certain variable selection techniques, such as; Direct Search on t Statistics, Backward Elimination Process, Forward Selection Method, Stepwise Regression Method, and All Possible Combination Method. Two opposing aims are involved in designing a regression model that includes only a subset of the available regressors; to construct a model that would include as many regressors as possible so that the knowledge quality of these variables will impact the expected value of the dependent variable; and to build a model that would include as few regressors as possible because the variance of prediction increases with variance in the number of regressors. Although it may not be preferred to construct the model with too many regressors, because it requires higher data collection and model maintenance costs. Selecting the "Best" regression equation is called the method of seeking a model that is a balance between these two goals. In a situation where both methods have different equations as the suitable subset model for estimation, the mean rank method of choosing the best will be applied; that is, the method will be graded on the basis of the parameters used. This will be used to pick the best All-Possible Regression Process model. Factors such as the Nigerian growth market capitalization, the Nigerian All-Share Index, Total Listing on the Nigerian Stock Exchange, Total New Issues, Transparency of Nigeria Trading Here, an estimation of Nigerian economic growth is to be obtained using the Nigerian Gross Domestic Product as a metric for growth and development, and seven (7) factors that affect its growth are to be considered here. One of the key issues with the use of different variable selection methods is that often the methods do not have the same subset models as the appropriate model. In such cases, where different methods have different sub-set regressors, each of the regressors has different contributions to the estimate of the dependent variable in such models. Therefore, it is important to know which of the different variable selection techniques obtained from these models is best for estimation. To obtain the best equation for the estimate of the Nigerian Gross Domestic Product using all variable selection techniques, to find out if all the variable selection tactics used will provide the same sub-set regressor model as the best model, to compare the best equation provided by these techniques, using their residual mean square and modified R^2 as a criterion [4-7]. It is not cost-effective to use the All-Possible approach, and it takes time to execute. For issues involving more than a few regressors, it is inefficient and it also involves the availability of high-speed computers that can build successful algorithms for it. This study is carried out in order to illustrate the use of other variable selection approaches to researchers and students who perform regression analysis, where they have to obtain the best subset regressor for estimating a given dependent variable, and probably at the end of this work they will be able to show them which of these techniques is better and more accurate. This study is also intended to inform students or young researchers about the use of mean ranks, where the need to compare the best performing method is required [8-11].

METHODOLOGY

Sources of Data

For this analysis, secondary data sources were employed. These include the Nigerian Stock Exchange Fact Books, the annual reports and accounts of the Nigerian Stock Exchange (for different years), the Statistical Bulletins of the Central Bank of Nigeria, the Statistical Bulletin of the Federal Office of Statistics. The variables used on the Nigerian Stock Market span the years 1981 to 2013, based on their authenticity and reliability [12-15]. Using the gross domestic product as the dependent variable and the independent variable are: Market Capitalization Rise, Total New Issues, Total Transaction Volume, Total Listed Equities and Government Stock, Total Market Turnover, Nigerian Economy All-Share Index, and Transparency. Accessible economic theories for theoretical support were also tested.

Limitations of the Data

There are actually several variables and determinants to consider when talking about the effect of the Nigerian stock market on its economic growth. As such, due to the accuracy and availability of data on an annual basis, the analysis was limited and data was collected on only seven of the variables.

Data Presentation

Table-1: Annual Report of the Nigerian Stock Exchange

YEAR	GDP	GMC	ASI	TLNSE	TNI	OOTE	VAL OF TRANS	TNOV
1982	315458.10	4464.20	88.00	157.00	423.50	0.047	388.70	0.23
1983	205222.10	4979.80	87.00	194.00	455.20	0.062	304.80	0.19
1984	199688.20	4025.70	94.00	205.00	533.40	0.077	214.80	0.21
1985	185598.10	5768.00	111.00	212.00	448.50	0.057	397.90	0.26
1986	183563.00	5514.90	100.00	213.00	159.80	0.099	418.20	0.25
1987	201036.30	6670.70	127.3	220.00	817.20	0.093	319.60	0.31
1988	205971.40	6794.80	163.8	240.00	833.00	0.072	494.40	0.49
1989	204806.5	8297.60	190.9	244.00	450.70	0.235	348.00	0.29
1990	219876.80	10020.80	233.6	253.00	400.00	0.239	137.60	0.25
1991	263729.60	12848.60	325.3	267.00	1629.90	0.375	521.60	0.65
1992	267660.00	16358.40	513.8	295.00	9964.50	0.582	265.50	0.31
1993	265379.10	23125.00	783.0	239.00	1870.00	0.795	136.00	0.23
1994	274833.30	31272.60	1107.60	251.00	3306.30	1.285	313.50	0.49
1995	275450.60	47436.10	1548.80	272.00	2636.90	1.399	402.30	0.66
1996	281407.40	663680.00	2205.00	276.00	2161.70	1.339	569.70	0.99
1997	293745.40	180305.10	5092.20	276.004	4425.60	6.373	1838.80	1.84
1998	302022.50	281815.80	6992.10	276.00	5858.20	6.373	7062.70	7.06
1999	310890.10	281887.20	6440.50	264.00	10875.70	6.911	11072.70	11.07
2000	312183.50	262517.30	5716.00	264.00	15018.10	5.112	13572.30	13.50
2001	329978.70	300041.10	5266.40	268.00	12038.50	6.571	14027.40	14.10
2002	356994.30	427290.00	8111.00	260.00	17207.80	8.903	28154.60	28.15
2003	433203.50	662561.30	10965.00	261.00	37198.80	9.037	57637.20	57.68
2004	477833.00	764975.80	12137.70	258.00	61284.00	7.518	60088.60	59.41
2005	527576.00	1359274.20	21222.60	277.00	180079.9	10.823	120703.00	120.40

2006	561931.40	2112549.60	23844.50	288.00	195418.4	12.491	225820.60	225.80
2007	595821.60	2900062.10	24085.80	294.00	552782.0	17.880	470257.00	262.94
2008	634251.00	5120000.00	33189.30	310.00	707400.0	18.020	1076020.40	470.25
2009	674889.00	13294059.0	57990.20	301.00	1935080.0	19.721	1679143.70	2086.29
2010	716949.70	9562970.00	31450.80	266.0	1509230.00	23.257	68572000.0	2379.14
2011	801700.00	9920000.00	46437.64	264.0	1894374.50	23.734	79755000.0	2388.34
2012	901300.00	10280000.0	59365.75	250.0	1735623.34	25.224	63492000.0	2511.67
2013	1067650.0	89000000.0	64768.55	198.0	1843274.87	27.555	62758000.0	2676.24

Source: Nigerian Stock Exchange Annual Reports and Accounts, various years; SEC Annual Reports and accounts; CBN Statistical Bulletin, Golden Jubilee Edition.

Model Specification

In line with the above specification, the research model is specified thus:

$$\mathbf{GDP} = \mathbf{f}(\mathbf{GMC}, \mathbf{ASI}, \mathbf{TLNSE}, \mathbf{TNI}, \mathbf{OOTE}, \mathbf{VALTRANS}, \mathbf{TNOV})$$

Where

$\text{GDP} = \text{Gross Domestic Product}$

$\text{GMC} = \text{Growth of Market Capitalization}$

$\text{ASI} = \text{All} - \text{Share Index}$

$\text{TLNSE} = \text{Total Listing on the Nigerian Stock Exchange}$

$\text{TNI} = \text{Total New Issues}$

$\text{OOTE} = \text{Openness of Nigerian Trade Economy}$

$\text{VALTRANS} = \text{Value of Transactions}$

$\text{TMT} = \text{Total Market Turnover}$

METHODOLOGY

There are more than one independent and one dependent variable in this analysis, and the interest is to evaluate the subset model that best estimates the dependent variable, Real GDP; in Regression analysis, such analysis is called the variable selection technique. In this analysis, we have different subtopics that we will exploit because they are needed in this analysis [16].

Correlation Theory

Correlation is an indicator of the power of two random variables to have a linear relationship. Multiple associations are referred to as the degree of connection linking more than one variable. When all points (X, Y) on the scatter diagram tend to cluster near a straight line [18], the association can be linear. The relationship is linear if all points in the dispersed diagram appear to lie near a line. There might be a positive correlation between two variables, a negative correlation or maybe uncorrelated.

If they appear to change in the same direction together, that is, if they decrease or grow together, two factors are said to be positively correlated. If they appear to change in the opposite direction, two variables are said to be negatively correlated, that is, when one increases, the other decreases, and vice versa. Two variables, if they are not inherently independent of each other, are said to be uncorrelated or have no association.

The Pearson Product Moment Correlation Coefficient between X and Y can be expressed as;

$$\widehat{\boldsymbol{\rho}} = \frac{\sum(\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{y_i} - \overline{\boldsymbol{y}})}{\sqrt{\sum(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2}\sqrt{\sum(\boldsymbol{y_i} - \overline{\boldsymbol{y}})^2}}$$

Where $\widehat{\boldsymbol{\rho}}$ is the popul ation correlation, $\boldsymbol{x_i}$ and $\boldsymbol{y_i}$ are the ith observation of the two variables of interest. While $\overline{\boldsymbol{x}}$ and $\overline{\boldsymbol{y}}$ are the mean of the ith observation of the two variables of interest.

Regression Theory

Regression is a statistical tool for evaluating the relationship between one or more dependent variables $X_1, X_2, \ldots, X_n$ and a single continuous dependent variable Y. it is most often used when the independent variables are not controllable, that is, when collected in a sample survey or other observational studies. There are so many types of regression model, but just one of the regression models will be used for this study and that is the linear regression model [17].

Linear Regression Model

If and only if a variable is a function of another variable whose power equals one, a regression model is linear. There are, and are, two kinds of linear regression; simple linear regression and multiple linear regressions. Only multiple regressions will be considered for the purpose of the study.

Multiple Linear Regressions

In order to infer a dependent variable from some independent variables, multiple linear regression is a statistical analysis that suits a model. More than one independent variable is involved in multiple regressions [19]. As follows, the relationship between the dependent and independent variable is expressed;—

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + \varepsilon_i$$

Where Y_i is the ith response or dependent variable?

x_1 is the ith independent variable

ε_i is the error term of the ith observation which is normally, independently distributed with mean zero and variance $[\sigma_\varepsilon^2]$.

$\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are the observation parameter.

Estimation of Parameters of the Model

By several methods, unbiased estimates of the parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ can be produced. The form of least squares is the most commonly used. This implies that the divergence of the observed value of Y from its predicted value is reduced by the number of squares. In other words, the sample estimates $b_0, b_1, \dots, b_k$ of β_0, $\beta_1, \dots, \beta_k$ are chosen in such a way by the process of least squares, respectively, that

Q = $\sum \varepsilon_i^2 = \sum (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \cdots - \beta_k X_{ik})^2$ is minimized.

As in simple regression, by solving the following series of normal equations, we obtain estimates, $b_0, b_1, \dots, b_k$ of the regression coefficients

$$\sum Y_i = nb_0 + b_1 \sum X_{i1} + b_2 \sum X_{i2} + \cdots + b_k \sum X_{ik}$$
$$\sum X_{i1} Y_i = b_0 \sum X_{i1} + b_i \sum X_{i1}^2 + b_2 \sum X_{i1} X_{i2} + \cdots + b_k \sum X_{i1} X_{ik}$$
$$\sum X_{i2} Y_i = b_0 \sum X_{i2} + b_i \sum X_{i1} X_{i2} + b_2 \sum X_{i2}^2 + \cdots + b_k \sum X_{i1} X_{ik}$$

$$.\sum X_{ik} Y_i = b_0 \sum X_{ik} + b_i \sum X_{i1} X_{ik} + b_2 \sum X_{i1} X_{i2} + \cdots + b_k \sum X_{ik}^2$$

Although these normal equations are mathematically obtained by finding estimates $b_0, b_1, \dots, b_k$ that would minimize equation Q, a simple procedure to note when obtaining them is as follows: $b_0, b_1, \dots, b_k$ as coefficients are written down with the regular regression equation. By summing up each term of this regression equation, the first normal equation is then obtained. By multiplying each term in the regression equation by X_{i1} and summing the result, the second normal equation is obtained. By multiplying every term in the regression equation by X_{i2} and summing the result, the third normal equation is obtained; and so on [20].

For the figures $b_0, b_1, \dots, b_k$ it would be too cumbersome to obtain separate expressions. Instead, these coefficients are obtained by calculating the sums needed for the different combinations of $X_1, X_2, \dots, and\ X_k$ from the data and substituting these sums into the usual equations that are solved at the same time.

Assumptions of the Regression Analysis

The following are the assumptions of the regression analysis above;

1. X values are fixed in repeated sampling.
2. Zero mean value of disturbanceμ_i. Given the value of X, the mean, or expected value of the random disturbance term μ_i is zero. Symbolically, we have $E(\mu_i / X_i) = 0$.
3. Homoscedasticity or equal variance of μ_i. Given the value of X, the variance of μ_i is the same for all observations. That is, the conditional variances of μ_i are identical.
4. No autocorrelation between the disturbances. Given any two X values, X_i and X_j(i≠j) is zero.
5. Zero covariance between μ_i and X_i, or $E(\mu_i X_i)$ =0.
6. The number of observations n must be greater than the number of parameters to be estimated.
7. Variability in X values. The X values in a given sample must not all be the same.
8. There is no perfect multicollinearity. That is, there are no perfect linear relationships among the explanatory variable.
9. The error term are normally distributed, that is, $\mu_i \sim N(0, \delta_u^2)$.

Hypothesis Testing

There are two types of hypothesis testing: null hypothesis and alternative hypothesis. The hypothesis being tested is the null hypothesis. With the aim of being rejected, it is always conceived. It is mentioned as being

$H_0: \beta = 0$ which shows That the coefficients are the same.

The hypothesis which contradicts the null hypothesis is the alternative hypothesis. It is mentioned as being
$H_1: \beta > 0$ or
$H_1: \beta < 0$ or
$H_1: \beta \neq 0$

Shows that the coefficients are not the same

Homoscedasticity

If they have equal or constant variance, observations are said to be homozcedastic. Since one of the regression assumptions is that the residuals have constant variance, to draw a conclusion on homoscedasticity, we will be using a scatter plot of the standardized predictors [21].

Autocorrelation

The term autocorrelation can be defined as a correlation of time-ordered observation series members, that is, time series data or space/cross-sectional data. The most celebrated test developed by statisticians Durbin and Watson d statistics for detecting serial correlation is that. A major benefit of the d statistic is that it is based on the approximate residuals regularly measured in regression analysis [22-23]. Durbin and Watson, however, succeeded in deriving a lower bound dl and an upper bound du such that a judgment on the existence of positive or negative serial correlation can be taken if the computed d is beyond these critical values. In addition, this limit only depends on the number of n-observations and the number of explanatory variables. Dublin and Watson have tabulated the thresholds, varying from 6 to 200 for n and up to 20 explanatory variables, and the limits of 0 and 4.

Test of hypothesis

H_0: There is no autocorrelation.
H_1: There is autocorrelation.

Decision: Reject the null hypothesis if the Durbin- Watson d statistics value falls outside the limit of d, that is, within the range of 0 and 4.

Level of Significance

The significance level is the distinction between the appropriate percentage and 100 percent. For example, if 95 percent is definitely required, then the significance level will be denoted as alpha = 0.05. This is the likelihood that a type one error will be committed, while a type one error will actually deny a true null hypothesis.

Test for Model Adequacy

Table-2: Test for Model Adequacy

SV	d.f	SS	MS	F-ratio
Regression	k-1	SSR	MSR	
Residual	n-(k+1)	SSE	MSE	F=$\frac{MSR}{MSE}$
Total	n-1	SST		

Test of Hypothesis
H_0: The model is not adequate.
H_1: The model is adequate.
Using a 5% level of significance

Test Statistic

$F_{cal} = \frac{MSR}{MSE} \sim F^{(\alpha)}_{k,n-(k+1)}$

Decision Rule

Reject H_0: if $F_{cal} > F_{tab}, accept\ if\ otherwise.$

Test for Parameter Significance

This is simply a test of the significance of the individual parameters in the model.

Test of hypothesis

$H_0: \beta = 0$ (The coefficient is not statistically significant)
$H_1: \beta \neq 0$ (The coefficient is statistically significant)

Using a 5% level of significance

$$t_{cal} = \frac{\hat{\beta}_i}{Se(\hat{\beta}_i)} \sim t^{\alpha-2}_{n-k}$$

Decision Rule

Reject H_0 if $/t_{cal}/ > t_{tab}$, accept if otherwise.

Critical Region

The critical region shows the importance of the test statistics, which means that the null hypothesis will be dismissed. It is also called the area of rejection. The acceptance region, which indicates the importance of test statistics that would mean acceptance of the null hypothesis, is the opposite of the critical region.

Multicollinearity

To denote the existence of linear or near linear dependency among the explanatory variables, multicollinearity is used. Multiple regression models with associated explanatory variables show how well the outcome of the variable is predicted by the entire package of predictors, but it does not provide reliable results on any particular predictor or on which predictors are redundant with others [24-26]. If the correlation between two independent variables is equal to +1 or -1, we have a perfect multicollinearity, so that if the following condition is met, we have an exact linear relationship;

$$a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_kx_k = 0$$

But if the X variables are intercorrelated to the Y variables, we have

$$a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_kx_k + v_i$$

Where α is the constant, v_i is the random term, and x_i represents the independent variables with i= 1,2,3, …, k.

Multicollinearity Diagnostics

Several techniques for multicollinearity detection have been suggested, but three techniques will be considered here. A diagnostic measure's desirable characteristics are that it clearly represents the degree of the issue of multicollinearity and provides information helpful in assessing which regressors are involved. We have the study of the matrix of correlation, variance inflation variables, X'X's own device analysis.

The variance inflation factor was used here to account for the impact of multicollinearity on the regression model of different subsets.

$$VIF = \frac{1}{1 - R^2_{ij}}$$

Variable Selection Techniques

Regression models that employ a subset of the candidate regressor variables are desirable to consider. It is normal to consider fitting models with different combinations of the candidate regressors in order to find the subset of variables to be included in the final equation. For generating subset regression models, there are many computational approaches, but our focus will be focused on four of these techniques and they are; all possible regressions, direct search on t, method of forward selection, method of backward elimination, and regression stepwise.

All Possible Regressions

This approach demands that all regression equations involving one candidate regressor, two candidate regressors and so on be fitted by the analyst. According to some relevant parameters, these equations are tested and the best regression model chosen. If we assume that the intercept term β 0 is used in all equations, there are 2^k total equations to be calculated if there are k candidate regressors. We consider the modified R^2 for the appropriate model, which is insensitive to the number of variables in the model, making it ideal for decision making in this process, where we have to choose the best model combination from the different model combinations, It is also important to take into account the number of variables in the model, but also the size of the model, since the more variables, the more information obtained from these variables, which actually has a major impact on the expected value of the independent variable. And the statement made with regard to the size must be careful, as the more the variable, the greater the forecast's uncertainty.

Direct Search on t

The test statistics for testing $H_0: B_J = 0$ for the full model with p=K+1 regressors is

$$t_{k,j} = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)}$$

Regressors that contribute significantly to the full model will have a large $/t_{k,j}/$ and will tend to be included in the best p-regressor subset, where best implies minimum residual sum of squares or C_p. Consequently ranking the regressors according to decreasing order of magnitude of the $/t_{k,j}/$, j=1,2, …, k, and then introducing the regressors into the model one at a time in this order should lead to the best or one of the best subset models for each p [Daniel and wood, 1980].

Forward Selection Method

This approach starts with the assumption that other than the intercept, there are no regressors in the model. By integrating regressors into the model one at a time, an attempt is made to find an optimal subset. The first regressor chosen for entry into the equation is the one that has the greatest simple correlation with the y response variable, and it is also the regressor that provides the greatest F- statistic for regression significance testing. If the F-statistics reaches a preselected F value, this regressor is entered, say F INN (or F- to- enter). The second regressor selected for entry is the one that now has the greatest association with y after correcting for the first regression effect. We refer to these correlations as partial correlations. They are simple correlation between the residuals from the equation $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$ and the residuals from the regressions of the other candidate regressors on x_1, say $\hat{x} = \hat{\alpha}_{0j} + \hat{\alpha}_{1j} x_1$, j=2,3, …k.

Here, the regressor with the highest partial correlation also means the largest F-statistic, and then the regressor joins the model if its F value reaches F-IN. In general, the regressor with the highest partial correlation with y is added to the model at each stage given the other regressor already in the model, if its partial F-statistic exceeds the preselected entry level F-IN. This process ends either when no F-IN is surpassed by the partial F statistics at a given point or when the last candidate regressor is applied to the model.

Backward Elimination Method

With no regressors in the model, forward selection starts and attempts to introduce variables before a suitable model is obtained. By working in the opposite direction, backward eliminations seek to find a successful model. That is, we start with a model containing all regressors for the K candidate. The partial F-statistic partial F-statistic is then compared to a preselected value, for example F-OUT (or F-to-remove), and if the smallest partial statistics are less than F-OUT, the regressor is removed from the model.

Criteria for Evaluating Subset Regression Models

In testing subset regression models, we make use of the following as a metric of adequacy in order to get the highest. We have the multi-determination coefficient, the residual mean square, the modified coefficient of multiple determinations.

Coefficient of Multiple Determinations R^2

The proportion/percentage of the overall variance in the dependent variable observed that can be clarified by the independent variables is the coefficient of multiple determinations. The intensity of the relation between the dependent and the independent variables is measured. It is given as the

$$R^2 = \frac{variation\ in\ Y\ explained\ by\ X_i'^s}{Total\ variation\ in\ Y}$$

$$R^2 = 1 - \frac{SSE}{SST}$$

Adjusted R^2

To avoid the difficulties of interpreting$\boldsymbol{R^2}$**,** the use of adjusted $\boldsymbol{R^2}$ statistics is preferable, defined for a p-term equation as

$$\bar{R}_p^2 = 1 - \left(\frac{n-1}{n-p}\right)\left(1 - R_p^2\right)$$

The $\bar{R}_p^2$ – statistics does not necessarily increase as additional regressors are introduced into the model, except the partial F- statistic for testing the significance of the s additional regressors exceeds one. The criterion for selection of an optimum subset model is to choose the model that has a maximum $\bar{R}_p^2$.

Residual Mean Square

As a model estimation criterion, the residual mean square for a sub-set residual model can be used.

The MSE(P) experiments with an initial decrease, then it stabilizes and can gradually increase as p increases, since SSE(p) always decreases as p increases as the amount of square error.

$$MS_E(P) = \frac{SS_E(P)}{n-p}$$

Choosing the model with the following is the criterion for choosing an optimal subset model;

1. The minimum $MS_E(P)$,
2. The value of p such that $MS_E(P)$, is approximately equal to MS_E.

Note: the subset regression model that minimizes $MS_E(P)$ will also maximize $\bar{R}_p^2$.

Proof;

Where

$$\bar{R}_p^2 = 1 - \left(\frac{n-1}{n-p}\right)(1 - R_p^2)$$
$$= 1 - \left(\frac{n-1}{n-p}\right)\frac{SS_E(P)}{S_{yy}}$$
$$= 1 - \left(\frac{n-1}{S_{yy}}\right)\frac{SS_E(P)}{n-p}$$

We recall;

$$MS_E(P) = \frac{SS_E(P)}{n-p}$$
$$\bar{R}_P^2 = 1 - \frac{n-1}{S_{yy}} MS_E(P)$$

Thus the criteria minimum residual mean square and maximum adjusted coefficient of multiple determinations are equivalent.

DATA ANALYSIS AND INTERPRETATIONS

Multiple Regressions

Regression is a statistical method to determine the relationship between one or more independent variable(s) $X_1, X_2, \dots, X_n$ and a single dependent continuous variable Y. It is most commonly used when independent variables, that is, when obtained in a sample survey or other observational studies, are not controllable. Multiple regressions require more than one separate variable. However, the multiple regression analysis in this work consists of seven (7) independent variables, including: gross market capitalization, total new problems, transaction size, and market turnover, total listing of the stock exchange of Nigeria, the All-share index, and the openness of Nigerian Trade Economy. Then the dependent variable is the Gross Domestic Product.

Variable Specification

From the data analysis, there are these following variable specifications

Y_i = Gross Domestic product(GDP)
X_1 = Growth of Market Capitalization.
X_2 = All − share Index
X_3 = Total Listing on the Nigerian stock exchange
X_4 = Total New Issues
X_5 = Openness of Nigerian Trade Economy
X_6 = Value of Transaction
X_7 = Total Market Turnover

Testing for Homoscedasticity

Using the scatter plot of the standardized residual against the standardized predictors on the entire model obtained to search for constant variance by the various variable selection techniques, we observe that only a few points less than four of the residuals differ from each obtained graph. This means that outliers are present. But because only a few points differ in the entire obtained graph, we can therefore assume that there is constant variation in the residuals.

Testing for Autocorrelation

Using Durbin Watson d Statistics, we note that the Durbin Watson d statistical value is 1.049, 0.986, 1.203, and 0.986 for the direct search on t, forward selection technique, backward exclusion technique, and stepwise regression method. Since all the values obtained fall within the range of 0 and 4, we accept the null hypothesis and conclude that there is no autocorrelation between residual overtime provided by the different techniques given by each subset regressor model.

Obtaining the Optimal Regression Model for the Estimation

Here, we have seen independent variables and an optimal regression model is required.

Using Direct Search on t

Table-3: Summary of regression coefficients direct search on t

Model		Unstandardized	Coefficient	T	Sig.
		B	Std. Error		
	(Constant)	232231.2	66927.63	3.47	0.002
	GMC	0.001	0.001	1.697	0.103
	ASI	9.066	2.099	4.32	0
	VAL OF TRANS	0.003	0.001	2.472	0.021
	TLNSE	-42.366	282.236	-0.15	0.882
	OOTE	11237.22	3352.694	3.352	0.003
	TNOV	-53.632	73.101	-0.734	0.47
1	TNI	-0.1	0.086	-1.161	0.257

From Table 4.5.1, we note that the growth market capitalization regression coefficients, Nigerian stock exchange total listing, Total new issues, transaction value, and market turnover are 0.001, -42.366, -0.100, 0.003, -53.632 and the likelihood associated with their t values are all higher than the preselected significance level of 0.05. This means that the contribution to the model is not statistically important for the five independent variables. While the All-shares index regression coefficient and the Nigerian trade economy's openness are 9,066 and 11237,220 and the likelihood associated with their t values are both lower than the preselected level of significance 0.05, which shows that the contribution of the All-share index and openness of Nigerian trade economy statistically significant.

Therefore, we have regressed the gross domestic product on the all-share index and transparency of the Nigerian trade economy by excluding the insignificant independent variables, and the model obtained is;

$$GDP = 220913.494 + 4.998ASI + 14968.78OOTE$$

Which have a R^2 value of 0.958 which shows that 95.8% of the total variation in the real GDP can be explained by the independent variables ASI and OOTE, and an adjusted R^2 value of 0.956 and a residual mean square value of 2361918160.279.

Testing for Model Adequacy

Hypothesis

$\mathrm{H}_0 =$ The model is not adequate.

$\mathrm{H}_1 =$ The model is adequate.

Using a 5% level of significance,

Table-4: Output on test for model adequacy using direct search on t

Model	Sum of squares	DF	Mean Square	F	Sig.
Regression	1.577E+12	2	7.886E+11	333.902	0.000
Residual	6.85E+10	29	2.362E+09		
Total	1.646E+12	31			

Interpretation: We reject the null hypothesis argument from the above result with an F-ratio of 333.902, which is important at 0.000<0.05, that the model is not adequate and conclude that the model is statistically adequate.

Test for Parameter Significance

Hypothesis

$\mathrm{H}_0 =$ The parameter is not significant.

$\mathrm{H}_1 =$ The parameter is significant.

Using a 5% level of significance,

Table-5: Summary on the significant regression coefficient using direct search on t

MODEL	Unstandardized Coefficients	T	Sig.	Collinearity Statistics VIF
Constant	220913.5	18.552	0	
ASI	4.998	3.513	0.001	9.985
OOTE	14968.78	4.77	0	9.985

Interpretation: From the outcome, we find that both independent variables X_2 and X_5 have unstandardized coefficients of 4.998 and 14968.78 with t-values of 3.513 and 4.770 respectively, and are both important at 0.001<0.05 and 0.000<0.05 respectively, respectively. We therefore reject the assertion of the null hypothesis that the parameters are not relevant and conclude that the parameters are statistically important.

Using the Backward Elimination Method

Using backward elimination method the appropriate model for this analysis, using a $F_{OUT} = 0.10$ is;

$$GDP = 220550.460 + 10.593ASI - 0.179TNI + 10016.59OOTE + 0.003VALTRANS$$

R^2=0.975 which shows that 97.5% of the total variation in real GDP can be explained by the independent variables ASI, VAL OF TRANS, OOTE, and TNI. With the adjusted R^2=0.972 indicating that the fit is good, and residual mean square of 1503272992.23.

Testing for Model Adequacy

HYPOTHESIS:

$\mathrm{H}_0 =$ The model is not adequate.

$\mathrm{H}_1 =$ The model is adequate.

Using a 5% level of significance,

Table-6: Output on test for model adequacy using backward elimination method

Model	Sum of squares	DF	Mean Square	F	Sig.
Regression	1.16E+12	4	4.01E+11	266.951	0.000
Residual	4.06E+10	27	1.5E+09		
Total	1.65E+12	31			

Interpretation: From the result in table 4.5.2.1 above, we observe that with a F-ratio of 266.951 which is significant at 0.000<0.05, we reject the null hypothesis statement that the model is not adequate and conclude that the model is adequate.

Test for Parameter Significance

HYPOTHESIS:

H_0 = The parameter is not significant.

H_1 = The parameter is significant.

Using a 5% level of significance,

Table-7: Summary on the significant regression coefficient using backward elimination method

MODEL	Unstandardized Coefficients		T	Sig.	Collinearity Statistics VIF
	B	Std. error			
Constant	220550	10020.92	22.009	0	
ASI	10.593	1.843	5.748	0	26.324
VALTRANS	0.003	0.001	4.024	0	4.557
OOTE	10016.59	2781.477	3.601	0.001	12.324
TNI	-0.179	0.046	-3.907	0.001	18.471

Interpretation: From the result in the table above, we observed that the independent variables X_2, $X_4, X_5,$ and X_6 have coefficients of 10.593, -0.179, 10016.59 and 0.003 respectively with t-values of 5.748, -3.907, 3.601, and 4.024 respectively with p-values all less than 0.05 indicating that the contribution of all the regressors to the model is statistically significant.

Using the Forward Selection Method

Using forward selection method, the appropriate model for this analysis, using a F_{IN}=0.05 is;

$$GDP = 223470.496 + 0.129GMC + 0.287ASI + 0.622OOTE$$

With a R^2=0.968; that is, 96.8% of the total variation in the GDP can be explained by the model.

Testing for Model Adequacy

HYPOTHESIS

H_0 = The model is not adequate.

H_1 = The model is adequate.

Using a 5% level of significance,

Table-8: Output on test for model adequacy using forward selection method

Model	Sum of squares	DF	Mean Square	F	Sig.
Regression	1.59E+12	3	5.31E+11	278.334	0.000
Residual	5.34E+10	28	1.91E+09		
Total	1.65E+12	31			

Interpretation: From the result in table 4.5.3.1 above, we observe that with a F-ratio of 278.334 which is significant at 0.000<0.05, we reject the null hypothesis statement that the model is not adequate and conclude that the model is adequate.

Test for Parameter Significance

HYPOTHESIS:

H_0 = The parameter is not significant.

H_1 = The parameter is significant.

Using a 5% level of significance,

Table-9: Summary on the significant regression coefficient using forward selection method

MODEL	Unstandardized Coefficients		T	Sig.	Collinearity Statistics
	B	Std. error			VIF
Constant	223470.5	10738.53	20.81	0	
ASI	3.406	1.398	2.436	0.021	11.941
OOTE	16315.85	2860.241	5.704	0	10.272
GMC	0.002	0.001	2.814	0.009	1.821

Interpretation: From the result in the table above, we observe that ASI, OOTE, and GMC have regression coefficient 3.406, 16315.85, and 0.002 respectively with t values of 2.436, 5.704, 2.814 respectively, with all contributing significantly to the estimates obtained from the model as the estimated values for the dependent variable, since all have p-values less than 0.05, the preselected level of significance.

Using the Stepwise Regression Method

Using stepwise regression method the appropriate model for this analysis, using F_{IN}=0.05 and F_{OUT}=0.10 is;

$$\mathrm{GDP} = 223470.496 + 0.129\mathrm{GMC} + 0.287\mathrm{ASI} + 0.62200\mathrm{TE}$$

With a R^2=0.968; that is, 96.8% of the total variation in the GDP can be explained by the model.

Testing for Model Adequacy

HYPOTHESIS:

H_0 = The model is not adequate.

H_1 = The model is adequate.

Using a 5% level of significance,

Table-10: Output on test for model adequacy using stepwise regression method

Model	Sum of squares	DF	Mean Square	F	Sig.
Regression	1.59E+12	3	5.31E+11	278.334	0.000
Residual	5.34E+10	28	1.91E+09		
Total	1.65E+12	31			

Interpretation: From the result in table 4.5.11 above, we observe that the model obtained using the stepwise regression method is also the as that obtained from using the forward selection method, having a F-ratio of 278.334 also which is significant at 0.000<0.05, we reject the null hypothesis statement that the model is not adequate and conclude that the model is adequate.

Test for Parameter Significance

HYPOTHESIS:

H_0 = The parameter is not significant.
H_1 = The parameter is significant.
Using a 5% level of significance,

Table-11: Summary on the significant regression coefficient using stepwise regression method

MODEL	Unstandardized Coefficients		T	Sig.	Collinearity Statistics
	B	Std. error			VIF
Constant	223470.5	10738.53	20.81	0	
ASI	3.406	1.398	2.436	0.021	11.941
OOTE	16315.85	2860.241	5.704	0	10.272
GMC	0.002	0.001	2.814	0.009	1.821

Interpretation: From the result in the table above, we observe that ASI, OOTE, and GMC have regression coefficient 3.406, 16315.85, and 0.002 respectively with t values of 2.436, 5.704, 2.814 respectively, with all contributing significantly to the estimates obtained from the model as the estimated values for the dependent variable, since all have p-values less than 0.05, the preselected level of significance.

Comparing Performance of the Method

Table-12: Ranked Performance of the four variable selection techniques

	Residual Mean Square	Total VIF	Adjusted R^2	Total Rank	Average Rank
Direct Search on t	2361918160.279(4)	19.97(1)	0.956(4)	9	3
Backward elimination	1388856407.982(1)	61.68(3)	0.974(1)	5	1.67
Forward Selection	1907056726.337(2.5)	24.03(2.5)	0.964(2.5)	7.5	2.5
Stepwise regression	1907056726.337(2.5)	24.03(2.5)	0.964(2.5)	7.5	2.5
Total Rank	10	9	10		
Average Rank	2.5	3	2.5		

Interpretation: From the result above, ranking the techniques based on their performance on the selected criteria and obtaining their average rank. We observed that the direct search on t method had an average rank of 3, the forward selection and stepwise regression method both had a tie of 2.5 in their average, and the backward elimination method had an average rank of 1.67.

Using the All Possible Regression Method

Table-13: Ranked Performance of all significant subset models using all possible combination method

S/N	Sig. Variable combinations	Residual mean square	TOTAL VIF	Adjusted R^2	Total Rank	Average Rank
1	X_1X_6	17787436868(14)	2.946(5)	0.665(14)	33	11
2	X_1X_3	22527885642(16)	2.094(3)	0.576(16)	35	11.67
3	X_1X_5	2231501763(6)	3.044(6)	0.958(7)	19	6.33
4	X_3X_5	2825482498(7)	2.228(4)	0.947(8)	19	6.33
5	X_3X_6	18467980497(15)	2.012(2)	0.652(15)	32	10.67
6	X_3X_7	11150343492(11)	2.004(1)	0.790(12)	24	8
7	$X_1X_3X_4$	84042211437(9)	4.828(8)	0.842(10)	27	9
8	$X_1X_3X_6$	11459147086(12)	4.065(7)	0.784(13)	32	10.67
9	$X_1X_3X_7$	8876538238(10)	4.899(9)	0.833(11)	30	10
10	$X_2X_4X_6$	2145840442(5)	31.1(10)	0.960(5)	20	6.67
11	$X_2X_4X_7$	3446858873(8)	84.637(15)	0.935(9)	32	10.67
12	$X_2X_6X_7$	2108072114(4)	36.213(11)	0.959(6)	21	7
13	$X_2X_4X_5X_6$	1503272992(1)	61.676(12)	0.972(2)	15	5
14	$X_2X_5X_6X_7$	1742103059(2)	78.316(14	0.967(3)	19	6.33
15	$X_2X_4X_5X_7$	1868177615(3)	70.661(13)	0.965(4)	20	6.67
16	$X_1X_2X_5X_6X_7$	14370771439(13)	84.988(16)	0.973(1)	30	10

Interpretation: From the outcome, we note that all P-values for their F-statistics were lower than the preselected level of significance according to the Anova tables obtained in the different tests running on the possible combinations of the seven independent variables, thus suggesting that all models are adequate. But using the t-statistic, we note that, with the exception of the following models with combinations above, some of the models have a regression coefficient that is not important. The rating was used in order to obtain the appropriate model for this method; that is to say, based on the chosen parameters to be used in these analyses. Judgment is conducted in this way here; the model with the lowest residual mean square is the best and rating increases as the residual mean square increases, the one with the lowest VIF is the best and rank increases as VIF increases, while the model with the highest adjusted R^2 is named best for adjusted R^2, but rank increases with respect to the decrease in their respective adjusted R^2. And it is summed up as the best model for estimation is called the model combination with the lowest average rank value.

The best model using the all possible regression process, judging by the average rank, is given below using this technique as;

$$\mathrm{GDP} = 220550.460 + 10.593\mathrm{ASI} - 0.179\mathrm{TNI} + 10016.591\mathrm{OOTE} + 0.003\mathrm{VALTRANS}$$

With a R^2 value of 0.975; that is, 97.5% of the total variation in the GDP estimated value can be explained by the model. But taking a good look at the VIF of the model assumed it becomes difficult to make conclusion on this result, it is observed that the VIF is extremely large stating there is a strong presence of multicollinearity, which has a huge effect on precision and confident interval/level.

CONCLUSION AND RECOMMENDATION

The following inferences can be deducted from the study carried out in chapter four on the effect of the Nigerian stock market on its economic growth using the various variable selection techniques to obtain models called the best equation based on these techniques;

SUMMARY

Using direct search on t, ASI and OOTE, i.e. the All Share Index (variable 2) and the Openness of the Nigerian Trade Economy (variable 5) respectively, the model was left as the appropriate subset of regressors for estimating the development and growth of the Nigerian economy. Using the anova table, the model is found to be satisfactory and 95.8 percent of the total variance in the gross domestic Nigeria could be clarified. ASI, OOTE, VALTRANS, and TNI, i.e. the All-Share Index, Transparency of the Nigerian Trade Economy, Transaction Value, and Total New Issues respectively, are left in the model as the required subset regressors for the estimation using the backward elimination process of the Nigerian economy development and growth. The model using the anova table is found to be adequate and 97.5 percent of the total variance in Nigerian gross domestic product could be clarified. Using the Forward and stepwise regression process of selection (which is the modified method of the forward selection method). It is noted that both methods generated exactly the same outcome as the appropriate subset for the calculation, i.e. GMC, ASI, and OOTE. Using the anova table, the model is found to be acceptable and it could explain 96.8 percent of the overall difference in Nigeria's gross domestic product.

CONCLUSION

Using the mean rank method to determine the best model from the study in chapter four, it is observed that the backward elimination method offers the best equation for estimating the Nigerian gross domestic product with a mean rank of 1.67, which is the lowest mean rank. However, its inflation variance factor is extremely high, implying strong collinearity among some of the independent variables. These may really have influenced the signs of the regressor coefficients, even though the residual mean square tends to be the lowest relative to other methods of variable selection. As observed, using the all-possible combination approach as a control. The best equation for the estimate with a mean rank of 5 is the same as that obtained from the backward selection containing the ASI, TNI, VALTRANS, and OOTE as the suitable sub-set regressors, using the lowest average rank as the best model judgment. With a R^2 value of 0.975, that is, the model will explain 97.5 percent of the total variance in the expected value of GDP. However, taking a good look at the model's VIF presumed that it becomes difficult to conclude on this outcome, it is observed that the VIF is extremely large stating that there is a heavy multicollinearity presence, which has a significant impact on accuracy and confident interval. These contradicts Richard Lockhart’s conclusion that the four variable selection techniques gives the same subset model as best.

RECOMMENDATION

While the variance inflation factor of the regression coefficients was applied to the performance assessment criterion, the various variable selection strategies were used to account for the negligence of the impact of multicolinearity prior to analysis. To ensure that the multicollinearity issue is solved in order to fulfill the presumption of no association between the independent variables, it is advisable for researchers and students who want to conduct similar analysis, and researchers may also carry out their findings in regression analysis on other variable selection techniques. A research should also be performed on the condition that all four variable selection methods provide the same model of subset regressors as the best model for estimation or not.

REFERENCE

1. Osuji GA, Obubu M, Nwosu CA. Stock investment decision in Nigeria; A PC Approach. World. 2016 Jan;2(1).
2. Osuji GA, Okoro CN, Obubu M, Obiora-Ilouno HO. Effect of akaike information criterion on model selection in analyzing auto-crash variables. Int J Sci Basic Appl Res. 2016;26(1):98-109.
3. Cyprian A. Oyeka. An introduction to applied statistical methods, eighth Edition. 2009.
4. Efroymson D. Selection of variables in multiple regression part i. A review and evaluation. Int. Statist. Rev. 1960; 46, 1-19.
5. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association. 2001 Dec 1;96(456):1348-60.
6. Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research. 2003;3(Mar):1157-82.
7. Kira K and L Rendell. A practical approach to feature selection in D. Sleeman and P. Edwards, editors, international conference on machine learning. 2006; 368-377, Aberdeen.
8. Montgomery DC and Peck EA. Introduction to linear regression analysis, 2nd edition. 1991; 4: 271; 302.
9. Richard Lockhart. Variable selection Method: an introduction, Milano chemometric and QSAR research group-Dept. of Environmental sciences, University of Milano-Boccoca P.za della Scienza; 2002.
10. Selena NG. Variable Selection in predictive regressions, Department of Economics, Columbia University; 2012. 420W 118 st., MC 3308, New York, NY 10027.
11. Wei C. On Predictive Least Squares Principle, “Annals of Statistics. 1992; 20:1,1-42.
12. Obubu M, Konwe CS, Nwabenu DC, Omokri PA, Chijioke M. Evaluation of the contribution of Nigerian stock market on economic growth; Regression approach. European Journal of Statistics and Probability. 2016 Oct;4(5):11-27.
13. Guobadia Emwinloghosa Kenneth. Statistical Application of Regression techniques in Modeling Road Accidents in Edo State, Nigeria. Sch J Phys Math Stat. 2021 Jan 8(1): 14-18
14. Maxwell O, Happiness OI, Alice UC, Chinedu IU. An empirical assessment of the impact of Nigerian all share index, Market Capitalization, and Number of Equities on Gross Domestic Product. Open Journal of Statistics. 2018 May 9;8(3):584-602.
15. Lydia OI, Maxwell O, Aideniosa OF, Ifeanyi AC, Victor EU. On the Relative Potency of Aframomum Melegueta Extract on Albino Rats. 2019.
16. Obubu Maxwell, Oyafajo Oyindamola Abubarkri, Anyawu Ifeyinwa Fidelia, Olayemi Joshua I. Modeling Typhoid Mortality with Box-Jenkins Autoregressive Integrated Moving Average Models. Scholars Journal of Physics, Statistics, and Mathematics. 2019, 6 (3): 29–34.
17. Maxwell O, Mayowa BA, Chinedu IU, Peace AE. Modelling count data; a generalized linear model framework. Am J Math Stat. 2018;8(6):179-83.
18. Guobadia Emwinloghosa Kenneth, Ibeakuzi Precious Onyedikachi & Uadiale Kenneth Kevin, Short Term Modeling of the Nigerian Naira/United States Dollar Exchange Rate Using ARIMA Model. Sch J Phys Math Stat, 2021 Jan 8(1): 8-13.
19. Okereke OE, Bernard CB. Forecasting Gross Domestic Product In Nigeria Using Box-Jenkins Methodology. Journal of Statistical and Econometric methods. 2014;3(4):33-46.
20. Uddin MM. Causal relationship between agriculture, industry and services sector for GDP growth in Bangladesh: An econometric investigation. Journal of Poverty, Investment and Development. 2015;8.
21. Maxwell O, Happiness OI, Alice UC, Chinedu IU. An empirical assessment of the impact of Nigerian all share index, Market Capitalization, and Number of Equities on Gross Domestic Product. Open Journal of Statistics. 2018 May 9;8(3):584-602.
22. Obubu M, Konwe CS, Nwabenu DC, Omokri PA, Chijioke M. Evaluation of the contribution of Nigerian stock market on economic growth; Regression approach. European Journal of Statistics and Probability. 2016 Oct;4(5):11-27.
23. Ekhosuehi N, Kenneth GE, Kevin UK. The Weibull Length Biased Exponential Distribution: Statistical Properties and Applications. Journal of Statistical and Econometric Methods. 2020;9(4):15-30.
24. Guobadia Emwinloghosa Kenneth. A Statistical Outlook into the Distribution of Crimes in Nigeria Using Principal Component Analysis. Sch J Phys Math Stat. 2021 Jan 8(1): 1-7.
25. Guobadia Emwinloghosa Kenneth, Momoh Besiru, Kelvin, Oghogho Iyenoma. A Time Domain Approach to Modeling Nigeria’s Gross Domestic Product. Sch J Phys Math Stat. 2021 Jan 8(1): 19-28.
26. Guobadia Emwinloghosa Kenneth. Statistical Application of Regression techniques in Modeling Road Accidents in Edo State, Nigeria. Sch J Phys Math Stat. 2021 Jan 8(1): 14-18.