

# A Model Selection on Economic Variable in Nigeria

Muritala Abdulkabir<sup>1</sup>, Omuku Ikechukwu Joshua<sup>1</sup>, Raji Surajudeen Tunde<sup>2</sup>

<sup>1</sup>Statistics Department, Lens Polytechnic Offa, Offa, Nigeria

<sup>2</sup>Mathematics and Statistics Department, Federal Polytechnic Offa, Offa, Nigeria

## Email address:

kaybeedydx@gmail.com (M. Abdulkabir)

## To cite this article:

Muritala Abdulkabir, Omuku Ikechukwu Joshua, Raji Surajudeen Tunde. A Model Selection on Economic Variable in Nigeria. *Biomedical Statistics and Informatics*. Vol. 1, No. 1, 2016, pp. 13-18. doi: 10.11648/j.bsi.20160101.12

**Received:** September 11, 2016; **Accepted:** October 21, 2016; **Published:** December 12, 2016

---

**Abstract:** This study is on model selection on economic variable on gross domestic product in Nigeria, the data used for this study were extracted from National Bureau of Statistics (NBS), the statistical tool is multiple regression model and model selection to select the best model and in the variable and to evaluate and test GDP as a determinant which will capture the effect on economic variables. At the end of the analysis and findings it were concluded that Import value import value from the export, production, petroleum and consumption plays the most significant role in the company's market. It can be used as a tool to estimate the company's future market price.

**Keywords:** Gross Domestic Product, Multiple Regression, Model Selection, Variance Inflation Factor (VIF), Tolerance

---

## 1. Introduction

The responsibility shouldered by the government of any nation, particularly the developing nations, is enormous. The need to fulfil these responsibilities largely depends on the amount of revenue generated by the government through various means. Taxation is one of the oldest means by which the cost of providing essential services for the generality of persons living in a given geographical area is funded. Globally, governments are saddled with the responsibility of providing some basic infrastructures for their citizens. Functions or obligations the government may owe her citizens include but are not restricted to: stabilization of the economy, redistribution of income and provision of services in the form of public goods [1].

Taxation is a major source of government revenue all over the world and governments use tax proceeds to render their traditional functions, such as: the provision of goods, maintenance of law and order, defence against external aggression, regulation of trade and business to ensure social and economic maintenance [7]. The primary function of a tax system is to raise enough revenue to finance essential expenditures on the goods and services provided by government; and tax remains one of the best instruments to boost the potential for public sector performance and repayment of public debt [9]. A system of tax avails itself as a veritable tool that mobilizes a nation's internal resources

and it lends itself to creating an environment that is conducive for the promotion of economic growth [3]. Therefore, taxation plays a major role in assisting a country to meet its needs and promote self-reliance.

In Nigeria, tax revenue has accounted for a small proportion of total government revenue over the years compared with the bulk of revenue needed for development purposes that is derived from oil [6]. The serious decline in the prices of oil in recent times has led to a decrease in the funds available for distribution to the federal, state and local governments [2]. Consequently, dependence on oil as a particular or main source of revenue in Nigeria has become risky and not beneficial for sustainable economic growth. It is worse for Nigeria where there are fluctuations in prices in the oil market; thereby creating concerns amongst Nigerians and indeed the Nigerian government on the need to diversify the economy.

Naturally, and globally, there is a paradigm shift to taxation revenue as an alternative source of revenue. Nigeria is not an exception. The machinery and procedures for implementing a good tax system in Nigeria are inadequate; hence tax evasion and avoidance of the self-employed individuals and organizations whose data base is not captured in the relevant tax authority's data system [8]. The need for the government to generate adequate revenue from internal sources has therefore become a matter of extreme urgency and importance [2]. The desire of any government to

maximize revenue from taxes collected from tax payers cannot be overemphasized.

This is because, as it well-known, the importance of tax lies in its ability to generate revenue for the government, influence the consumption trends and grow and regulate economy through its influence on vital aggregate economic variables [2]. In the light of the above, and in broad spectrum, this paper examines the impact of indirect taxes on economic growth in Nigeria. This topic is formed and informed against the backdrop of the need for a paradigm shift to indirect taxation in the face of the dwindling oil prices and the relative paucity of studies, using inflation-factored GDP in Nigeria. To this end, and in order to set a direction for

This paper, aim to evaluate and test GDP as a determinant which will capture the effect on economic variables and identify the best model in the fitted model using the model selection approach

## 2. Methodology

### *Regression analysis:*

Regression analysis is a technique used in statistics for investigating and modeling the relationship between variables [4].

### *Simple linear regression:*

Simple linear regression is a model with a single regressor  $x$  that has a relationship with a response  $y$  that is a straight line. This simple linear regression model can be expressed as

$$y = \beta_0 + \beta_1 x_1 + \ell_i$$

where the intercept  $\beta_0$  and the slope  $\beta_1$  are unknown constants and  $\varepsilon$  is a random error component.

### *Multiple linear regression:*

If there is more than one regressor, it is called multiple linear regression. In general, the response variable  $y$  may be related to  $k$  regressors,  $x_1, x_2, \dots, x_k$ , so that

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

### *Least Squares Estimation:*

The method of least squares is used to estimate  $\beta_0, \beta_1, \dots, \beta_k$ . That is, we estimate  $\beta_0$  and  $\beta_1$  so that the sum of the squares of the differences between the observations  $y_i$  and the straight line is a minimum [4].

### *R-squared:*

R-squared is a measure in statistics of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determinations for multiple regression. It is the percentage of the response variable variation that is explained by a linear model.

$$R - \text{Square} = \frac{\text{Explained Variation}}{\text{Total variation}}$$

R-squared is always between 0 and 100%. 0% means the model explains none of the variability of the response data

around its mean. 100% indicates that the model explains all the variability of the response data around its mean. Generally, the higher the R-squared, the better the model fits the data (Frost, 2013).

### *Analysis of variance (ANOVA):*

Analysis of variance (ANOVA) is a collection of statistical models used in order to analyze the differences between group means and their associated procedures. In the ANOVA setting, the observed variance in a particular variable is partitioned into components attributable to different sources of variation. The following equation is the Fundamental Analysis-of-Variance Identity for a regression model

$$\sum_{i=1}^n (y - \bar{y})^2 = \sum_{i=1}^n (\hat{y} - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y})^2$$

$$SST = SSR + SSRes$$

### *Statistical Hypothesis:*

Statistical hypothesis are statements about relationships. The statistical hypothesis testing is the use of statistics to determine the probability that a given hypothesis is true [5]. The null hypothesis is denoted by

$H_0$ . The alternative hypothesis is the negation of the null hypothesis, denoted by

$H_1$  or  $H_a$

### *Testing Significance of Regression:*

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$

$H_1: \text{at least one } \beta_i \neq 0$

The hypotheses are related to the significance of regression. Failing to reject  $H_0$  implies that there is no linear relationship between  $x$  and  $y$ . On the other hand, if  $H_0$  is rejected, it implies that at least one  $\beta_i$  show a significant relationship to  $y$

### *F-test:*

An F-test is a statistical test in which the test statistic is based on the F-distribution under the null hypothesis. It is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled. In this research, the F-test is used to test the significance of the model.

The test statistics  $F_0$  can be computed by  $\frac{MS_R}{MS_E}$  follows the

$F_{k, n-k-1}$  distribution. Reject  $H_0$ , if  $F_0 > F_{k, n-k-1}$ . The test statistic  $F_0$  can usually be obtained from the ANOVA table

### *Test on Individual Regression Coefficients (t Test):*

The t-test is used to check the significance of individual regression coefficients in the multiple linear regression model. Adding a significant variable to a regression model makes the model more effective, while adding an unimportant variable may make the model worse. The hypothesis statements to test the significance of a particular regression coefficient,  $\beta_j$ , are:

$H_0: \beta_j = 0$

$H_1: \beta_j \neq 0$

The test statistic for this test has the t-distribution:

$$T_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \text{ where the standard error, } se(\hat{\beta}_j), \text{ is obtained.}$$

One would fail to reject the null hypothesis if the test statistic lies in the acceptance region:  $-t_{\alpha/2, n-2} < T_0 < t_{\alpha/2, n-2}$

This test measures the contribution of a variable while some other variables are included in the model

*P-value:*

P-value or calculated probability is the estimated probability of rejecting the null hypothesis (H0) of a study question when that hypothesis is true.

*The variance Inflation Factor (VIF):*

The variance inflation factor for each term in the model measures the combined effect of the dependences among the regressors on the variance of the term. Practical experience indicates that if any of the VIFs exceeds 5 or 10, it is an indication that the associated regression coefficients are poorly estimated because of multicollinearity

*Model Selection:*

1.) Forward Selection: This procedure begins with the assumption that there is no regressor in the model other than the intercept. The goal is to find an optimal subset by inserting regressors into the model one at a time.

2.) Backward Elimination: This procedure is the opposite approach from the forward selection. First, we begin with a full model with K candidate regressors. Then, the partial F statistic (or a t statistic) is computed for each regressor. If the regressor with the smallest partial F or t value is less than the preselected F value, that regressor is removed from the model. Fit model with K-1 predictors and the procedure is repeated.

3.) Stepwise Regression: It is a method that allows moves in either direction, dropping or adding variables at the various steps. It combines both forward selection and backward elimination. We perform two steps in forward selection and a backward step. Then, perform another forward step and another backward step. We continue until no action can be taken in either direction

*Residuals:*

The difference between the observed value of the dependent variable (y) and the predicted value ( $\hat{y}$ ) is called

theresidual(e). Each data point has one residual.

$$\text{Residual} = \text{Observed value} - \text{Predicted value}$$

$$e = y - \hat{y}$$

Both the sum and the mean of the residuals are equal to zero. That is,

$$\sum e = 0 \text{ and } \bar{e} = 0 [4].$$

*Residual Diagnostics:*

A residual plot is a graph that shows the residuals on the vertical axis and the independent variable on the horizontal axis. If the points in a residual plot are randomly dispersed around the horizontal axis, a linear regression model is appropriate for the data; otherwise, a non-linear model is more appropriate.

*Checking normality:*

*Histogram:*

The Histogram of the Residual can be used to check whether the variance is normally distributed. A symmetric histogram as shown in the figure 1 below, which is evenly distributed around zero, indicates that the normality assumption is likely to be true [10]. The typical "bell-curve" is the ideal indication as to normality. When this cannot be obtained, a symmetrical histogram is sufficient.

### 3. Analysis

*Table 1. Descriptive statistics.*

Descriptive Statistics			
	Mean	Std. Deviation	N
Market price	224200.1333	184297.60461	15
Export	57251.4667	29926.93610	15
Petroleum	51906.6667	28290.41561	15
Import	33395.9333	20508.49817	15
Production	2030.1600	189.77164	15
Consumption	269.0867	60.73901	15

The descriptive statistics in table 1 shows the summary of market price, export, petroleum, import, production and consumption with average and standard deviation with sample size of 15

*Table 2. Model coefficient Ordinary Least Square.*

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients		Sig.
		B	Std. Error	Beta	t	
1	(Constant)	-215284.381	149631.884		-1.439	.184
	Export	6.145	2.591	.998	2.372	.042
	Petroleum	-5.890	2.966	-.904	-1.986	.078
	Import	4.428	.994	.493	4.454	.002
	Production	-61.378	51.157	-.063	-1.200	.261
	Consumption	1375.489	283.863	.453	4.846	.001

a. Dependent Variable: Market price

$$\text{Market Price} = -215284.381 + 6.145\text{Export} - 5.890\text{Petroleum} + 4.428\text{Import} - 61.378 + 1375.489\text{consumption}$$

*Test of Hypothesis*

H<sub>0</sub>:  $\beta_1 = 0$  (The linear model is inadequate).

H<sub>1</sub>:  $\beta_1 \neq 0$  (The linear model is adequate).

Level of significance  $\alpha=0.05$

## Computation

**Table 3.** Model summary.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.994 <sup>a</sup>	.987	.980	26149.66007

a. Predictors: (Constant), Consumption, Production, Export, Import, Petroleum

**Table 4.** Analysis of Variance (ANOVA).

ANOVA <sup>a</sup>						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	469364256419.686	5	93872851283.937	137.280	.000 <sup>b</sup>
	Residual	6154242494.047	9	683804721.561		
	Total	475518498913.733	14			

a. Dependent Variable: Market price

b. Predictors: (Constant), Consumption, Production, Export, Import, Petroleum

Decision rule: Reject Ho if P-value is less than level of significant  $\alpha=0.05$  otherwise do not reject Ho

Decision: Ho is rejected

Conclusion: It can be concluded the model is adequate

## 4. Model Selection

### Forward Selection Method

**Table 5.** Model coefficient.

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-63778.521	27238.050		-2.342	.036
	Import	8.623	.701	.960	12.293	.000
	(Constant)	-276539.812	52673.251		-5.250	.000
2	Import	5.939	.775	.661	7.659	.000
	Consumption	1123.797	261.839	.370	4.292	.001

a. Dependent Variable: Market price

The final regression equation from the above model which indicate the best model from the Variance Inflation Factor is  
 Market price = -63778.521 + 8.623 Import

#### Backward Elimination:

The market price is perfectly explained by all of the variables combined, so the standard error is zero. The test of statistics is undefined when regressing market price with the five (5) main factors on linear regression. In order to further use Backward Elimination in pinpointing a factor's contribution to the market price of a company, every variable, but one (K-1) are selected to generate the initial model for this method.

**Table 6.** Model Coefficient.

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-215284.381	149631.884		-1.439	.184
	Export	6.145	2.591	.998	2.372	.042
	Petroleum	-5.890	2.966	-.904	-1.986	.078
	Import	4.428	.994	.493	4.454	.002
	Production	-61.378	51.157	-.063	-1.200	.261
	Consumption	1375.489	283.863	.453	4.846	.001
	(Constant)	-384224.238	51731.458		-7.427	.000
2	Export	7.565	2.354	1.228	3.213	.009
	Petroleum	-7.702	2.608	-1.182	-2.953	.014
	Import	4.626	1.002	.515	4.618	.001
	Consumption	1563.062	242.084	.515	6.457	.000

a. Dependent Variable: Market price

The model for the backward selection is

$$y = -215284.381 + 6.145 \text{Export} - 5.890 \text{Petroleum} - 61.378 \text{Production} + 1375.48 \text{Consumption}$$

Table 7. Stepwise Selection.

Coefficients <sup>a</sup>		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model		B	Std. Error	Beta		
1	(Constant)	-63778.521	27238.050		-2.342	.036
	Import	8.623	.701	.960	12.293	.000
2	(Constant)	-276539.812	52673.251		-5.250	.000
	Import	5.939	.775	.661	7.659	.000
	Consumption	1123.797	261.839	.370	4.292	.001

a. Dependent Variable: Market price

The model equation from the above model which indicate the best model from the Variance Inflation Factor is Market price = -63778.521 + 8.623 Import which is the same as the forward selection

Findings of the best model selection using simple linear regression (i.e Market price vs Import)

Table 8. Model Coefficient.

Coefficients <sup>a</sup>		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model		B	Std. Error	Beta		
1	(Constant)	-63778.521	27238.050		-2.342	.036
	Import	8.623	.701	.960	12.293	.000

a. Dependent Variable: Market price

The model is Market price = -63778.521 + 8.623 Import

Table 9. Test of Hypothesis.

ANOVA <sup>a</sup>		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	437853853802.784	1	437853853802.784	151.126	.000 <sup>b</sup>
	Residual	37664645110.950	13	2897280393.150		
	Total	475518498913.733	14			

a. Dependent Variable: Market price

b. Predictors: (Constant), Import

$H_0: \beta_1 = 0$  (The relationship between Market price and value of Import is not significant).

$H_1: \beta_1 \neq 0$  (The relationship between Market price and value of Import is significant).

Level of significance  $\alpha = 0.05$

Computation

Decision rule: Reject  $H_0$  if P-value is less than level of significance  $\alpha = 0.05$  otherwise do not reject  $H_0$

Decision:  $H_0$  is rejected

Conclusion: from the result of the analysis it can be concluded that there is relationship between Market price and value of Import is significant

Therefore the equation that best depict the market price is

$$\text{Market price} = -63778.521 + 8.623 \text{Import}$$

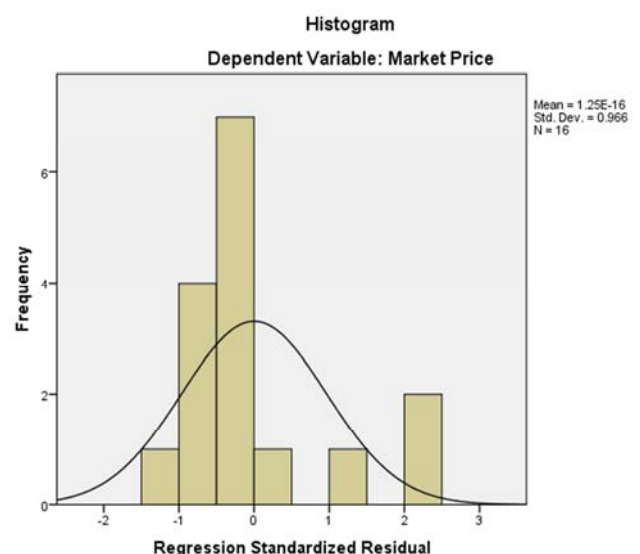


Fig. 1. Normality plot.

The histogram (as shown in figure 1) appears to be symmetric, which seems to depict normality.

## 5. Conclusions

Based on the result of the analysis, model selection method is processed; the consensus has shown that only the Import plays a statistically significant role in the market of the company.

After the significant category is found deeper analysis is conducted of the import category. Each individual of the independent (i.e. export, petroleum, import, production and consumption) is then regressed on the market price. After running these additional models, it was discovered that the import value from the export, production, petroleum and consumption remained the most significant model.

Futhermore, it can be concluded that Import value import value from the export, production, petroleum and consumption plays the most significant role in the company's market. It can be used as a tool to estimate the company's future market price.

---

## References

- [1] Abiola, J., & Asiweh, M. (2012). Impact of tax administration on government revenue in a developing economy- a case study of Nigeria. *International Journal of Business and Social Science*. 3(8), 99-113.
- [2] Afubero, D., & Okoye, E. (2014). The impact of taxation on revenue generation in Nigeria. A study of Federal Capital Territory and selected states. *International Journal of Public Administration and Management Research*. 2(2), 22-47.
- [3] Ayuba, A.J. (2014). Impact of non-oil revenue on economic growth: the Nigerian perspective. *International Journal of Finance and Accounting*. 3(5), 303-309. Berman, H. (n.d.). Residual Analysis in Regression. *Stat Trek*. Retrieved from <http://stattrek.com/regression/residual-analysis.aspx>
- [4] Douglas Montgomery, Peck, E., & Vinning, G. (2012). *Introduction to Linear Regression Analysis* (5th ed.). Wiley. *Experiment Design and Analysis Reference*. (n.d.). ReliaSoft. Retrieved from [http://reliawiki.org/index.php/Experiment\\_Design\\_and\\_Analysis\\_Reference](http://reliawiki.org/index.php/Experiment_Design_and_Analysis_Reference)
- [5] Iyanaga, S., & Kawada, Y. (1980). *Statistical Estimation and Statistical Hypothesis Testing*. (Vol. Appendix A, Table 23). Cambridge, MA: MIT Press.
- [6] Otu, O. H., & Adejumo, T. O. (2013). The effects of tax revenue on economic growth in Nigeria (1970-2011). *International Journal of Humanities and Social Science invention*. 2(6), 16-26.
- [7] Edame, G. E., & Okoi, W. W. (2014). The impact of taxation on investment and economic development in Nigeria. *Academic Journal of Interdisciplinary Studies*. 3(4). 209-218.
- [8] Fasoranti, M. M. (2013). Tax productivity and economic growth. *Loem Journal of Business and Economics*. 1(1), 1-10.
- [9] Okoye, P.V.C., & Ezejiofor, R. (2014). The impact of e-taxation on revenue generation in Enugu, Nigeria. *International of Advanced Research*. 2(2), 449-458.
- [10] Residual Analysis. (n.d.). *DePaul University*. Retrieved from <http://facweb.cs.depaul.edu/sjost/csc423/documents/resid-anal.htm>