GSJ: Volume 8, Issue 1, January 2020, Online: ISSN 2320-9186
www.globalscientificjournal.com

# A MODEL OF SOME NIGERIA SOCIO-ECONOMIC PROBLEMS: MULTIPLE REGRESSION APPROACH

AKPOJARO, O.O AND ONWUBUYA, M. N.
Department of Statistics, Delta State Polytechnic, Otefe-Oghara
Email: owenssmith@gmail.com, +234(0)7067000466

## ABSTRACT

*A model of some Nigeria socio-economic problems: multiple regression approach is a research aimed at employing regression diagnostics in solving Nigeria socio-economic problems. A secondary data was collected via the World Bank Golden Jubilee edition and Fact-Fish Data Base in the International Monetary Fund (IMF)bulletin, with a sample size of 12 years ranging from 2005-2016. The data set was analyzed with SPSS 20.0 at 5% level of significance.It was discovered that the model PPP GDP = 1985.365 – 34.862 Unemployment rate $x_1$ – 14.741 inflation rate $x_2$ + 0.027 electricity consumption $x_3$ was adequate; this was seen via the ANOVA, coefficient of determination $R^2$ = 71%. It was also observed that the data set follows or met the regression assumption of linearity, normality and residual mean which was zero. Conclusion was drawn.*

***Keywords;Multiple Regression, Model Adequacy, Residual, Socio-economics, Coefficient of Determination***

## 1.0 INTRODUCTION

A technique that allows us to predict someone's score on one variable on the basis of their scores on several other variables, an example might help. Suppose we were interested in predicting how much an individual's enjoys their job. Variables such as salary, extent of academic qualifications, age, sex, number of years in full-time employment and socio-economic status might all contribute. Much of scientific studies are directed towards discovering the form of relationships between variable, and predicting the values of a variable from some functional relationship is one of the most important areas of Applied Statistics.

Regression analysis is a statistical tool which helps to predict one variable from the other variables, on the basis of assumed nature of the relationship between the variables. It's a tool that is used to measure the linear relationships between two or more variables.

The variables being predicted is usually to all unknown or dependent variable, because its value are dependent on the values of the other variable or variables variously referred to as the independent variable, explanatory variable, predictor or predetermined variable. The values of these later variables are fixed and controlled only by the investigator.

Multiple regression is a statistical techniques job satisfaction. If we collected data on all of these variables, perhaps by surveying a few hundred members of the public, we would be able to see how many and which of these variables gave rise to the most accurate prediction of job satisfaction. We might find that job satisfaction is most accurately predicted by type of occupation, salary and years in full-time employment, with the other variables not helping us to predict job satisfaction.

When using multiple regression in psychology, many researchers use the term "independent variables" to identify those variables that they think will influence some other "dependent variable". We prefer to use the term "predictor variables" for those variables that may be useful in predicting the scores on another variable that we call the "criterion variable". Thus, in our example above, type of occupation, salary and years in full-time employment would emerge as significant predictor variables, which allow us to estimate the criterion variable – how satisfied someone is likely to be with their job. As we have pointed out before, human behavior is inherently noisy and therefore it is not possible to produce totally accurate predictions, but multiple regression allows us to identify a set of predictor variables which together provide a useful estimate of a participant's likely score on a criterion variable.

## SCOPE OF THE STUDY

Socio-economic is an aspect of economics which looks at the life of people socially and economically. Since this indicators are many, this research will focus on sampled sector such as electricity consumption per capita in Nigeria, purchasing power parity gross domestic product PPP GDP, inflation rate and unemployment rate in Nigeria from 2005 to 2016.

## OBJECTIVES OF THE STUDY

The overall aim of this research is the application of multiple regression in modeling socio-economic data in Nigeria. The specific objectives are to;

i. Use the multiple regression technique to model key socio-economic indicator data set.
ii. Access the socio-economic impact on the Nigeria economy
iii. Check if the socio economic data applied met the regression assumption.

## LITTERATURE REVIEW

Chatterjeeand Hadi,(2006), regression analysis is a conceptually method for investigating functional relationships among variables. A real estate appraiser may wish to relate the sale price of a home from selected physical characteristics of the building and taxes (local, school, county) paid on the building. We may wish to examine whether cigarette consumption is related to various socioeconomic and demographic variables such as age, education, income, and price of cigarettes. The relationship is expressed in the form of an equation or a model connecting the response or dependent variable and one or more explanatory or predictor variables. In the cigarette consumption example, the response variable is cigarette consumption (measured by the number of packs of cigarette sold in a given state on a per capita basis during a given year) and the explanatory or predictor variables are the various socioeconomic and demographic variables.

Abrams (1999), regression analysis is used when you want to predict a continuous dependent variable from a number of independent variables. If the dependent variable is dichotomous, then logistic regression should be used.

Mudbeck (2012), multiple linear regression is a means to express the idea that a response variable, Y, varies with a set of independent variables $X_1$, $X_2$. . .$X_m$. The variability that *y* exhibits has two components: a systematic part and a random part. The systematic variation of *y* can be modeled as a function of the *X*variables. The model relating *Y*to $X_1$, $X_2$. . .$X_m$is called the regression equation. The random part takes into account the fact that the model does not exactly describe the behavior of the response.

Formally, multiple linear regression fits a response variable *y* to a function of regressor variables and parameters. The general linear regression model has the form;

$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m + \varepsilon$

Where;

*Y* is the response, or dependent variable,$\beta_0$ , $\beta_1$. . . $\beta_m$are unknown parameters (constant).

$X_1$, $X_2$. . .$X_m$are the regressor or independent variables.

ε is a random error term.

Least squaresis a technique used to estimate the parameters based on a set of observed values of these variables. The goal is to find estimates of the parameters $\beta_0$, $\beta_1$ . . . $\beta_m$that minimize the sum of the squared differences between the actual *y* values and the values of *y* predicted by the equation. These estimates are called the least-squares estimates, and the quantity minimized is called the error sum of squares.

Typically, you use regression analysis to do the following:

i. obtain the least-squares estimates of the parameters
ii. estimate the variance of the error term
iii. estimate the standard error of the parameter estimates
iv. test hypotheses about the parameters
v. calculate predicted values using the estimated equation
vi. Evaluate the fit or lack of fit of the model.

Miranda andChing, (2000), opiated that multiple regression is a statistical technique that allows us to predict someone's score on one variable on the basis of their scores on several other variables. An example might help. Suppose we were interested in predicting how much an individual enjoys their job. Variables such as salary, extent of academic qualifications, age, sex, number of years in full-time employment and socioeconomic status might all contribute towards job satisfaction. If we collected data on all of these variables, perhaps by surveying a few hundred members of the public, we would be able to see how many and which of these variables gave rise to the most accurate prediction of job satisfaction. We might find that job satisfaction is most accurately predicted by type of occupation, salary and years in full-time employment, with the other variables not helping us to predict job satisfaction.

Stockburger(2010)is of the opinion that multiple regression is an extension of simple linear regression in which more than one independent variable (X) is used to predict a single dependent variable (Y). The predicted value of Y is a linear transformation of the X variables such that the sum of squared deviations of the observed and predicted Y is a minimum. The computations are more complex, however, because the interrelationships among all the variables must be taken into account in the weights assigned to the variables. The interpretation of the results of a multiple regression analysis is also more complex for the same reason.

With two independent variables the prediction of Y is expressed by the following equation: $\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i}$

Note that this transformation is similar to the linear transformation of two variables discussed in the previous chapter except that the w's have been replaced with b's and the $X'_i$ has been replaced with $Y'_i$. The "b" values are called regression weights and are computed in a way that minimizes the sum of squared deviations:

$$\sum_{i=1}^{N}(Y_i - Y'_i)^2$$

In the same manner as in simple linear regression, the difference is that in simple linear regression only two weights, the intercept ($b_0$) and slope ($b_1$), were estimated, while in this case, three weights ($b_0$, $b_1$, and $b_2$) are estimated.

## ASSUMPTIONS OF THE MODEL

The following are the assumptions of regression analysis given below

i. The relationship between x and y is linear or can at least be approximated by a straight line.
ii. Y depends on X
iii. The errors are independently distributed. That is, the errors are uncorrelated with each other. $E(e_ie_j) = 0$, for $i \neq j$
iv. The error term are normally and identically distributed with mean equal to zero and variance $\sigma^2$. That is $e \sim NID(0, \sigma^2)$
v. The x is a fixed variable and y is a random variable. (Jude e'tal 2005)

## WHEN TO USE MULTIPLE REGRESSION

1. You can use this statistical technique when exploring linear relationships between the predictor and criterion variables – that is, when the relationship follows a straight line. (To examine non-linear relationships, special techniques can be used.)

2. The criterion variable that you are seeking to predict should be measured on a continuous scale (such as interval or ratio scale). There is a separate regression method called logistic regression that can be used for dichotomous dependent variables (not covered here).

3. The predictor variables that you select should be measured on a ratio, interval, or ordinal scale. A nominal predictor variable is legitimate but only if it is dichotomous, i.e. there are no more than two categories. For example, sex is acceptable (where male is coded as 1 and female as 0) but gender identity (masculine, feminine and androgynous) could not be coded as a single variable. Instead, you would create three different variables each with two categories (masculine/not masculine; feminine/not feminine and androgynous/not androgynous). The term dummy variable is used to describe this type of dichotomous variable.

4. Multiple regression requires a large number of observations. The number of cases (participants) must substantially exceed the number of predictor variables you are using in your regression. The absolute minimum is that you have five times as many participants as predictor variables. (Kemstone, e'tal, 2011)

Akpojaro O.O. and Omokaro B. E. (2020), a common problem in the sciences and industry is to compare several treatments to determine which produce a superior outcome. For instance, suppose a manufacturer wants to examine the effect on sales due to package design. A reasonable way to proceed is to select a group of stores with comparable sales volumes and randomly and independently assign each store to carry one of the package designs to be tested.Assume several stores carry each package design andconditions that could affect sales,

such as price, shelf space, and promotional efforts, are the same for all stores. When thedata gathering is finished, it may turn out that one package design is clearly superior to the others. In this case, there is noneed for statistical analysis of the data. On the other hand, the average or mean sales for each design may be close enoughthat it is not easy to decide whether their differences are real or are due to the inherent variation in sales among the stores.A common method for investigating such differences is called analysis of variance, often abbreviated to ANOVA

## NIGERIA FROM SOCIO-ECONOMIC PERSPECTIVE

According to Foluke and Remi (2016), Nigeria contains an interesting paradox – due to its oil reserve, it is Africa's largest economy, but 70% of Nigerians live under the international poverty line of US $1.25 a day. The majority of Nigerians live without indoor plumbing and electricity, and many lack access to clean water and sanitation facilities. This lack of infrastructure helps explain Nigeria's high infant mortality rate – ranking the ninth highest of all countries. A few more statistics give a brief outline of the birth to death experiences: a full 24% Nigerians age 5-14 are involved in child labor; 70% of Nigerians between the ages of 15-24 have attained literacy (with many more males than females able to read and write); and life expectancy in Nigeria is 52.1 years.

Ogla (2001), Socio-economic is a field of study that examines social and economic factors to better understand how the combination of both influences something. Social economics is a branch of economics that focuses on the relationship between social behavior and economics, and it examines how social norms, ethics and other social philosophies that influence consumer behavior shape an economy, and uses history, politics and other social sciences to examine potential results from changes to society or the economy.

## METHODOLOGY

### Data Source and Method

Data was collected secondarily from the World Bank, Central Bank of Nigeria (CBN) golden jubilee edition, and fact-fish data base.

## TARGET POPULATION

As an indicator, the target sectors this research focused on are the inflation rate, unemployment rate, purchasing power parity gross domestic product PPP GDP and the electricity consumption from 2005-2016

## SAMPLING TECHNIQUE AND SIZE

The sampling technique used in this research work is the stratified sampling and the sample size of 13 years will be observed.

## MODEL SPECIFICATION

Consider the multiple regression model given as;

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ ……..1

Where,

Y = PPP GDP

$X_1$ = Inflation rate

$X_2$ = Unemployment rate

$X_3$ = Electricity consumption rate

$\beta_0, \beta_1, \beta_2$ and $\beta_3$ are model parameter

$\varepsilon$ is the stochastic disturbance.

$\beta_j = (X^I X)^{-I}. X^1 Y$……...…….2

The above equation(2) is the matrix approach to multiple linear regression which helps in estimating the model parameters.

| ***YEAR*** | ***PPP GDP (Y)*** | ***INFLATION RATE ($X_1$)*** | ***UNEMPLOYMENT RATE ($X_2$)*** | ***Electricity consumption per capita (kWh) ($X_3$)*** |
|---|---|---|---|---|
| 2016 | 5439.00 | 15.40 | 5.00 | 144000.00 |
| 2015 | 5671.00 | 9.00 | 4.30 | 142000.00 |
| 2014 | 5672.00 | 8.00 | 4.80 | 156000.00 |
| 2013 | 5479.00 | 8.50 | 7.10 | 150000.00 |
| 2012 | 5339.00 | 12.20 | 7.60 | 136000.00 |
| 2011 | 5259.00 | 10.80 | 7.30 | 120000.00 |
| 2010 | 5150.00 | 13.70 | 7.30 | 127000.00 |
| 2009 | 4905.00 | 12.50 | 7.20 | 138000.00 |
| 2008 | 4711.00 | 11.60 | 7.20 | 111000.00 |

| | | | | |
|---|---|---|---|---|
| 2007 | 5552.00 | 5.40 | 7.10 | 129000.00 |
| 2006 | 4374.00 | 8.20 | 7.10 | 123000.00 |
| 2005 | 4149.00 | 17.90 | 7.10 | 101000.00 |

Source: World Bank, Central Bank of Nigeria CBN and Fact-Fish Database

2.0 The data was analyzed using SPSS 20. The outputs are presented in appendix

**DISCUSSION OF RESULT**

From the analysis above, we derived a multiple regression model in table (4) given as;

PPP GDP = 1985.365 – 34.862 Unemployment rate $x_1$ – 14.741 inflationrate $x_2$ + 0.027 electricity consumption $x_3$

The above estimated multiple regression model is interpreted thus; the mean value of PPP GDP when unemployment rate is zero is 1985.365. For any additional unit change in PPP GDP, unemployment rate will decrease by 34.862, given that inflation rate and electricity consumption is held constant. For any additional unit change in PPP GDP, inflation will decrease by 14.741 provided that electricity consumption and unemployment is held constant. Also, for additional unit change in PPP GDP, electricity consumption will increase by 0.027, provided that unemployment rate, and inflation rate is held constant. It is seen from table (3)in the ANOVA table that the model is adequate since the probability value 0.009 is less than the level of significance 0.05. This can also be revealed in table (1), the coefficient of determination $r^2$ = 71% which indicate a model adequacy and supports the analysis of variance that the model is adequate. The residual table(5),revealed that the mean of the error term is 0. The histogram and the normal plot indicate that the data set met the regression assumption of linearity and normality.

**SUMMARY AND CONCLUSSION**

In conclusion, the multiple regression analysis is highly plausible in handling socio-economic data in Nigeria. From the model, it was discovered that there is a decrease in inflation rate, increase in electricity consumption per capita kWh and a decrease in unemployment rate in Nigeria over the years observed.

**RECOMMENDATION**

**(1)** the multiple regression analysis is hereby recommended for all scholars as it's a powerful tool for modeling socio-economic data both in Nigeria and the world

**(2)** I also recommend for further research via the regression diagnostics the contribution of each variable in the model performances.

**REFFERENCES**

Akpojaro O. O. and Omokaro B. E. (2020).Multiple Comparisons of Means UsingBonferroni Approach.*Innovative Journal of Science*. Vol. 1, No. 1, pp. 06-12

Abraham,P. (2001).*Statistical standards, advanced measure andapplication, second edition*, McCoy ltd, Brooklyn.

David W.S. (2010).*Pegression Analysis and Its Behavior to AppliedSciences,* Gritz publishers Inc, Denver

Deborah, R. A. (1999).*Multivariate statistics, department of psychology,* Phareal publishers inc, New York.

Foluke, B.O. and Remi J.A, (2016). "Beyond the Violence: Nigeria from a Socio-economic Perspective", lekanje publishing ltd, Ibadan

Jose, M. A. (2012).*First Course to Least Square Regression,* Jason's Lobmann Publishers, Nairobi

Jose Mudbeck A. (2012).*Regression Approach to Sociological Problems,*Mcgramm ltd, Oslo

Eze J. I, Obigbu M. E. and Eze E. J. .N. (2005).*Statistics and Quantitative Methods For Construction and Business Managers,* the Nigerian institute of building, lagos

Mirianda H. e'tal, (2000). "A Statistical Analysis of life Expectancy Across Countries Using Multiple Regression" Tonielsinc, California.

Ogla K. (2001). Socio-economic Performance in any Growing Economy; AnIndicator to Foreign Investors, Jorddein-freider Publishing Company, New Orleans.

Samprit C. and Ali S. H, (2006).Statistical Methods for Sciences, Harper-Collins publishers, Delhi



## Appendix

Table (2)**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | R Square Change | F Change | df1 | df2 | Sig. F Change |
| 1 | .840[a] | .706 | .609 | 409.36485 | .706 | 7.217 | 3 | 9 | .009 |

Table (3)**ANOVA**

| Model | Sum of Squares | Df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Regression | 3628354.559 | 3 | 1209451.520 | 7.217 | .009[b] |
| Residual | 1508216.211 | 9 | 167579.579 | | |
| Total | 5136570.769 | 12 | | | |

Table (4) **Coefficients**

| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|
| | B | Std. Error | Beta | | |
| (Constant) | 1985.365 | 2070.841 | | .959 | .363 |
| UNEMPLOYMENT RATE | -34.862 | 39.082 | -.189 | -.892 | .396 |
| INFLATION RATE | -14.741 | 128.270 | -.025 | -.115 | .911 |
| ELECTRICITY | .027 | .010 | .717 | 2.871 | .018 |

Table (5)**Residuals Statistics**

| | Minimum | Maximum | Mean | Std. Deviation | N |
|---|---|---|---|---|---|
| Predicted Value | 4013.1167 | 5893.1846 | 5019.3077 | 549.87533 | 13 |
| Residual | -646.56439 | 482.77457 | .00000 | 354.52036 | 13 |
| Std. Predicted Value | -1.830 | 1.589 | .000 | 1.000 | 13 |
| Std. Residual | -1.579 | 1.179 | .000 | .866 | 13 |

a. Dependent Variable: PP GDP

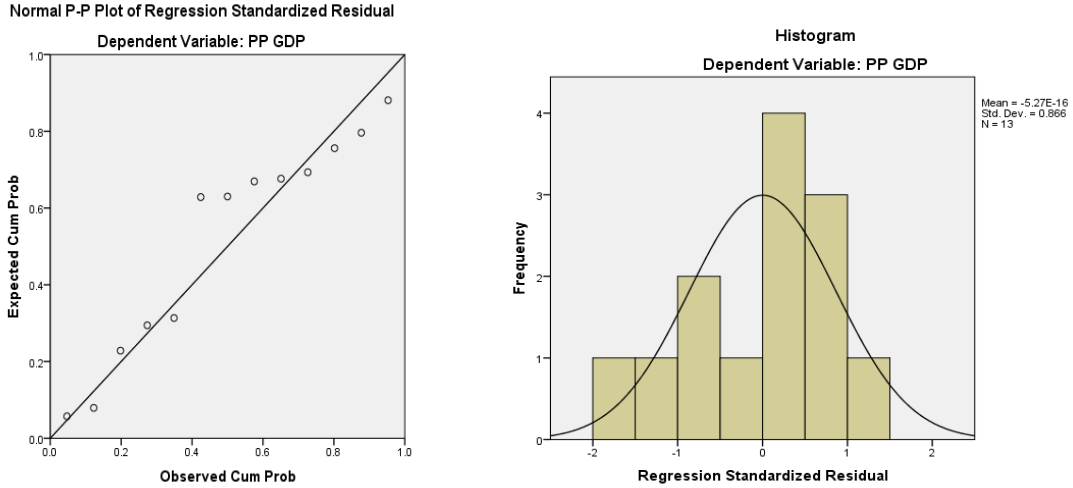


Figure (1)shows the histogram with normal distribution and normal probability plot standardized regression residual.