

---

# Statistical Analysis of Electricity Generation in Nigeria Using Multiple Linear Regression Model and Box-Jenkins' Autoregressive Model of Order 1

Imo Enoidem Ebukanson<sup>1</sup>, Chukwu Benedict Chidi<sup>2</sup>, Abode Innocent Iriaoghuan<sup>2</sup>

<sup>1</sup>Department of Electrical/Electronic and Computer Engineering, Faculty of Engineering, University of Uyo, Uyo, Nigeria

<sup>2</sup>Department of Electrical/Electronic Engineering Imo State Polytechnic, Umuagwo, Owerri, Nigeria

## Email address:

iniobongetokebe61@yahoo.com (I. E. Ebukanson)

## To cite this article:

Imo Enoidem Ebukanson, Chukwu Benedict Chidi, Abode Innocent Iriaoghuan. Statistical Analysis of Electricity Generation in Nigeria Using Multiple Linear Regression Model and Box-Jenkins' Autoregressive Model of Order 1. *International Journal of Energy and Power Engineering*. Vol. 6, No. 3, 2017, pp. 28-33. doi: 10.11648/j.ijep.20170603.12

**Received:** January 8, 2017; **Accepted:** January 18, 2017; **Published:** June 7, 2017

---

**Abstract:** This study presents statistical analysis of electricity generation in Nigeria using two different statistical models, namely; multiple linear regression model and box-Jenkins' autoregressive model of order 1. Two climatic variables (rainfall and temperature) were used as the explanatory variables. Data on electricity generation in Nigeria between 2002 and 2014 were obtained from the Central Bank of Nigeria Statistical Bulletin while Data on rainfall and temperature between 2002 and 2014 were extracted from the National Bureau of Statistics (NBS) abstract. Test of model fitness and forecasting accuracy were done using generic statistical approach which include coefficient of determination and root mean square error. The prediction accuracy of the two statistical models was compared and the best model was selected. Furthermore, correlation between power generation and the two climatic variables (rainfall and temperature), were carried out and the result reveals that the amount of rainfall has significant and positive relationship with power generation in Nigeria. Specifically, rainfall has correlation value of  $r = 0.927$  with the power generation at probability,  $p = 0.000$  and the relationship was significant at 1% ( $p < 0.01$ ). However, temperature although it is positively correlated, does not significantly affect power generation. Temperature has correlation value of  $t = 0.136$  with power generation at probability,  $p = 0.658$  ( $p > 0.05$ ) and the relationship was significant at 5% ( $p < 0.05$ ). Among the two statistical models, multiple linear regression model was selected as the best model as it gave the highest value of coefficient of determination ( $r^2 = 99.77\%$ ) and the least Root Mean Square Error (60.27%).

**Keywords:** Electricity, Box-Jenkins' Autoregressive Model, Electricity Generation, Multiple Linear Regression Model, Statistical Analysis of Electricity

---

## 1. Introduction

Electricity generation and supply is indispensable in modern living in any industrial or commercial society. Electricity production in Nigeria over the years varied from gas – fired, oil – fired, hydroelectric power stations to coal-fired with hydroelectric power system and gas – fired system taking precedence [1-4]. Presently, Nigeria mostly employs gas-fired and hydroelectric turbines for bulk generation, oil being too expensive and coal-fired stations having gone moribund.

Nigeria being rapidly growing in development, industry

and commerce, is battling with the problems of continually expanding consumption and increase in demand for electricity. Madueme [5] observed that in Nigeria, the national peak demand for electric energy is on the increase as a result of many development activities. He went further to state that in spite of this, the total power generation has not matched this increase. It is a fact that most communities in Nigeria regard the construction of access roads, installation of pipe borne water, building of town halls and so forth as the rudiments to their development, yet electricity promotes instantly their social and economic life as many services and facilities become readily available [6].

In this regard, adequate electricity generation constitute a central development issue which cannot be over-emphasized [7-11]. Apart from serving as the pillar of wealth creation in Nigeria, it is also the nucleus of operations and subsequently the engine of growth for all sectors of the economy. This study examines how the power generation will be affected by climate change or variables such as temperature and rainfall which is regarded as one of the greatest threat to humanity in the 21<sup>st</sup> century. Considerable efforts have been made in assessing the causal link between energy consumption and economic growth, but very few studies have examined how environmental challenges such as the incidence of climate change will significantly impact energy production which is regarded as the bane of economic growth. Most researches in the aspect of impacts of climate change on the energy sector are in form of technical reports; some others examine how the energy sector contributes to climate change and not so much on how the energy sector will be affected by the incidence of climate change. This study will attempt to fill this gap in addition to examining the bivariate relationship between two climate variables (rainfall and temperature) and electric power generation in Nigeria between 2002 and 2014, conducting statistical analysis of electric power generation in Nigeria between 2002 and 2014 based on two different statistical models using rainfall and temperature as the explanatory variables.

## 2. Methodology

In this study, analytical and simulation research methodologies are used. In the analytical method, statistical models are developed for analysing and predicting electricity generation in Nigeria. Specifically, two different statistical models are used in this study. These models are multiple linear regression model [12-14] and Box-Jenkins' autoregressive model of order 1 [15-19]. It also provides the source and method of data collection.

### 2.1. Source and Method of Data Collection

The source of data collection for this study is secondary. Data on electricity generation in Nigeria between 2002 and 2014 were obtained from the Central Bank of Nigeria statistical bulletin [20] while data on rainfall and temperature were extracted from the National Bureau of Statistics (NBS) abstract [21].

### 2.2. Multiple Linear Regression

The multiple linear regression can be represented mathematically as:

$$G_i = \beta_0 + \beta_1 R_i + \beta_2 T_i + e_i \quad (1)$$

Making the residual the subject in equation (1) gives;

$$e_i = G_i - \beta_0 - \beta_1 R_i - \beta_2 T_i \quad (2)$$

The sum of square error ( $S_2$ ) is given as;

$$S_2 = \sum_{i=1}^n (G_i - \beta_0 - \beta_1 R_i - \beta_2 T_i)^2 \quad (3)$$

$$\frac{\partial S_2}{\partial \beta_0} = 2(-1) \sum (G_i - \beta_0 - \beta_1 R_i - \beta_2 T_i) \quad (4)$$

$$\frac{\partial S_2}{\partial \beta_1} = 2(-1) \sum_{i=1}^n [R_i (G_i - \beta_0 - \beta_1 R_i - \beta_2 T_i)] \quad (5)$$

$$= 2(-1) \sum_{i=1}^n (R_i G_i - \beta_0 R_i - \beta_1 R_i^2 - \beta_2 R_i T_i)$$

$$\frac{\partial S_2}{\partial \beta_2} = -2 \sum_{i=1}^n T_i (G_i - \beta_0 - \beta_1 R_i - \beta_2 T_i)$$

$$= -2 \sum_{i=1}^n (T_i G_i - \beta_0 T_i - \beta_1 T_i R_i - \beta_2 T_i^2) \quad (6)$$

To find estimate of the parameters ( $\beta_0, \beta_1, \beta_2$ ), equations (3), (5), (6) are set to zero. Then,

$$\sum_{i=1}^n G_i - n\beta_0 - \beta_1 \sum_{i=1}^n R_i - \beta_2 \sum_{i=1}^n T_i = 0 \quad (7)$$

$$\sum_{i=1}^n R_i G_i - \beta_0 \sum_{i=1}^n R_i - \beta_1 \sum_{i=1}^n R_i^2 - \beta_2 \sum_{i=1}^n R_i T_i = 0 \quad (8)$$

$$\sum_{i=1}^n T_i G_i - \beta_0 \sum_{i=1}^n T_i - \beta_1 \sum_{i=1}^n T_i R_i - \beta_2 \sum_{i=1}^n T_i^2 = 0 \quad (9)$$

Rearranging equations (7), (8) and (9) gives;

$$n\beta_0 + \beta_1 \sum_{i=1}^n R_i + \beta_2 \sum_{i=1}^n T_i = \sum_{i=1}^n G_i \quad (10)$$

$$\beta_0 \sum_{i=1}^n R_i + \beta_1 \sum_{i=1}^n R_i^2 + \beta_2 \sum_{i=1}^n R_i T_i = \sum_{i=1}^n R_i G_i \quad (11)$$

$$\beta_0 \sum_{i=1}^n T_i + \beta_1 \sum_{i=1}^n T_i R_i + \beta_2 \sum_{i=1}^n T_i^2 = \sum_{i=1}^n T_i G_i \quad (12)$$

Solving the systems of the equations, 10 to12, give the estimate of the parameters of the model.

### 2.3. Box Jenkins' Autoregressive Model of Order 1

The Box Jenkins' Autoregressive Model of Order 1 can be expressed mathematically in the form:

$$g_i = \gamma_0 + \gamma_1 g_{i-1} + e_i \quad (13)$$

The residual term  $e_i$  can be obtained from Equation (12) as:

$$e_i = g_i - (\gamma_0 + \gamma_1 g_{i-1}) \quad (14)$$

$$e_i = (g_i - \gamma_0 - \gamma_1 g_{i-1}) \quad (15)$$

The sum of square error can be expressed as:

$$RMSE = \sqrt{MSE} \tag{26}$$

$$S_4 = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (g_i - \gamma_0 - \gamma_1 g_{i-1})^2 \tag{16}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (G_i - \hat{G}_i)^2 \tag{27}$$

Differentiating the Equation (16) with respect to  $\gamma_0, \gamma_1$

$$\frac{\partial S_4}{\partial \gamma_0} = -2 \sum_{i=1}^n (g_i - \gamma_0 - \gamma_1 g_{i-1}) \tag{17}$$

$$\frac{\partial S_4}{\partial \gamma_1} = -2 \sum_{i=1}^n g_{i-1} (g_i - \gamma_0 - \gamma_1 g_{i-1}) \tag{18}$$

$$\frac{\partial S_4}{\partial \gamma_1} = -2 \left[ \sum_{i=2}^n g_i g_{i-1} - \gamma_0 \sum_{i=2}^n g_{i-1} - \gamma_1 \sum_{i=2}^n g_{i-1}^2 \right] \tag{19}$$

To obtain the parameters  $(\gamma_0, \gamma_1)$  we set the partial derivation with respect to each parameter to 0. Then:

$$n\gamma_0 + \gamma_1 \sum_{i=2}^n g_{i-1} = \sum_{i=1}^n g_i \tag{20}$$

$$\gamma_0 \sum_{i=2}^n g_{i-1} + \gamma_1 \sum_{i=2}^n g_{i-1}^2 = \sum_{i=2}^n g_i g_{i-1} \tag{21}$$

Solving the set of equations 20-21, with two unknowns  $(\gamma_0, \gamma_1)$ , this will give the estimates of the model parameters.

**2.4. Test of Model Fitness**

**2.4.1. Coefficient of Determination**

SST is the sum of square total which is given as;

$$SST = \sum_{i=1}^n (G_i - \bar{G})^2 \tag{22}$$

Where  $G_i$  is the electricity generated during the year

$$SS (\text{Error}) = \sum_{i=1}^n (G_i - \hat{G})^2 \tag{23}$$

$$SS (\text{Regression}) = \sum_{i=1}^n (\hat{G}_i - \bar{G})^2 \tag{24}$$

Coefficient of Determination ( $r^2$ ) is given as;

$$r^2 = \frac{SSR}{SST} \tag{25}$$

Where  $\bar{G}$  and  $\hat{G}$  are the mean and estimated electricity generation.

**2.4.2. Root Mean Square Error (RMSE)**

The root mean square error was used to assess the forecasting accuracy of the four models. The RMSE is defined as:

Where  $G_i$  and  $\hat{G}$  are the electricity generated during the year and estimated electricity generation and n is the number of observation.

**3. Results and Discussion**

The data on electricity generation obtained from central bank of Nigeria statistical bulletin and the two climatic variables (rainfall and temperature) extracted from National Bureau of statistics abstract between 2002 and 2014 were analysed based on the two different statistical models. Correlation between power generation and climatic variables was carried out. Graphpad prism 5.0 Econometric View (E-View) software was used to plot the graph of actual and predicted power generation in Nigeria between 2002 and 2014 for the two statistical models. Generic statistical approach was used to assess the goodness of fit and forecasting accuracy of the two models. The prediction accuracy of the two models was compared and the best model was selected.

Table 1 shows the result of bivariate relationship between power generation and the climatic variables. The result reveals that the amount of rainfall was found to be significantly related with power generation ( $r = 0.927, p = 0.000, p < 0.01$ ). The relationship was significantly positive which means that as the amount of rainfall increases, power generation also increases. Also, the relationship between the temperature and power generation was also positive but not significant ( $t = 0.136, p = 0.658, p > 0.05$ ). This relationship was not significant at 5 percent ( $p < 0.05$ ).

**Table 1.** Correlation between power generation and climatic variables.

	G	R	T
G	1		
R	0.927** (0.000)	1	
T	0.136 (0.658)	-0.117 (0.703)	1

G = Power generation, R = rainfall, T = temperature. \*\*Significant at 1% ( $p < 0.01$ ), \*significant at 5% ( $p < 0.05$ ).

**3.1. Multiple Linear Regression**

Table 2 shows that the parameters of the multiple linear regression are as follows:

$$\beta_0 = -2609.781, \beta_1 = 0.4600, \beta_2 = 136.1937.$$

Therefore, the estimated model for power generation based on the multiple linear regression is:

$$G = -2609.781 + 0.4600 R + 136.1937T \tag{28}$$

The coefficient of determination is 0.9977. This indicates that rainfall and temperature explained for 99.77% of the variation in power generation while the remaining 0.23

percent of the variation in power generation could be due to other variables not accounted for in the multiple linear regression model. Furthermore, the coefficient of rainfall ( $\beta_1 = 0.46$ ) is positive meaning that as rainfall increases, power generation also increases.

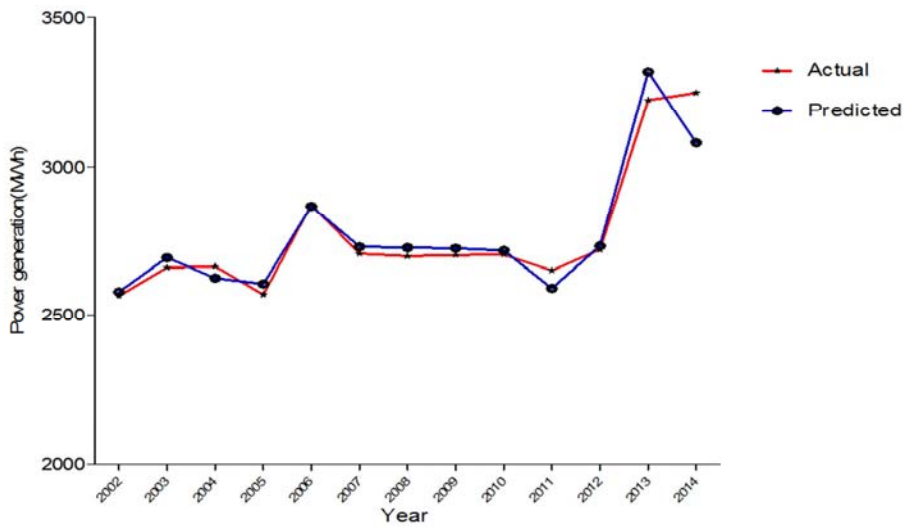
**Table 2.** Summary result of the estimates of the multiple linear regression model.

Model Parameters	$\beta$	SSR	SST	r-square	RMSE (%)
Constant	-2609.781	47229.207	20833613	0.9977	60.2745
Rainfall	0.4600				
Temperature	136.1937				

$\beta$  = coefficients, SSR = regression sum of square, SST = Total Sum of Square, RMSE = Root Mean Square Error.

**Table 3.** Actual and Predicted power generation in Nigeria between 2002 and 2014 using multiple linear regression model.

Year	Actual generation (MWh)	Predicted generation (MWh)
2002	2564.3	2577.93
2003	2660.0	2693.14
2004	2663.6	2624.14
2005	2569.3	2605.05
2006	2871.3	2866.97
2007	2706.3	2729.27
2008	2698.1	2726.33
2009	2701.4	2724.31
2010	2703.5	2717.15
2011	2650.0	2590.67
2012	2719.9	2732.22
2013	3225.0	3319.79
2014	3249.0	3080.48



**Figure 1.** Graph of actual and predicted power generation in Nigeria between 2002 and 2014 using multiple linear regression model.

**3.2. Box Jenkins's Autoregressive Model of Order 1**

The Box's Jenkin's AR (1) model for power generation is:

$$G_t = 684.1545 + 0.770103G_{t-1} \tag{29}$$

From Table 4, Box Jenkin's AR (1) model gave an  $r^2$  value of 98.45% which means that the model explained 98.45 percent of the variation in power generation in Nigeria. The coefficient of power generation at one period lag ( $\beta = 0.770103$ ) is positive which means that power generation will increase with time.

The Box's Jenkin's AR (1) model for power generation is:

$$G_t = 684.1545 + 0.770103G_{t-1} \tag{30}$$

From Table 4, Box Jenkin's AR (1) model gave an  $r^2$  value of 98.45% which means that the model explained 98.45 percent of the variation in power generation in Nigeria. The coefficient of power generation at one period lag ( $\beta = 0.770103$ ) is positive which means that power generation will increase with time.

**Table 4.** Summary result of the estimates of the box Jenkins's autoregressive model of order 1.

Model Parameters	$\beta$	SSR	SST	r-square	RMSE (%)
Constant	684.1545	21827609.17	22170749.38	0.9845	169.1006
$G_{t-1}$	0.770103				

$\beta$  = coefficients, SSR = regression sum of square, SST = Total Sum of Square, RMSE = Root Mean Square Error.

**Table 5.** Actual and Predicted power generation in Nigeria between 2003 and 2014 using box's Jenkins' autoregressive model of order 1.

Year	Actual generation (MWh)	Predicted generation (MWh)
2003	2660.0	2658.93
2004	2663.6	2732.63
2005	2569.3	2735.40
2006	2871.3	2662.78
2007	2706.3	2895.35
2008	2698.1	2768.28
2009	2701.4	2761.97
2010	2703.5	2764.51
2011	2650.0	2766.13
2012	2719.9	2724.93
2013	3225.0	2778.76
2014	3249.0	3167.74

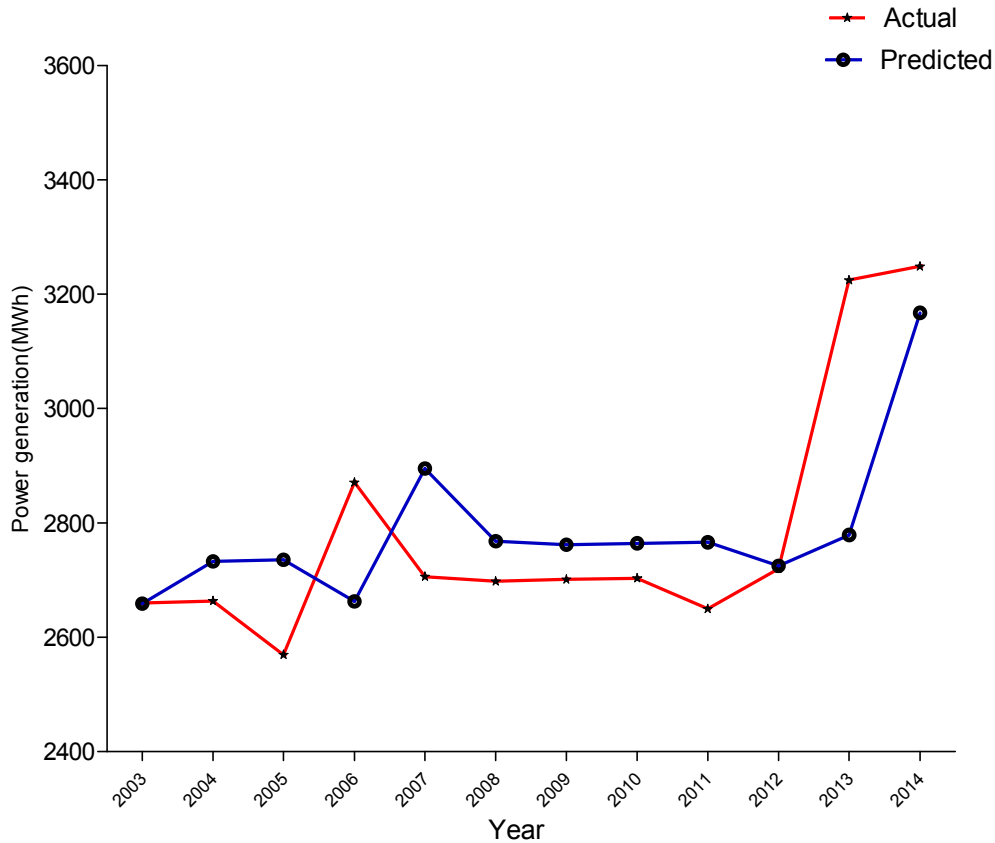


Figure 2. Graph of actual and predicted power generation in Nigeria between 2003 and 2014 using box's Jenkins' autoregressive model of order 1.

### 3.3. Comparison of the Forecasting Accuracy of the Two Models

Table 6. Comparison of the forecasting accuracy of the two models.

S/N	Models	r-square (%)	RMSE	Rank
1	Multiple linear regression	99.77	60.27	1
2	Box-Jenkins's Autoregressive Model of order 1	98.45	169.10	2

Based on the result in Table 6, the multiple linear regression model has the highest coefficient of determination (r-square = 99.77%) and least root mean square error (RMSE = 60.27). Hence, the multiple linear regression model is recommended as the best of the two competing models.

## 4. Conclusion

In this study, correlation between power generation and two climatic variables are presented. The result reveals that the amount of rainfall has significant and positive relationship with power generation in Nigeria. The relationship was significantly positive which means that as the amount of rainfall increases, power generation also increases. However, temperature although positive, does not significantly affect power generation. From the results obtained from model 1 shows that the two independent variables (rainfall and temperature) explained for 99.77% of the variation in electricity generation. The multiple linear regression model

was selected as the best model as it gave the highest value of coefficient of determination ( $r^2=99.77\%$ ) and the least root mean square error (60.27%).

## References

- [1] Wara, S. T. (2012). Electricity Provision And Managemel In Nigeria: Challenges And Prospects.
- [2] Oji, J. O., Idusuyi, N., & Kareem, B. (2012). Coal power utilization as an energy mix option for Nigeria: a review. *American Academic & Scholarly Research Journal*, 4 (4), 1.
- [3] Anumaka, M. C. (2012). Scenario of Electricity in Nigeria. *International Journal of Engineering and Innovative Technology (IJEIT)*, 1 (6), 176-183.
- [4] Anumaka, M. C. (2012). Scenario of Electricity in Nigeria. *International Journal of Engineering and Innovative Technology (IJEIT)*, 1 (6), 176-183.
- [5] Madueme, T. I. (2002). Analysis of Electricity Load Demand in Nigeria. *Nigerian Journal of Engineering Management*. 3 (2): 76.
- [6] Ogumodede O. B. (2005). Consumers' Expectations on Service Delivery of PHCN: A Study of Lagos and Ibadan Metropolis. Unpublished MBA Thesis, Imo State University.
- [7] Ayodele, A. S. (2004). Improving and sustaining power (electricity) supply for socio-economic development in Nigeria.

- [8] Lionel, E. (2013). The dynamic analysis of electricity supply and economic development: Lessons from Nigeria. *Journal of Sustainable Society*, 2 (1), 1-11.
- [9] Wara, S. T., Abayomi-Alli, A., Umo, N. D., Oghogho, I., & Odikayor, C. (2009). An impact assessment of the Nigerian power sector reforms. In *Advanced Materials Research* (Vol. 62, pp. 147-152). Trans Tech Publications.
- [10] Egboh, H. I. (2011). Clean energy in Norway: a case study for Nigerian electricity development.
- [11] Iwayemi, A. (2008). Investment in electricity generation and transmission in Nigeria: issues and options. *International Association for Energy Economics*, 37-42.
- [12] Hinman, J., & Hickey, E. (2009). Modeling and forecasting short-term electricity load using regression analysis. *Journal of Institute for Regulatory Policy Studies [электронный ресурс]*.
- [13] Cho, H., Goude, Y., Brossat, X., & Yao, Q. (2013). Modeling and forecasting daily electricity load curves: a hybrid approach. *Journal of the American Statistical Association*, 108 (501), 7-21.
- [14] Safa, M., Allen, J., & Safa, M. (2014, January). Predicting Energy Usage Using Historical Data and Linear Models. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction* (Vol. 31, p. 1). Vilnius Gediminas Technical University, Department of Construction Economics & Property.
- [15] Adhikari, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. *arXiv preprint arXiv: 1302.6613*.
- [16] Roken, R. M., & Badri, M. A. (2006). Time Series Models for Forecasting Monthly Electricity Peak Load for Dubai. *Chancellor's Undergraduate Research Award*.
- [17] Bennett, C., Stewart, R. A., & Lu, J. (2014). Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks. *Energies*, 7 (5), 2938-2960.
- [18] Singh, A., & Mishra, G. C. (2015). Application of Box-Jenkins method and Artificial Neural Network procedure for time series forecasting of prices. *Statistics in Transition new series*, 1 (16), 83-96.
- [19] Mohamed, Z., & Bodger, P. S. (2004). Forecasting electricity consumption: A Comparison of models for New Zealand.
- [20] CBN. (2006). Central Bank of Nigeria, *Statistical Bulletin*, Vol. 17.
- [21] NBS. (2012). National Bureau of Statistics Abstract Available at: [http://www.nigerianstat.gov.ng/pdfuploads/annual\\_abstract\\_2012.pdf](http://www.nigerianstat.gov.ng/pdfuploads/annual_abstract_2012.pdf). Accessed on 12<sup>th</sup> November 2016.