



Energy Consumption & Economic Growth in Nigeria (1971 - 2021): A Multifaceted Analysis Using Descriptive Statistics, Charts, Logistic Regression, and Neural Networks

Victory Ikpe ^{a*} and Emmanuel Oyedeji ^a

^a Western Illinois University, United States of America.

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2023/v25i4572

Open Peer Review History:

This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: <https://www.sdiarticle5.com/review-history/110212>

Received: 09/10/2023

Accepted: 13/12/2023

Published: 19/12/2023

Original Research Article

Abstract

This study examines the intricate interplay between electricity consumption and economic growth, specifically Gross Domestic Product (GDP), in Nigeria. Despite the acknowledged influence of electricity consumption on GDP, the research takes a comprehensive approach by scrutinizing various economic factors that may contribute to Nigeria's remarkable GDP performance within the African context. Utilizing a meticulous methodology, the investigation explores the relationships between GDP and variables such as unemployment, population, lending interest rates, importation costs, and inflation rates across a 50-case observational dataset. The findings highlight the substantial impact of electric power consumption, population, and lending interest rates on Nigeria's GDP, while revealing the relatively minor effects of other predictors.

*Corresponding author: Email: vi-ikpe@wiu.edu;

Keywords: Nigeria's GDP; electric power consumption; electricity consumption; economic growth.

1 Introduction

This research aims to elucidate the complex relationship between energy consumption and economic growth, specifically focusing on Nigeria [1,2]. As the largest economy in Africa since 2013, Nigeria's GDP consistently surpasses that of its continental counterparts [3-5]. Despite the acknowledged influence of electricity consumption on GDP, the paradoxical scenario of inadequate electricity supply challenges conventional assumptions [6-8]. This study seeks to unravel the economic factors contributing to Nigeria's GDP growth, examining the roles and significance of these factors in the context of suboptimal electricity consumption.

Purpose of the Study: Given the paradox of low electricity consumption and Nigeria's high GDP ranking in Africa, this research aims to discern the multifaceted economic factors contributing to this phenomenon. The study endeavors to delineate the roles and significance of these factors in shaping Nigeria's augmented GDP, scrutinizing whether electricity consumption remains the predominant determinant.

Research Questions:

Aligned with the research objectives, the study addresses pivotal questions:

1. What is the impact of electricity consumption on Nigeria's GDP during the periods under examination?
2. To what extent does electricity consumption significantly influence Nigeria's GDP?
3. In the absence of a substantial impact from electricity consumption, what alternative factors exert influence on Nigeria's GDP, and what are their respective roles and significance?

Statement of Hypothesis: Considering the research queries, the study formulates hypotheses to assess the significance of electricity consumption on GDP: A. H₀; Null hypothesis: B₁ = 0 (Electricity consumption has no significant effect on Nigeria's GDP) B. H₁; Alternative hypothesis: B₁ ≠ 0 (Electricity consumption has a significant effect on Nigeria's GDP).

Data Source & Scope: The statistical models employed in this study were implemented using Solver Excel XLMiner tool, providing a robust analytical framework for investigating the relationships between energy consumption and various economic variables impacting Nigeria's GDP. The primary dataset for this analysis is derived from the World Bank Development Indicators (WDI) database, encompassing the years 1971 to 2021. The choice of this temporal span allows for a comprehensive exploration of energy consumption trends and their correlations with key variables influencing Nigeria's economic dynamics. The reliability and relevance of the data are ensured through its careful curation from <https://databank.worldbank.org/source/world-development-indicators>.

2 Model Specification

The modeling process involves a meticulous examination of the selected variables, including electric power consumption, population, lending interest rates, unemployment, importation costs, and inflation rates. Leveraging the Solver Excel XLMiner, the study aims to optimize the coefficients and parameters within the selected models to best fit the observed data. The XLMiner tool iterative optimization capabilities enable the identification of the most fitting relationships and their respective magnitudes, contributing to a nuanced understanding of the intricate economic dynamics at play. The regression equation that captures the effect of energy consumption, vis-a-vis other macroeconomic variables, on GDP is captured by the equation as stated below:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \epsilon$$

$$\text{GDP} = \beta_0 + \beta_1\text{ElectricityConsumption} + \beta_2\text{InflationRate} + \beta_3\text{TotalPopulation} + \beta_4\text{ImportationCost} + \beta_5\text{BankLendingInterestRate} + \beta_6\text{UnemploymentRate} + \epsilon$$

Where; Y is the dependent variable that captures GDP

X₁ is one of the independent variables that captures electricity consumption

X₂ captures inflation rate

X₃ captures total population

X₄ captures importation cost

X₅ captures bank lending interest rate

X₆ captures Unemployment rate

€ is the error term

Table 1. Descriptive Statistics

	GDP (current US\$)	Electric P.C (kWh per capita)	Inflation	Population	Imports	Lending Interest Rate (%)	Unemployment
Mean	1.71892E+11	95.78868672	18.30377629	1.2E+08	2.425E+10	15.56737308	4.534017785
Standard Error	23832729935	4.673295749	2.208795361	6323714	3589937982	0.841491651	0.22380281
Median	73615090340	90.47039342	12.77548619	1.12E+08	1.0829E+10	16.82041667	3.953499979
Mode	#N/A	#N/A	#N/A	#N/A	#N/A	7	#N/A
Standard Deviation	1.68523E+11	33.04519114	15.61854178	44715413	2.5385E+10	5.950244529	1.582524847
Sample Variance	2.84E+22	1091.984658	243.9388473	2E+15	6.4438E+20	35.40540995	2.504384891
Kurtosis	-0.774588375	-0.885095983	3.396652081	-0.93366	-0.11722403	-0.296946667	5.053147881
Skewness	0.916593677	-0.071079085	1.99645577	0.460345	1.07639954	0.113007429	2.529328089
Range	5.34402E+11	124.0694028	69.37785255	1.53E+08	8.7293E+10	25.65	6.088000059
Minimum	12274416018	32.72774895	3.457649752	58665813	1447830923	6	3.700000048
Maximum	5.46676E+11	156.7971517	72.8355023	2.11E+08	8.8741E+10	31.65	9.788000107
Sum	8.59462E+12	4789.434336	915.1888145	6E+09	1.2125E+12	778.3686542	226.7008892
Count	50	50	50	50	50	50	50

Table 2. Correlation

	GDP (current US\$)	Electric power consumption (kWh per capita)	Inflation, consumer prices (annual %)	Population, total	Imports of goods and services (current US\$)	Lending interest rate (%)	Unemployment
GDP (current US\$)	1						
Electric power consumption (kWh per capita)	0.827166	1					
Inflation, consumer prices (annual %)	-0.31591	-0.087994963	1				
Population, total	0.886983	0.885343236	-0.196945341	1			
Imports of goods and services (current US\$)	0.938179	0.811001679	-0.31355436	0.85721268	1		
Lending interest rate (%)	0.080489	0.460680449	0.317217485	0.390184579	0.118099376	1	
Unemployment, total (% of total labor force) (modeled ILO estimate)	0.522103	0.377872203	-0.073582303	0.644571049	0.464806877	-0.002106582	1

Analysis: Each cell in the above table shows the correlation between two specific variables. The increase in GDP is strongly related to the Electric power consumption, Imports and Population at 0.83, 0.94 and 0.88. The correlation between "Inflation" and the "GDP (current)" is -0.316, which indicates that they are weakly negatively correlated. That is, the increase in GDP is associated with inflation

Table 3. Pivot table

GDP (current US\$)	Average of Electric power consumption (kWh per capita)
5.46676E+11	144.5254385
5.08693E+11	142.729212
4.86803E+11	122.5882514
4.55502E+11	156.7971517
4.4812E+11	132.6985635
4.40777E+11	133.5680311
4.32294E+11	133.3029686
4.04994E+11	150.1980159
4.0465E+11	125.7947602
3.9719E+11	130.7752289
3.75746E+11	129.1430924
3.61457E+11	136.4262652
3.39476E+11	127.2446137
2.95009E+11	120.6350712
2.75626E+11	138.9094166
2.36104E+11	111.7524163
1.76134E+11	129.3270312
1.64475E+11	50.90103784
1.42769E+11	81.89592594
1.36386E+11	123.6324865
1.04912E+11	101.9252114
97094911792	81.73535221
95385819321	104.6610449
74030364472	75.56977626
73745821158	80.4544837
73484359521	62.06356225
69448756933	74.49061979
64201788123	68.0570943
59372613486	75.76719632
54805852581	90.88641197
54604050168	76.96886384
54457835193	82.00415836
54035795388	87.07894453
52676041931	89.30353228
51075815093	85.90498293
49648470440	87.13950744
49118433048	89.59979699
47794925815	90.05437487
47259911894	59.82247397
44062465800	91.48820715
44003061108	97.07070111
36527862209	60.68837566
36308883249	51.5781796
36035407725	59.1787877
33833042988	95.56127199
27778934625	45.77566148
27752204320	100.8853382
24846641318	32.84846076
15162871287	35.2992363
12274416018	32.72774895

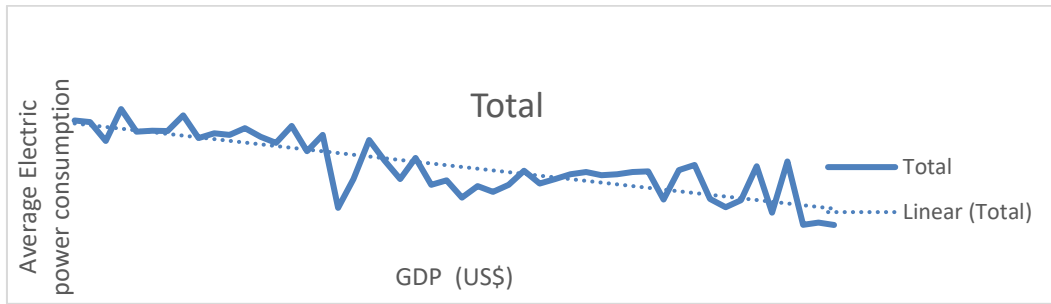


Fig. 1. GDP vs Avg Electric Power Consumption Trendline (1972-2021)

From the Pivot table, the GDP (Y Variable) and the Average Electric Consumption is the explanatory variable. Broadly speaking, Average Electric power consumption (KWh per capita) decreases as the GDP decreases and also Increases as GDP increase. In Year 2012, the GDP had the highest Electric power consumption at 157. We see a positive correlation between these two variables in the Economic growth of Nigeria.

HISTOGRAM

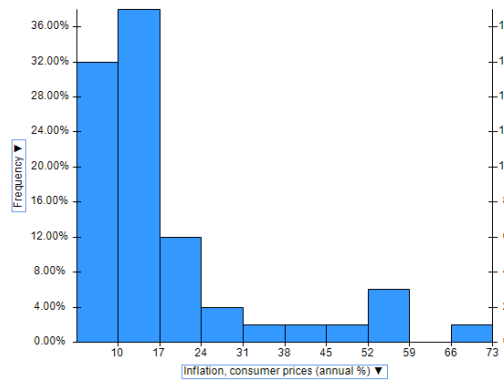


Fig. 2. Histogram

Analysis of Histogram: The GDP shows a slightly skewed distribution with a minor long tail to the right with an outlier at 71. There is an increase in inflation at the 10 to 17 range.

SCATTERPLOT MATRIX

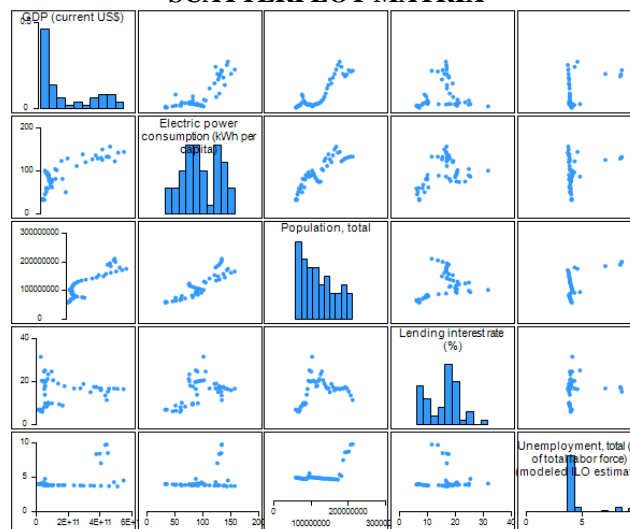


Fig. 3. Scatterplot Matrix

From the Scatterplot matrix above, we can make the following statement:

1. There is a positive relationship between the GDP and Electric power consumption. That is, an increase in GDP also leads to a corresponding increase in Electric power consumption
2. Also we can see a positive relationship between the GDP and Population. That is, an increase in population yields a corresponding increase in GDP
3. There is a horizontal line in the relationship between GDP and Unemployment. That is, there is no relationship between GDP and Unemployment.

Because the line of best fit slants from left to right, the above scatter plot shows a positive relationship between GDP and Populations. That means in General, the higher the GDP, the higher the population. (Note: Birth rate increases population)

3. Exhaustive Search Analysis

Table 4. Regression analysis

Regression Summary							
Metric	Value						
Residual DF	43						
R2	0.93751171						
Adjusted R2	0.92879241						
Std. Error Estimate	4.497E+10						
RSS	8.6959E+22						

Feature Selection							
Best Subsets							
Subset ID	Intercept	Electric power consumption (kWh per capita)	Inflation, consumer prices (annual %)	Population, total	Imports of goods and services (current US\$)	Lending interest rate (%)	Unemployment, total (% of total labor force) (modeled ILO estimate)
Subset 1	1	0	0	0	0	0	0
Subset 2	1	0	0	0	1	0	0
Subset 3	1	0	0	1	1	0	0
Subset 4	1	0	0	1	1	1	0
Subset 5	1	1	0	1	1	1	0
Subset 6	1	1	0	1	1	1	1
Subset 7	1	1	1	1	1	1	1

Best Subsets Details						
Subset ID	#Coefficients	RSS	Mallovs's Cp	R2	Adjusted R2	Probability
Subset 1	1	1.3916E+24	640.1288733	-8.88178E-16	-8.88178E-16	3.00897E-24
Subset 2	2	1.66741E+23	36.45168576	0.880179878	0.877683625	2.42093E-05
Subset 3	3	1.30795E+23	20.67678759	0.906010647	0.9020111	0.001266495
Subset 4	4	1.00696E+23	7.793017901	0.927639982	0.92292085	0.094568962
Subset 5	5	9.04448E+22	4.723913267	0.935006486	0.929229285	0.429502967
Subset 6	6	8.71289E+22	5.084240256	0.937389286	0.930274432	0.773026474
Subset 7	7	8.69586E+22	7	0.937511705	0.928792408	N/A

The variable “Imports of goods and services” entered the model first. Imports accounts for 94.0% of the variability in the GDP. We then move on to select the best two models. We make this selection from the best two models with the highest Adjusted R-Square.

The best two models in this Featured selection are Subset 6 and Subset 5 with 5 and 4 regressors respectively. They have the highest Adjusted R-square at 0.9303 and 0.9292 respectively.

We then selected the best model for our analysis.

See analysis below:

Table 5. One of the best model with five (5) regressors

Feature Selection						
Best Subsets						
Subset ID	Intercept	Electric power consumption (kWh per capita)	Population, total	Imports of goods and services (current US\$)	Lending interest rate (%)	Unemployment, total (% of total labor force) (modeled ILO estimate)
Subset 1	1	0	0	0	0	0
Subset 2	1	0	0	1	0	0
Subset 3	1	0	1	1	0	0
Subset 4	1	0	1	1	1	0
Subset 5	1	1	1	1	1	0
Subset 6	1	1	1	1	1	1

Best Subsets Details						
Subset ID	#Coefficients	RSS	Mallows's Cp	R2	Adjusted R2	Probability
Subset 1	1	1.3916E+24	654.7551198	-2.22045E-16	-2.22045E-16	2.58212E-25
Subset 2	2	1.66741E+23	38.20420441	0.880179878	0.877683625	7.23552E-06
Subset 3	3	1.30795E+23	22.05149904	0.906010647	0.9020111	0.00042564
Subset 4	4	1.00696E+23	8.851373371	0.927639982	0.92292085	0.041428554
Subset 5	5	9.04448E+22	5.674524421	0.935006486	0.929229285	0.202404856
Subset 6	6	8.71289E+22	6	0.937389286	0.930274432	N/A

Coefficients						
Predictor	Estimate	Confidence Interval: Lower	Confidence Interval: Upper	Standard Error	T-Statistic	P-Value
Intercept	-6.8108E+10	-1.29741E+11	-6474646782	30581505448	-2.2270853	0.0311059
Electric power consumption (kWh per capita)	814346178	-206729392.7	1835421750	506644834.5	1.607331454	0.1151351
Population, total	2095.53435	964.8297997	3226.238901	561.0413529	3.735080025	0.0005366
Imports of goods and services (current US\$)	2.66476204	1.389408582	3.940115505	0.63281432	4.210969883	0.0001237
Lending interest rate (%)	-7295940536	-10467894333	-4123986739	1573883512	-4.63562931	3.179E-05
Unemployment, total (% of total labor force) (modeled ILO estimate)	-8918179447	-22807619278	4971260385	6891765059	-1.29403417	0.2024049

ANOVA					
Source	DF	SS	MS	F-Statistic	P-Value
Regression	5	1.30447E+24	2.60894E+23	131.751024	2.58212E-25
Error	44	8.71289E+22	1.9802E+21	N/A	N/A
Total	49	1.3916E+24	2.84E+22	N/A	N/A

Table 6. Training: Prediction Summary

Metric	Value
SSE	8.71289E+22
MSE	1.74258E+21
RMSE	41744201762
MAD	32828810350
R2	0.937389286

i. From the above Model with 5 Regressors), the estimated regression equation is: $Y^{\wedge} = -0.0000 + 814346178 \text{ ELECTRIC CONS.} + 2095.53 \text{ POPULATION} + 2.664 \text{ IMPORT} - 72959 \text{ LENDING RATE} - 89181 \text{ UNEMPLOYMENT}$

At 20 % Significant level:

- ii. Reject $H_0: \beta_2 = 0$ since p-value of 0.0005 is less than 20%
- iii. Reject $H_0: \beta_3 = 0$ since p-value of 0.00001 is less than 20%
- iv. Reject $H_0: \beta_4 = 0$ since p-value of 0.00000 is less than 20%

That is, the regression coefficient (β_2, β_3 & β_4 all have p-value less than 20%, thus, the coefficients are significant at 20%. That is, they all reject the $H_0: \beta_1 = 0$ at 20% significance level, and that the regression coefficients is statistically different from 0.

At 20% Significant level, the following explanatory variables are **not** statistically significant:

- Electric Consumption &
- Unemployment

The above variables have p-value > 20%

Removing an insignificant variable at the 20% level and re-run the regression gives the following printout:

Table 7. Data statistic result

Coefficients

Predictor	Estimate	Confidence Interval: Lower	Confidence Interval: Upper	Standard Error	T-Statistic	P-Value
Intercept	-8.6696E+10	-1.41479E+11	-31914004603	27199336110	-3.18744102	0.0026109
Electric power consumption (kWh per capita)	1065103873	115217324.5	2014990421	471617571.1	2.258405831	0.0288157
Population, total	1550.54924	798.4012247	2302.697264	373.4406207	4.152063698	0.0001449
Imports of goods and services (current US\$)	2.94177534	1.733407513	4.150143162	0.599953225	4.903341148	1.268E-05
Lending interest rate (%)	-6474055939	-9396101702	-3552010176	1450792337	-4.46242772	5.375E-05

ANOVA

Source	DF	SS	MS	F-Statistic	P-Value
Regression	4	1.30115E+24	3.25288E+23	161.8441963	4.2928E-26
Error	45	9.04448E+22	2.00988E+21	N/A	N/A
Total	49	1.3916E+24	2.84E+22	N/A	N/A

$$Y^{\wedge} = -0.0000 + 106510 \text{ ELECTRIC CONS.} + 1550.5492 \text{ POPULATION} + 2.9418 \text{ IMPORT} - 647406 \text{ LENDING INTEREST}$$

Regression Summary

Metric	Value
Residual DF	45
R2	0.93500649
Adjusted R2	0.92922929
Std. Error Estimate	4.4832E+10
RSS	9.0445E+22

Training: Prediction Summary

Metric	Value
SSE	9.04448E+22
MSE	1.8089E+21
RMSE	42531121976
MAD	31436674669
R2	0.935006486

At Population, The Sum of squared estimate is 904448E+22. Also, there is a minor decrease in SSE from 1.30115E+24 to 9.04448E+22. Therefore, this deviation speaks to the better fit of the line to the data.

The MSE = 18089
 The RMSE = 42531
 The MAD = 31436

Comparing these two models, the model with 4 regressors which gives the highest Adjusted R-square seems to be the best. This model also gives us the lowest RMS error and total sum of squared errors in the training data.

Owing to the above analysis and significance level, it is important to state that electricity consumption is not a major contributor to the current GDP buoyancy that Nigeria as a country currently enjoys. It is also worthy to note that the Unemployment rate in the country has been high from time immemorial. One would have concluded that a country with low electricity consumption and high unemployment rate would rank among countries with the lowest GDPs however, that is not the case for Nigeria.

This therefore means that since other variables like population, bank lending interest rates, importation cost and inflation rate contribute significantly to the GDP, the government should implement policies geared towards;

1. Increased birth rate through provisions of outstanding health care centers for pregnant women.
2. Reduced cost of borrowing loans of businesses from commercial banks, reducing the interest rates to a one digit percent to a figure less than 5%
3. Reduced importation costs by reducing tariffs and import duties
4. Implement effective price legislation aimed at reducing the general price levels of goods and services.

4 Comparing Logistic Regression and Neural Network Model in Selecting the Best Model that Performs Better

Logistic Regression Model Specification:

The logistic regression equation that captures the effect of energy consumption, vis-a-vis other macroeconomic variables, on GDP is captured by the function as stated below:

First;

$$\hat{p} = \frac{1}{1 + e^{-(a+bX)}}$$

Explicitly;

$$P = 1 / 1 + e^{-(\beta_0 + \beta_1 \text{ElectricityConsumption} + \beta_2 \text{InflationRate} + \beta_3 \text{TotalPopulation} + \beta_4 \text{ImportationCost} + \beta_5 \text{BankLendingInterestRate} + \beta_6 \text{UnemploymentRate} + \epsilon)}$$

Where; P is the probability value of dependent variable that captures GDP

- X1 is one of the independent variables that captures electricity consumption
- X2 captures inflation rate
- X3 captures total population
- X4 captures importation cost
- X5 captures bank lending interest rate
- X6 captures Unemployment rate
- € is the error term

For report 1, the linear regression model was adopted and Y, the dependent variable was numerical. However, for project 2, the logistic regression is adopted and instead of Y to be numerical, Y would be 1 or 0 hence, the basis for the logistic regression model.

Logistic regression is similar to the linear regression adopted in project 1 but it is different such that logistic regression is used when a categorical response is involved.

The same data set used in previous analysis would be further used in the logistic regression model but Y would be grouped into two classes (1 & 0) by using some meaningful criteria.

For this study, the mean of the GDP will be used and for GDP value above the mean, it would be tagged “1” and below the mean, it will be tagged “0” hence, high GDP/low GDP.

The mean of the GDP is 171,892,422,082.72 hence, if (GDP > 171,892,082.72), GDP is 1 otherwise, zero.

The new GDP column was created to account for the new GDP value (1,0) for logistic regression purpose.

4.1 Analysis

Table 8. Logistics regression summary

Metric	Value
# Iterations Used	7
Residual DF	46
Residual Deviance	6.5022606
Multiple R2	0.89856629

Coefficients

Predictor	Estimate	Confidence Interval: Lower	Confidence Interval: Upper	Odds	Standard Error	Chi2-Statistic	P-Value
Intercept	3.47900552	-2373.209646	2380.167657	32.42746	1212.618533	8.2312E-06	0.9977109
Electric power consumption (kWh per capita)	-0.34859146	-39.24684646	38.54966354	0.705681	19.84641315	0.00030851	0.9859863
Inflation, consumer prices (annual %)	0.14888575	-106.9390203	107.2367918	1.16054	54.63769075	7.4254E-06	0.9978258
Population, total	9.2744E-07	-7.03935E-05	7.22484E-05	1.000001	3.63889E-05	0.00064958	0.9796666
Imports of goods and services	1.5956E-10	-3.86599E-08	3.8979E-08	1	1.98062E-08	6.4903E-05	0.9935721

Predictor	Estimate	Confidence Interval: Lower	Confidence Interval: Upper	Odds	Standard Error	Chi2-Statistic	P-Value
(current US\$)							
Lending interest rate (%)	-3.16257576	-117.2854124	110.9602609	0.042317	58.22700701	0.00295007	0.9566845
Unemployment, total (% of total labor force) (modeled ILO estimate)	-10.3785383	-585.1604265	564.40335	3.11E-05	293.2614542	0.00125246	0.9717687

All the variables were insignificant at 10% or 20% or 30% or 40% at the first logistic regression output hence, we decided to drop a couple of the variables and rerun the logistic regression.

Table 9. Removing some variables and re-running the regression gives the following printout

Coefficients

Predictor	Estimate	Confidence Interval: Lower	Confidence Interval: Upper	Odds	Standard Error	Chi2-Statistic	P-Value
Intercept	-22.7561983	-62.44530653	16.93290983	1.31E-10	20.24991709	1.262853355	0.2611117
Electric power consumption (kWh per capita)	0.23779383	-0.02736459	0.502952255	1.268448	0.135287395	3.089487916	0.0787996
Lending interest rate (%)	-0.48161399	-1.574837056	0.611609076	0.617785	0.55777712	0.74555004	0.3878888
Unemployment, total (% of total labor force) (modeled ILO estimate)	1.03777522	-8.712239312	10.78778975	2.82293	4.97458862	0.043520337	0.8347486

At the second logistic regression output, electric power consumption and lending rate are significant at 40% because their probability values (0.078 and 0.387) are lesser than 40% significance level. To achieve this, we dropped the inflation rate variable, population variable and imports of goods variable.

The estimated logistic regression function is given as;

$$\hat{P} = \frac{1}{1 + e^{-(a+bX)}}$$

and for the purpose of this project, the logistic regression function is written as:

$$\hat{P} = 1 / 1 + e^{-(\beta_0 + \beta_1X_1 - \beta_2X_2 + \beta_3X_3)}$$

Where, $\beta_0 = -22.756$, $\beta_1 = 0.237$, $\beta_2 = -0.481$, $\beta_3 = 1.037$

From the above Model with 3 Regressors, the estimated regression equation is:

$$\hat{P} = 1 / 1 + e^{-(22.756 + 0.237 \text{ ELECTRIC CONS} - 0.481 \text{ LENDING RATE} - 1.037 \text{ UNEMPLOYMENT RATE})}$$

At 40 % Significant level:

- v. Reject H0: $\beta_1 = 0$ since p-value of 0.078 is less than 40%
- vi. Reject H0: $\beta_2 = 0$ since p-value of 0.387 is less than 40%
- vii. Do not reject H0: $\beta_3 = 0$ since p-value of 0.834 is greater than 40%

That is, the regression coefficient β_1, β_2 except β_3 all have p-value less than 40%, thus, the β_1 and β_2 coefficients are significant at 40%. That is, we all reject the $H_0: \beta_1 = 0, \beta_2 = 0$ at 40% significance level, and that the regression coefficients (β_1, β_2) are statistically different from 0.

Therefore, in a case, say year 2021, where Nigeria’s electric power consumption (Kwh per capita) was 133.56, lending interest rate (converted from percentage) was 0.114 and unemployment rate (converted from percentage) was 0.0978, the logistic equation would be estimated as;

$$\hat{P} = 1 / 1 + e^{-(22.75 + 0.237*133.56 - [-0.481*0.114] + 1.037*0.0978)} = 14.395$$

Hence, the estimated probability $\hat{P} = 1 / 1 + e^{-14.395} = 0.99$

We estimated that the probability of a higher GDP in Nigeria is 0.99 given that Nigeria as a country consumes 133.56 (Kwh per capita), fixes interest rate at 11.4% where unemployment rate stands at 9.78 in 2021.

5 Training & Validation Data

For the purpose of this project, the data set adopted is divided randomly into training (70%) and validation (30%) data.

Table 10. Partition Summary

Partition	# Records
Training	35
Validation	15

Detailed Report Output:

Training: Classification Details: This table reflects the calculation of GDP for each of the selected years in the data using the logistic regression to predict the chance of high GDP (that is, calculating the probability that the GDP for a particular year is in group 1)

In the table below, there are 50 observations. Xlminer employs the logistic regression form to plug in the x values and find the probabilities to be in class one for each of the observations.

It should be noted that the default cut-off is 0.5, so a probability value exceeding 0.5 assigns “1” to the group and “0” if probability value is less than 0.5.

For instance, record 1 to 32 have probability values less than 0.5 as such these records have low chance to be in group 1 hence, they are classified into group 2 as “Low GDP”. Hence, comparing the columns of zeros and ones in the table show how good the model classifies.

Overall, the table explains how many GDP observations are zeros (low GDP) and are classified correctly. It also explains the number of GDP observations that are ones (high GDP) and are classified correctly and by extension shows observations that are misclassified as indicated by the “red” print in the table.

Table 11. Prediction of Training: Classification Details

Record ID	Y = NEW GDP FOR LOGISTIC REGRESSION)	Prediction: Y = NEW GDP FOR LOGISTIC REGRESSION)	PostProb: 1	PostProb: 0
Record 1	0	0	7.24751E-07	0.999999275
Record 2	0	0	1.31021E-06	0.99999869
Record 3	0	0	7.1867E-07	0.999999281
Record 4	0	0	2.1747E-05	0.999978253
Record 5	0	0	7.49916E-05	0.999925008
Record 6	0	0	0.000568917	0.999431083
Record 7	0	0	0.000557531	0.999442469
Record 8	0	0	0.000227091	0.999772909
Record 9	0	0	0.001377473	0.998622527

Record ID	Y = NEW GDP FOR LOGISTIC REGRESSION)	Prediction: Y = NEW GDP FOR LOGISTIC REGRESSION)	PostProb: 1	PostProb: 0
Record 10	0	0	1.81099E-05	0.99998189
Record 11	0	0	0.020515237	0.979484763
Record 12	0	0	0.015771486	0.984228514
Record 13	0	0	0.000129458	0.999870542
Record 14	0	0	0.014664131	0.985335869
Record 15	0	0	0.119153286	0.880846714
Record 16	0	0	0.014413318	0.985586682
Record 17	0	0	0.00255789	0.99744211
Record 18	0	0	0.006212442	0.993787558
Record 19	0	0	9.47936E-05	0.999905206
Record 20	0	0	0.001085578	0.998914422
Record 21	0	0	0.000120671	0.999879329
Record 22	0	0	5.81506E-05	0.999941849
Record 23	0	0	0.003477128	0.996522872
Record 24	0	0	0.001455254	0.998544746
Record 25	0	0	0.000451259	0.999548741
Record 26	0	0	0.000471194	0.999528806
Record 27	0	0	0.000116074	0.999883926
Record 28	0	0	3.13442E-05	0.999968656
Record 29	0	0	1.38756E-05	0.999986124
Record 30	0	0	6.20124E-06	0.999993799
Record 31	0	0	0.003110725	0.996889275
Record 32	0	0	0.011555641	0.988444359
Record 33	0	1	0.806512021	0.193487979
Record 34	1	1	0.966754704	0.033245296
Record 35	1	0	0.421512505	0.578487495
Record 36	1	1	0.997760829	0.002239171
Record 37	1	1	0.984863894	0.015136106
Record 38	1	1	0.673315965	0.326684035
Record 39	1	1	0.994157048	0.005842952
Record 40	1	1	0.999894524	0.000105476
Record 41	1	1	0.999967213	3.27868E-05
Record 42	1	1	0.999061141	0.000938859
Record 43	1	1	0.999769126	0.000230874
Record 44	1	1	0.940057118	0.059942882
Record 45	1	1	0.998273344	0.001726656
Record 46	1	1	0.999727128	0.000272872
Record 47	1	1	0.99987356	0.00012644
Record 48	1	1	0.999964479	3.55212E-05
Record 49	1	1	0.999996095	3.90525E-06
Record 50	1	1	0.999998799	1.20051E-06

Table 12. Two way table

Actual	Predicted	
	0	1
0	True negative 32	False Positive 1
1	False negative 1	True Positive 16

Table 13. Logistics Regression- Prediction of Training Data

Confusion Matrix		
Actual\Predicted	0	1
0	32	1
1	1	16

The Classification summary report is based on the detailed report of the model classification as reported by Xlminer. It gives us insight of how good or bad the model performed.

6 Error Report

The error report table records that just two observations out of 50 cases are misclassified.

In our data, there are 33 cases that are “zeros” (Low GDP), 17 cases are ones (High Gdp) and out of the 33 cases, one is misclassified ($1/33 = 3.03\%$ -> False Positive Percentage of the model). Similarly, there are 17 cases that are “ones” (high GDP) and 1 case is misclassified ($1/17 = 5.88\%$ -> False Negative Percentage of the model)

This translates to the fact that our classification rule correctly classified 96% of the observations and about just 4% error in the entire model.

So, the overall misclassification error is 4% (2/50).

Table 14. Error report

Error Report			
Class	# Cases	# Errors	% Error
0	33	1	3.03030303
1	17	1	5.882352941
Overall	50	2	4

False Positive Error - This is a scenario where our model captures “1” when it is actually supposed to be “0” and that accounts for 3.03%. This is just one case out of 33 according to the table above.

False Negative Error - On the other hand, the false negative error is a scenario where it is actually “1” but the model predicts “0”. That is one case out of 17 which is 5.8%

7 Metrics

Table 15. Metrics results

Metrics	
Metric	Value
Accuracy (#correct)	48
Accuracy (%correct)	96
Specificity	0.96969697
Sensitivity (Recall)	0.941176471
Precision	0.941176471
F1 score	0.941176471
Success Class	1
Success Probability	0.5

Accuracy: Explains the correctness of the classification of the model and in this case, it is (33 & 17) as such, we correctly classified 48 cases out of 50 cases. The model records 96% accuracy, which is not a bad one.

Sensitivity: This is also known as “Recall of the Model”. The table records True Positive percentage of about 94% (33 out of 34 - positive cases were correctly classified). This is a good score for the model.

Specificity: Explains the True Negative percentage which is 96% (17 out 18 negative cases that were correctly classified). Overall, the model is okay.

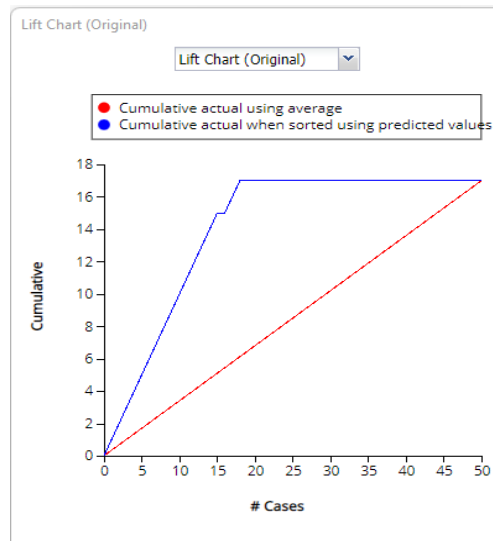


Chart 1. Lift chart

7.1 Lift Chart

The lift chart measures the model performance as such, the higher the lift chart, the better the model.

Here, the lift chart is high and by implication, the model is good

For the ROC curve, we look at the area under the curve and in this case, that is 99% which is also very good.

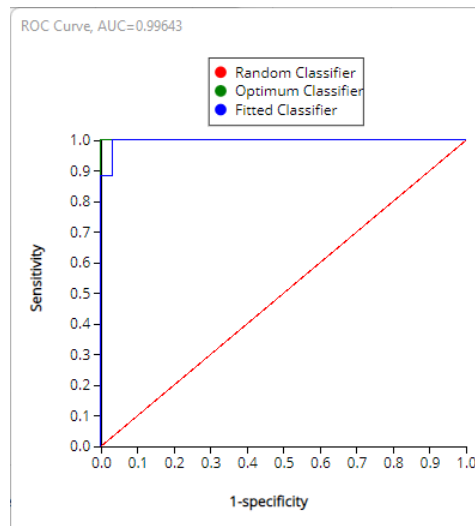


Chart 2. ROC Curve

Overall, from the outputs, it can be concluded that there is some sort of high level of consistency which also spans through the validation data.

Logistics Regression: Overall, an increased energy consumption with low interest rate and low unemployment rate leads to a high GDP in Nigeria.

Therefore, central bank policies aimed at reducing bank lending rate should be implemented to serve as incentive for firms to borrow more and increase output/production.

Policies geared towards increased power/electricity generation and supply should be implemented, thus making power available to firms at reduced costs. This would increase the overall GDP in the country. Owing to the above, as firms produce more, they would also need to hire more people to meet the increased supply of goods and services

8 Neural Network Model

8.1 Building the best model

After a series of trials and errors, in the final model; the Neurons were reduced to just 4 and one significant change we made was “Changing the variables”. The explanatory variables used in the logistics regression where: Electric Power Consumption, Lending Interest rate and Unemployment but in this final trial, the variables adopted are: Electric power consumption, Lending Interest Rate and Population.

Table 16. Neural network classification- prediction of training data

Confusion Matrix			
Actual\Predicted		0	1
0		23	1
1		0	11

Error Report			
Class	# Cases	# Errors	% Error
0	24	1	4.166666667
1	11	0	0
Overall	35	1	2.857142857

Metrics	
Metric	Value
Accuracy (#correct)	34
Accuracy (%correct)	97.14285714
Specificity	0.958333333
Sensitivity (Recall)	1
Precision	0.916666667
F1 score	0.956521739
Success Class	1
Success Probability	0.398752

Table 17. Neural network classification- prediction of validation data

Confusion Matrix			
Actual\Predicted		0	1
0		9	0
1		0	6

Error Report			
Class	# Cases	# Errors	% Error
0	9	0	0
1	6	0	0
Overall	15	0	0

Metrics	
Metric	Value
Accuracy (#correct)	15
Accuracy (%correct)	100
Specificity	1
Sensitivity (Recall)	1
Precision	1
F1 score	1
Success Class	1
Success Probability	0.398752

Training Dataset:

- Partitioned: 70-30
- Probability Cut-off: 0.398792 (Re-adjusted based on the Average of previous result)
- 3 Explanatory Variables (Electric, Population, Unemployment)- A different variable from the Logistics regression model
- Neurons: 4
- Hidden Layer: 1
- Rescale Data: Yes
- True positive (Sensitivity)- 100%
- True Negative (Specificity)- 96%
- Correctly classified 97%
- False Positive and False Negative is 4.2% and 0% respectively

Validation Dataset:

- Partitioned: 70-30
- Probability Cut-off: 0.398792 (Re-adjusted based on the Average of previous result)
- 3 Explanatory Variables (Electric, Population, Unemployment)- A different variable from the Logistics regression model
- Neurons: 4
- Hidden Layer: 1
- Rescale Data: Yes
- True positive (Sensitivity)- 100%
- True Negative (Specificity)- 100%
- Correctly classified 100%
- False Positive and False Negative is 0% and 0% respectively

Taking into account the three (3) explanatory variables from our data, and setting the probability cutoff at the default which is 0.398752, we can see:

In the training data, the Overall Misclassification Error is 1 out of 35 observations = 2.8% while in the Validation data, the Overall Misclassification Error is 0 out of 15 = 0%

On classification: our classification rule correctly classified 97% in the training and 100% in the validation data

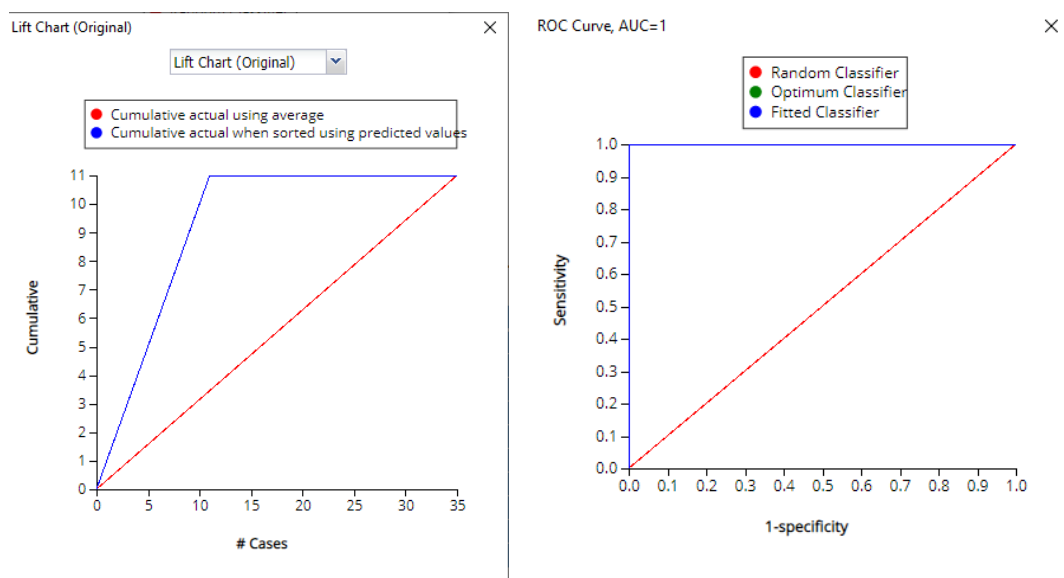


Chart 3. Training data lift chart and AUC

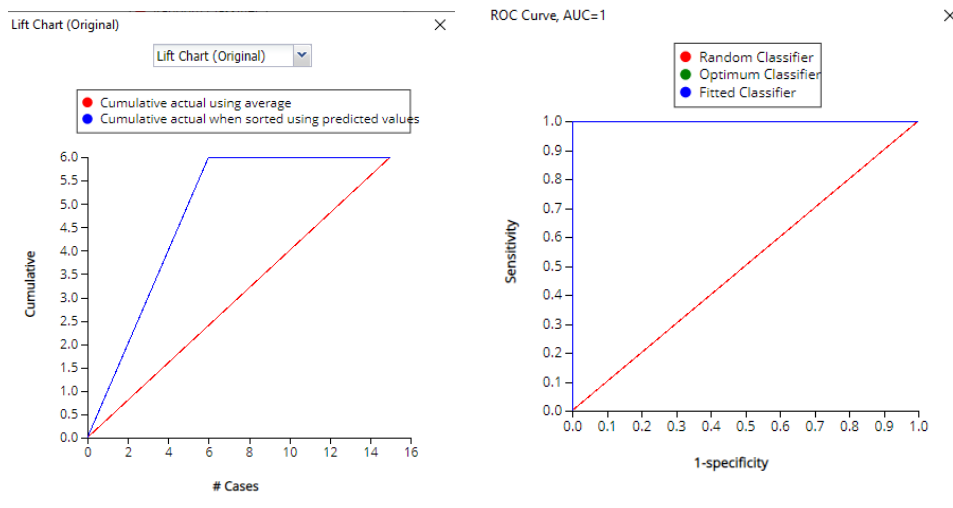


Chart 4. Validation lift chart and AUC

LIFT CHART

The lift chart measures the model performance as such, the higher the lift chart, the better the model.

Here, the lift chart in the training and validation data is high and by implication, the model is good

ROC CURVE

For the ROC curve, we look at the area under the curve and in this case, that is 100% and 100% for both the training and validation data which is a very good model performance. See charts for both Training and Validation Dataset:

9 Comparison

In the Logistic Regression model, after series of tests, Electric power consumption and lending rate are significant at 40% because their probability values (0.078 and 0.387) are lesser than 40% significance level. To achieve this, we dropped inflation rate variable, population variable and imports of goods variable. See coefficient below:

On the other hand, results from the Neural Network classification model revealed the coefficients as presented in the table 19.

9.1 Logistics Regression

Training: Classification Summary: The Logistic Regression model had an overall % Error of 5.71%, Accuracy of 94.29%, and Sensitivity of 91%. This shows the model is good. There were four (4) cases of False Positives and nine (9) cases of False Negatives. However, the Specificity stood at 95%.

On the Training data set, the Neural network seems to have a better Sensitivity figure (100%) when compared to the Logistic Regression model (91%) , and overall, the % mis-classification error for the Neural Network model is also lower (2.71%) compared to the overall mis-classification %error (5.71%) in the logistics regression model. This is impressive owing to the complexity of Neural Network models.

Therefore, the Neural Network model performs better for our dataset owing to the overall misclassification % and accuracy metrics.

Recall, the ROC curve speaks to the better performance of the model and this aims closer to 1. The AUC in the Neural Network is at exactly 100% while the AUC in the Logistics regression model is 99.6% (See below) in both the training and Validation dataset. Judging from this, the NN model performs better for our dataset.

Table 18. Correlation coefficient result

Predictor	Estimate	Confidence Interval: Lower	Confidence Interval: Upper	Odds	Standard Error	Chi2-Statistic	P-Value
Intercept	-18.4229734	-36.30049004	-0.545456741	9.98E-09	9.121349572	4.079446569	0.0434079
Electric power consumption (kWh per capita)	0.23234233	-0.023443657	0.488128314	1.261552	0.130505452	3.169562844	0.0750225
Lending interest rate (%)	-0.46715088	-1.515746415	0.581444651	0.626786	0.535007552	0.762420423	0.3825721

Table 19. Neuron weight result

Neuron Weights: Input Layer - Hidden Layer 1					
Neurons	Electric power consumption (kWh per capita)	Lending interest rate (%)	Unemployment, total (% of total labor force) (modeled ILO estimate)		Bias
Neuron 1	-0.095485075	0.362711386	-1.325130979		-0.0347214
Neuron 2	0.521630007	0.542443597	0.491132982		-0.0125665
Neuron 3	-0.144760377	0.605782742	-0.093453618		-0.0435881
Neuron 4	0.833419481	-0.49237841	-0.37513595		-0.010458

Neuron Weights: Hidden Layer 1 - Output Layer					
Neurons	Neuron 1	Neuron 2	Neuron 3	Neuron 4	Bias
0	0.286756588	0.315081249	-0.868064399	0.49326575	0.397142
1	0.024789237	0.331743867	-0.604229897	0.39580746	-0.53512

Table 20. Logistics Regression- Prediction of Training Data

Confusion Matrix			
Actual\Predicted		0	1
0		23	1
1		1	10

Error Report			
Class	# Cases	# Errors	% Error
0	24	1	4.166666667
1	11	1	9.090909091
Overall	35	2	5.714285714

Metrics	
Metric	Value
Accuracy (#correct)	33
Accuracy (%correct)	94.28571429
Specificity	0.958333333
Sensitivity (Recall)	0.909090909
Precision	0.909090909
F1 score	0.909090909
Success Class	1
Success Probability	0.5

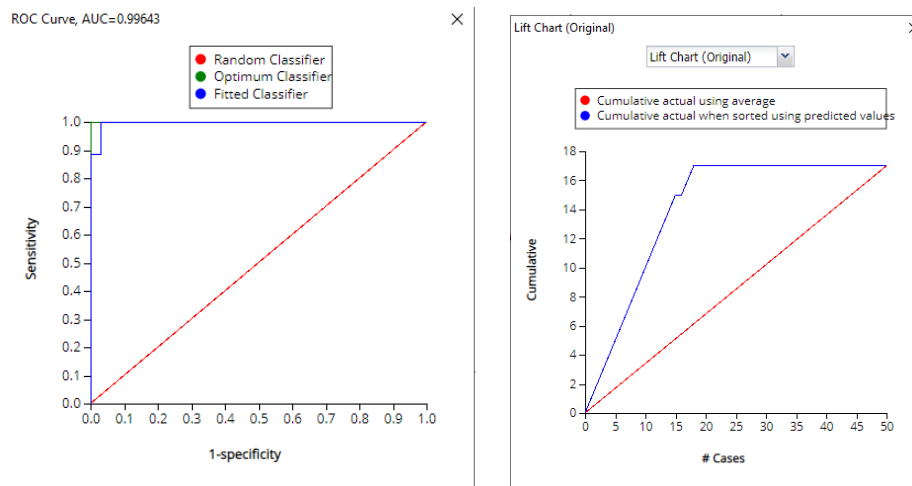


Chart 5. ROC and Lift Chart

10 Conclusion

The results of the Neural Network model performed better than the Logistics Regression model in this study. There is need for the improvement of the Logistic Regression model for a better GDP classification.

11 Recommendation

Electric power consumption and Lending interest rate are the best regressors to study Nigeria's Gross Domestic Product (GDP) in the span of 50 years

Using the Neural Network model, Electric power consumption, Population and Lending interest Rate when compared to the performance of the Logistic Regression model, the Neural Network model seems to be a better model to be used in this classification project. The Neural Network model performed better across the metrics including Confusion Matrix, Accuracy, Specificity, AUC curve, Lift Chart, and the Sensitivity report.

Out of the many predictor variables we tried, only the Electric power consumption, Population and Lending interest Rate are used to determine the high or low GDP rate in our 50 case observation and it is very interesting to find that all other predictors like Unemployment, Importation cost and Inflation rate do not have much effect.

Competing Interests

Authors have declared that no competing interests exist.

References

- [1] Iyke BN. Electricity consumption and economic growth in Nigeria: A revisit of the energy-growth debate. *Energy Economics*. 2015 Sep 1;51:166-76.
- [2] Omisakin DO. Energy consumption and economic growth in Nigeria: A bounds testing cointegration approach. *Journal of Economic Theory*. 2008;2(4):118-23.
- [3] Ogundipe AA. Electricity consumption and economic growth in Nigeria. *Energy*. 2013 Jul;100000:120000.
- [4] Alege PO, Adediran OS, Ogundipe AA. Pollutant emissions, energy consumption and economic growth in Nigeria. *International Journal of Energy Economics and Policy*. 2016 Jan 6;6(2):202-7.
- [5] Akinlo AE. Electricity consumption and economic growth in Nigeria: evidence from cointegration and co-feature analysis. *Journal of Policy Modeling*. 2009 Sep 1;31(5):681-93.
- [6] Mustapha AM, Fagge AM. Energy consumption and economic growth in Nigeria: A causality analysis. *Journal of Economics and Sustainable Development* Vol. 2015;6.
- [7] Abalaba BP, Dada MA. Energy consumption and economic growth nexus: new empirical evidence from Nigeria. *International Journal of Energy Economics and Policy*. 2013 Jan 12;3(4):412-23.
- [8] Ekeocha PC, Penzin DJ, Ogbuabor JE. Energy consumption and economic growth in nigeria: A test of alternative specifications. *International Journal of Energy Economics and Policy*. 2020;10(3):369-79.

© 2023 Ikpe and Oyedeji; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

<https://www.sdiarticle5.com/review-history/110212>