SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (SARIMA) OF NIGERIAN GROSS DOMESTIC PRODUCT

[1]A.U. Kinafa [1]S.I. Kamara [1]Abubakar Bello, [1]M. A. Aliyu [1]M.S Tabra, [1]A. Adamu

[1]Department of Mathematics, Gombe State University, Gombe.

[2] Correspondence Email: usmankinafaaliyu@yahoo.com

Phone: 08035492070.

Abstract

This paper presents a research on gross domestic product; gross domestic product (GDP) of Nigeria in 2015 records the lowest growth in four years. In this paper, the features of quarterly data of Nigeria's GDP obtained from aggregated quarterly data of Central Bank of Nigeria starting from 2010 to 2015 were studied. Time series plot of the data shows that the data is not stationary, after decomposing the data into its component parts, it was discovered that Nigeria's GDP has a trend component with seasonal component and enough residuals component. To model the GDP, a class of SARIMA (Seasonal autoregressive integrated moving average) models is built following Box-Jenkins method. Then, the efficient model is SARIMA (1,0,0)(0,1,0)[4]. According to the recognition rules and stationary test of time series under the AIC and BIC criterion, the fitted SARIMA model was then used to forecast the GDP for the next three years. The forecast showed that there is an upward trend and the relative, predicted values were within the 95% confidence interval and also Ljung-Box and Jarque- Bera Test were used to test the adequacy of the fitted model and we recommend a need for economic diversification to solve recession issued.

Keywords: Forecasting, GDP, SARIMA.

Introduction

Gross domestic product (GDP) a basic measure of an economy's economic performance, is the market value of all final goods and services produced within the borders of a nation in a year. GDP can be defined in three ways, all of which are conceptually identical. First, it is equal to the total expenditure for all final goods and services produced within the country in stipulated period of time. Second it is equal to the sum of value added at every stage of production by all industries within a country, plus taxes less subsidies on products in the period. Third it is equal to the income generated by the production in the country in the period that is compensation of employees, taxes on production and imports less subsidies and gross operating surplus or profits, the primary indicators used to measure the healthiness of a country's economy. It is also used to determine the standard of living of individuals in an economy. However, Gross Domestic Product could be defined as the market value of all officially recognised final goods and services produced within a country in a given period of time. This implies that Gross Domestic Product takes into account the market value of each good or service rather than adding up the quantities of the goods and services directly. Gross Domestic Product is important in an economy because it is used to determine if an economy is growing more quickly or more slowly. Also, it is used to compare the size of economies throughout the world. Again, the Gross Domestic Product is used in the comparison of relative growth rate of economies throughout the world. For instance, the Federal Reserve's in the United States uses it as one of the indicators of whether the economy needs to be restrained or stimulated. The components of Gross Domestic Product using the expenditure method includes; Consumption, Investment, Government expenditure, Gross export and Gross import. Where this could be expressed mathematically as GDP = C + I + G + (X - M). There are two other methods of calculating the Gross Domestic Product which are the Value Added (or Production) approach and the Income (or By Type) approach. In calculating Gross Domestic Product using either of the three approaches, it does not include intermediate goods, but only "new" products (final goods) and services, this is to avoid double counting which may lead to the presentation of an inaccurate value of GDP. There are two types of GDP in existence which includes; Real GDP and Nominal GDP. Where Real GDP is the measurement of economic output of a country minus the effect of inflation and Nominal GDP is the measurement that leaves the price changes in the estimate. In this work, gross domestic product will be used in analyzing the Nigerian economy. Also, models will be fitted for future predictions of the state of the economy based on its previous GDP data with an aim of describing the pattern in which Gross Domestic Product in Nigeria.

Methodology

The basic methodology approach was to develop models, of increasing sophistication, for quarterly series data, the models were then tested to gauge their forecasting accuracy.
To model the GDP, a class of SARIMA (Seasonal autoregressive integrated moving average) models is built following Box-Jenkins method. Then, the efficient model is SARIMA (1,0,0) (0,1,0)[4]. According to the recognition rules and stationary test of time series under the AIC and BIC criterion, the fitted SARIMA model was then used to forecast the GDP for the next three years. The forecast showed that there is an upward trend and the relative, predicted values were within the 95% confidence interval and also Ljung-Box and Jarque- Bera Test were used to test the adequacy of the fitted model.
The analysis of variable under study was done using time series analysis. Time series analysis consists of several models and the use of each model depends on the nature of data. For the purpose of this study time series analysis was adopted.

Seasonal autoregressive integrated moving average models (sarima)

Seasonality in a time series is a regular pattern of changes that repeats over S time periods, where S defines the number of time periods until the pattern repeats again. For example, there is seasonality in monthly data for which high values tend always to occur in some particular months and low values tend always to occur in other particular months. In this case, S = 12 (months per year) is the span of the periodic seasonal behavior. For quarterly data, S = 4 time periods per year. In a seasonal ARIMA model, seasonal AR and MA terms predict x_t using data values and errors at times with lags that are multiples of S (the span of the seasonality).

- With monthly data (and $S = 12$), a seasonal first order autoregressive model would use x_{t-12} to predict x_t. For instance, if we were selling cooling fans we might predict this August's sales using last August's sales. (This relationship of predicting using last year's data would hold for any month of the year.)
- A seasonal second order autoregressive model would use x_{t-12} and x_{t-24} to predict x_t. Here we would predict this August's values from the past two Augusts.
- A seasonal first order MA(1) model (with $S = 12$) would use w_{t-12} as a predictor. A seasonal second order MA(2) model would use w_{t-12} and w_{t-24}.
- Differencing
  Almost by definition, it may be necessary to examine differenced data when we have seasonality Seasonality usually causes the series to be nonstationary because the average values at some particular times within the seasonal span (months, for example) may be different than the average values at other times. For instance, our sales of cooling fans will always be higher in the summer months. Seasonal differencing is defined as a difference between a value and a value with lag that is a multiple of S.
- With $S = 12$, which may occur with monthly data, a seasonal difference is $(1-B^{12})x_t = x_t - x_{t-12}$. The differences (from the previous year) may be about the same for each month of the year giving us a stationary series.
- With $S = 4$, which may occur with quarterly data, a seasonal difference is $(1-B^4)x_t = x_t - x_{t-4}$. Seasonal differencing removes seasonal trend and can also get rid of a seasonal random walk type of nonstationarity.
  Non-seasonal differencing: If trend is present in the data, we may also need non-seasonal differencing. Often (not always) a first difference (non-seasonal) will "detrend" the data. That is, we use $(1-B)x_t = x_t - x_{t-1}$ in the presence of trend.
  Differencing for Trend and Seasonality; when both trend and seasonality are present, we may need to apply both a non-seasonal first difference and a seasonal difference. That is, we may need to examine the ACF and PACF of $(1-B^{12})(1-B)x_t = (x_t - x_{t-1}) - (x_{t-12} - x_{t-13})$.

Removing trend doesn't mean that we have removed the dependency. We may have removed the mean, μ_t, part of which may include a periodic component. In some ways we are breaking the dependency down into recent things that have happened and long-range things that have happened.

Non-seasonal Behaviour Will Still Matter. With seasonal data, it is likely that short

run non-seasonal components will still contribute to the model. In the monthly sales of cooling fans mentioned above, for instance, sales in the previous month or two, along with the sales from the same month a year ago, may help predict this month's sales.
We'll have to look at the ACF and PACF behaviour over the first few lags (less than S) to assess what non-seasonal terms might work in the model.

Seasonal ARIMA Model

The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model is

ARIMA $(p, d, q) \times (P, D, Q)S$,

With p = non-seasonal AR order, d = non-seasonal differencing, q = non-seasonal MA order, P = seasonal AR order, D = seasonal differencing, Q = seasonal MA order, and S = time span of repeating seasonal pattern.

Without differencing operations, the model could be written more formally as

$$\Phi(B^S)\varphi(B)(x_t - \mu) = \Theta(B^S)\theta(B)w_t$$

The non-seasonal components are: AR:

$\varphi(B) = 1 - \varphi_1 B - ... - \varphi_p B^p$, MA: $\theta(B) = 1 + \theta_1 B + ... \theta_q B^q$

The seasonal components are: Seasonal AR: $\Phi(B^S) = 1 - \Phi_1 B^S - ... - \Phi_P B^{PS}$ Seasonal MA: $\Theta(B^S) = 1 + \Theta_1 B^S + ... + \Theta_Q B^{QS}$

Note that on the left side of equation (1) the seasonal and non-seasonal AR components multiply each other, and on the right side of equation (1) the seasonal and non-seasonal MA components multiply each other.

Example 1: SARIMA $(0, 0, 1) \times (0, 0, 1)_{12}$

The model includes a non-seasonal MA(1) term, a seasonal MA(1) term, no differencing, no AR terms and the seasonal period is $S = 12$.

The non-seasonal MA (1) polynomial is $\theta(B) = 1 + \theta_1 B$.

The seasonal MA (1) polynomial is $\Theta(B^{12}) = 1 + \Theta_1 B^{12}$.

The model is $(x_t - \mu) = \Theta_1(B^{12})\ \theta_1(B)w_t = (1 + \Theta_1 B^{12})(1 + \theta_1 B)w_t$.

When we multiply the two polynomials on the right side, we get

$$(x_t - \mu) = (1 + \theta_1 B + \Theta_1 B^{12} + \theta_1\Theta_1 B^{13})w_t$$

$$= w_t + \theta_1 w_{t-1} + \Theta_1 w_{t-12} + \theta_1\Theta_1 w_{t-13}.$$

Thus the model has MA terms at lags 1, 12, and 13. This leads many to think that the identifying ACF for the model will have non-zero autocorrelations only at lags 1, 12, and 13. There's a slight surprise here. There will also be a non-zero autocorrelation at lag 11. We supply a proof in Appendix 1 for this document.

The Box–Jenkins Methodology for SARIMA Models

The basis of the Box-Jenkins approach to modelling time series consists of three phases: identification, estimation and testing, and application (Spyros et al., 1998). Box and Jenkins effectively put together in a comprehensive manner the relevant information required to understand and use univariate time series ARIMA models. The theoretical underpinnings described by Box and Jenkins and later by Box, Jenkins, and Reinsell (1994) are quite sophisticated, but it is possible for the non-specialist to get a clear understanding of the essence of ARIMA methodology. Application of a general class of forecasting methods involves two basic tasks: (*a*) Analysis of the data series and (*b*) Selection of the forecasting model that best fits the data series. Thus, in using a smoothing method, analysis of the data series for seasonality, aids in selection of a specific smoothing method that can handle the seasonality.

Data analysis

The analysis of this work is based on the data collected from quarterly data of gross domestic product (GDP) of Nigeria From 2010 to 2015 sourced from the CBN statistics data base. The data were subjected to time series analysis using R Statistical package version .3.3.4 in order to determined pattern presence in the series of GDP

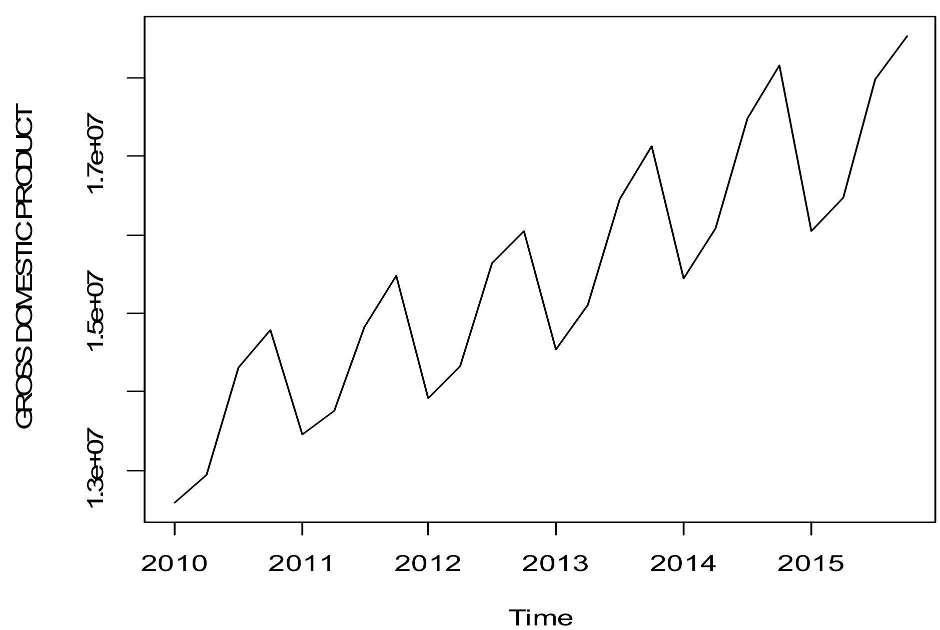

Figure 4.1 Plot of the original series of GDP

In order to determine the pattern presence in the series of gross domestic product the above plot indicate a trend as well as seasonal pattern at every fourth quarter of every year, which shows that the series of GDP is non stationary the plot also suggest a seasonal differencing in order to make the series stationary. To do this first of all we check the presence of auto correlation.

Test for autocorrellation

In order to check for autocorrelation between the error terms, the Durbin Watson test for autocorrelation was performed. The hypothesis is stated as;

H_0: $\rho = 0$ or $d = 2$ (no autocorrelation) vs.
H_1: $\rho \neq 0$ or $d \neq 2$ (autocorrelation exist)

Table 4.1: Summary statistics for gross domestic product

Statistic	Values
Mean	15477026
Standard deviation	1634457
Akaike criterion	546.43
Schwarz criterion	368.044
Hannan-quinn	3683.388
Rho	0.921613
Durbin Watson	0.133109

Decision: if $d < 2$ there is positive serial autocorrelation, If $d = 2$ no serial autocorrelation, when $d > 2$ negative autocorrelation exist. Since Durbin-watson statistic (d) =0.133109 is substantially less than 2 we reject H_0 and conclude that there is an evidence of positive autocorrelation.

Test for stationarity (unit root test)

Augmented dickey fuller (ADF) test and The Kwiatkowski-Phillips-Schmidt-shin (kpss) test are used to test for Stationary in the series of gross domestic product the hypothesis is stated as follows;

H_0: $\varphi = 0$ (The series is not stationary)

H_1: $\varphi \neq 0$ (The series is stationary)

Table 4.2 Adf test statistics

Test statistic	Values	p-value	Decision
ADF	-2.711	0.2869	Accept

Since ADF p-value (0.2869)>level of significant (0.05) we do not reject H_0 and conclude that the series is not stationary (unit root exist), and therefore series is stationary at integrated of order one I(1) with p-value 0.01598 which is less than level of significance 0.05

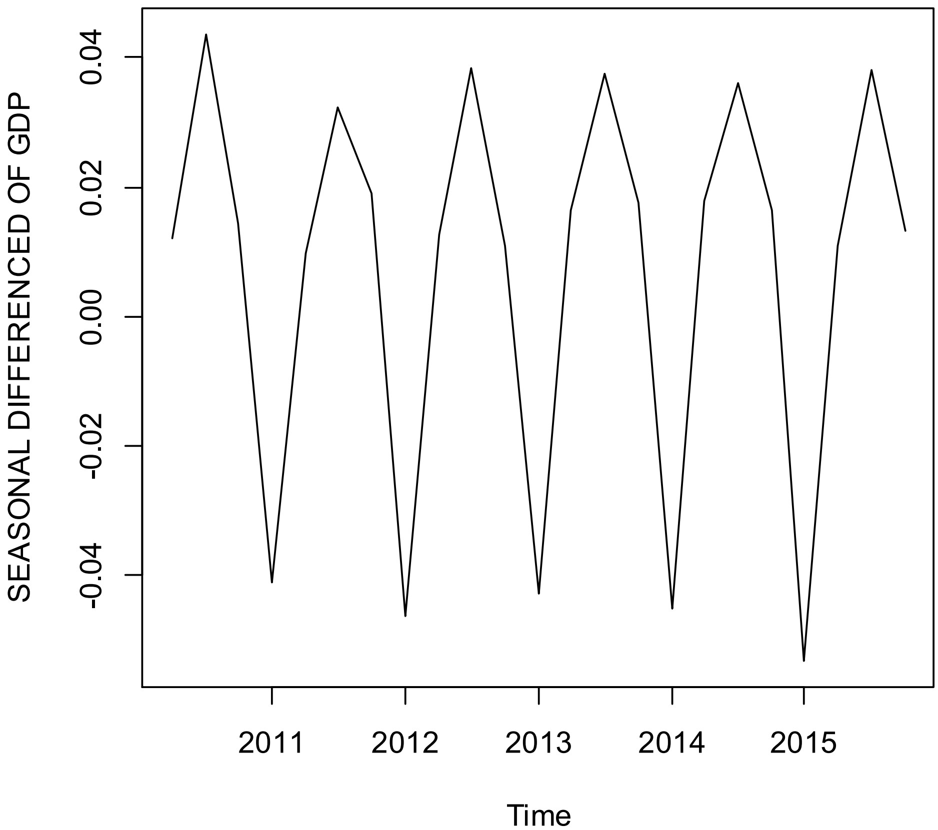

Figure 4.3: Plot of seasonal differenced of GDP

Now we clear the trend as well as seasonality in the series of gross domestic product by making the series stationary in term of mean and variance

Plots of autocorrelation and partial autocorrelation function

Before we can determine the estimate of parameters for time series models, we firstly identify the order of the models, the autocorrelation function (ACF) and partial autocorrelation function (PACF) factor plot help to identify the pattern in the stationary series of gross domestic product, the idea is to identify the presence of AR and MA components .

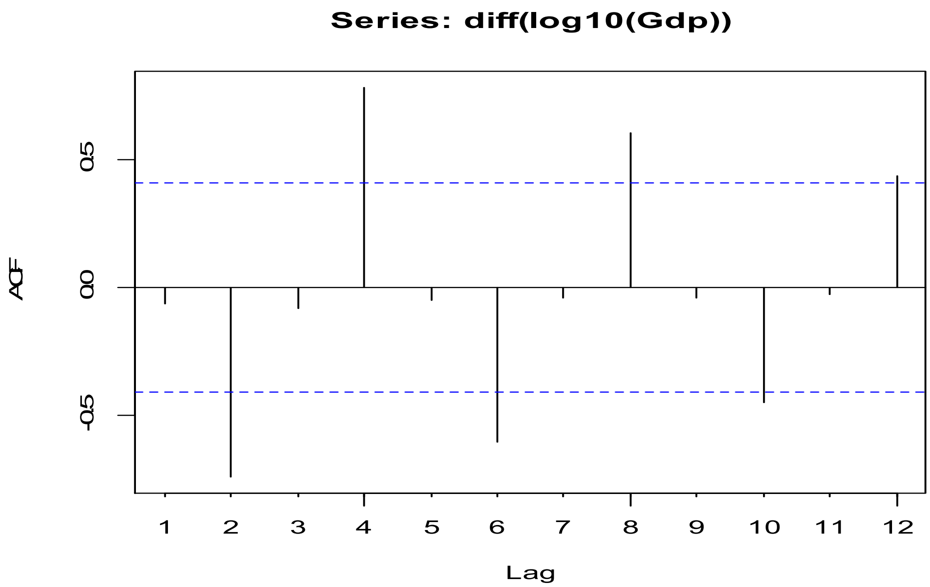

Figure 4.3: Autocorrelation function of GDP

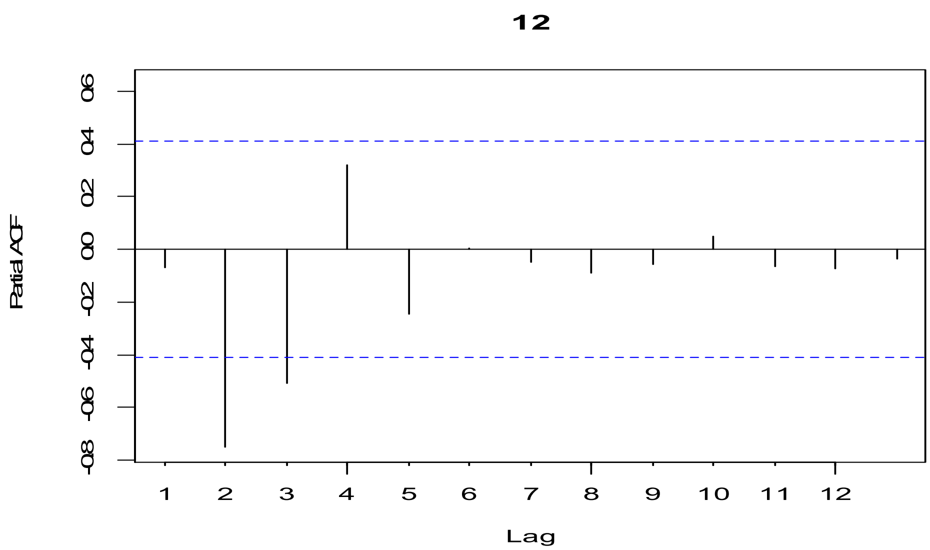

Figure 4.4 Partial autocorrelation function of GDP

Since there are spikes in the plots outside the insignificant zone we conclude that the residuals are not random. This implies that there is information available in residuals to be extracted by AR and MA models, 12 postulated models were examine from which, based on the model selection criterion of AIC and BIC, the appropriate model is selected. The postulated models together with selection criteria are presented in table below

Table 4.3: Postulated models and evaluation

Models	AIC	BIC
SARIMA(1,0,0)(0,1,0)[4]	546.15	549.14
SARIMA(1,0,2)(0,1,0)[4]	581.43	557.40
SARIMA(1,0,3)(0,1,0)[4]	554.22	559.33
SARIMA(2,0,0)(0,1,0)[4]	551.46	554.68
SARIMA(2,0,1)(0,1,0)[4]	554.03	558.46
SARIMA(3,0,1)(0,1,0)[4]	553.89	557.58
SARIMA(3,0,2)(0,1,0)[4]	552.84	553.23
SARIMA(3,0,3)(0,1,0)[4]	556.12	557.74
SARIMA(3,0,4)(0,1,0)[4]	555.54	556.22
SARIMA(4,0,2)(0,1,0)[4]	558.44	559.02
SARIMA(4,0,0)(0,1,0)[4]	557.91	558.36
SARIMA(4,0,3)(0,1,0)[4]	559.17	559.74

Using the forecast package In R and model selection criterion of AIC and BIC. SARIMA (1, 0, 0) (0, 1, 0)[4] efficient model with minimum value of AIC and BIC, hence the selected model to fit the series of gross domestic product. The chosen SARIMA of the form:

$$\widehat{Y} = \delta + y_{t-4} + \varphi_1 (y_{t-1} - y_{t-5})$$

Table 4.4 Parameters estimate of the fitted sarima model

PARAMETER	ESTIMATE	STD ERROR	Z	P-VALUE
Δ	177060.72	22827.48		
Φ_1	0.598	0.185	31.75	0.006

$$\widehat{Y} = 177060.72 + y_{t-4} + 0.598 (y_{t-1} - y_{t-5})$$

Table 4.5 Predicted gross domestic products

YEAR	Q1	Q2	Q3	Q4
2016	16564587	17055419	18615011	19200454
2017	17247989	17748807	19314370	19903385
2018	17953055	18455150	20021477	20610949

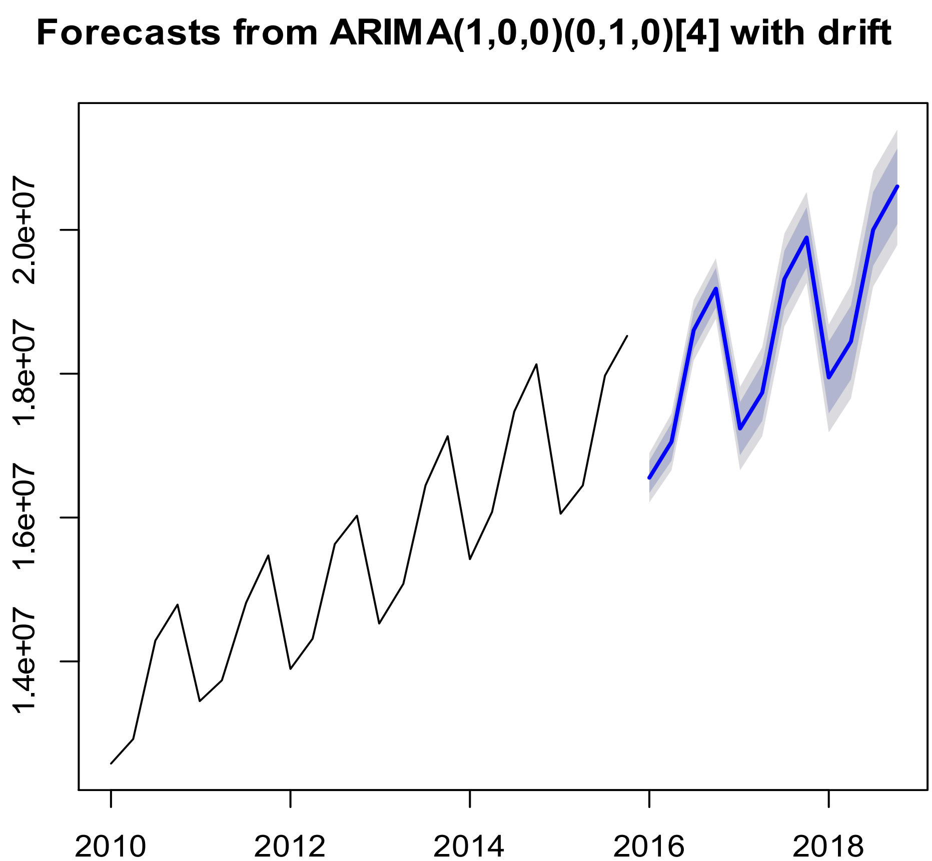

Figure 4.5 Plot of the forecasted GDP

From the above plot the result shows that there is an upward trend which indicates that Nigeria is experiencing economic growth of 1.7% in 2016, 1.9% in 2017 and 2% in 2018 which also implies that in the nearest future Nigeria will experience a better economy than it has experienced in the past

After fitting the model, we should check whether the model is appropriate that is we want to show that $\mathcal{E}_t$ is independent and identically distributed ($\mathcal{E}_t \sim$ iid N (0,1))

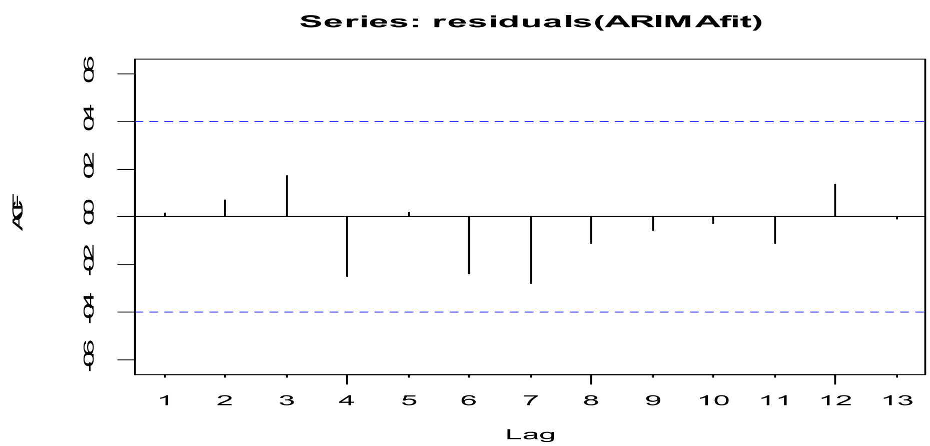

Figure 4.6: ACF plots of the residuals

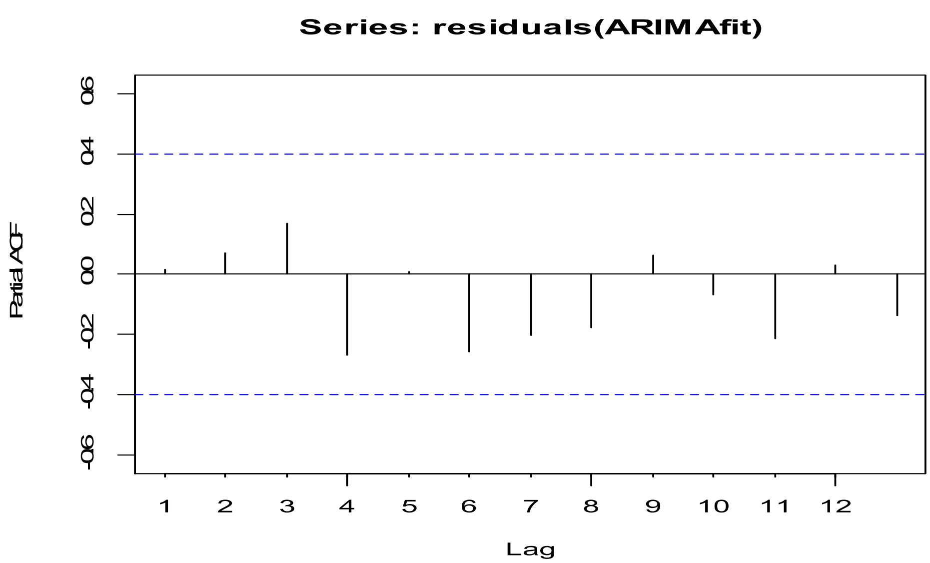

Figure 4.7: PACF plots of the residuals

Table 4.6: ljung box and jaque bera statistic

Test statistic	Value	p-value
Jaque-Bera test	0.8355	0.6585
Ljung-Box	4.126	0.4357

H_0: error terms are normally distributed
H_1 error terms are not normally distributed

From the Jaque bera and Ljung-box test statistic with p-value of 0.6585 and 0.4357 which are greater than 0.05 level of significant we fail to reject the null hypothesis of normality distribution of the residuals, and also the ACF and PACF PLOTS of the residual in figure 4.6 and 4.7 shows that the residuals are normally distributed since there is no spike in the plots out side the insignificance zone which indicate the adquacy of the model.

Conclusion

Based on the above summary, we therefore conclude that there is an upward trend in the Nigerian GDP. By fitting SARIMA (1, 0, 0) (0,1,0)[4] comparing with other models, we found that the above model performed efficient. The model was then used to forecast the GDP for the next three years, the forecast showed that the relative and predicted values were within the 95% confidence interval and also Ljung-Box and Jarque- Bera were used to test the adequacy of the model.

Recommandation

Based on the above conclusion, we therefore recommend that Since the price of crude oil is nearly halved between the third quarter of 2014 and second quarter of 2015, Government should diversifies the economy to prevent the country from being in recession in near future. There is a need for government to ensure that Nigerians Naira appreciate foreign currencies in the exchange rate market

References

Ayinde, K. and Abdulwahab, H. (2013) *Modelling and forecasting Nigerian crude oil Exportation:* Seasonal Autoregressive Integrated Moving Average Approach, International Journal of Science and Research, 2(12):245-249.

AfDB, OECD, UNDP, UNECA (2011). *African Economic Outlook*. Available online at http://www.africaneconomicoutlook.org/fileadmin/uploads/aeo/Country_Notes/2011/ Full/Nigeria.pdf.

Ajakaiye, O. and Fakiyesi, T. (2009). *Global Financial Crisis Discussion Series; Paper 8: Nigeria*. Overseas Development Institute. London.

Archer, D. and Halliday, J. (1998). "The Rationale for Holding Foreign Currency Reserves" *Of Reserve Bank New Zealand Bulletin,* 61(4):346-354.

Box, G. E., & Pierce, D. A. (1970). *Distribution of residual autocorrelations* in autoregressive-integrated moving average time series models. Journal of the American statistical Association, 65(332), 1509-1526.

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control, revised.*

Central Bank of Nigeria. (2007). The Bullion: Building and Management External Reserves for Economic Development. 31(2). Central bank of Nigeria. (2005), *Statistical Bulletin*, Vol.16,December.

Central bank of Nigeria. (2014), *Statistical Bulletin*, Vol.20, December.

Dooley, M.P., Folkerts-Landau, D. and Garber, P. (2004), "The Revived Bretons Woods System", *International Journal of Finance and Economics*, 9(4):307–313.

Etuk, H.E, Uchendu, B., Udo, E.O. (2012). Box-Jenkins Modeling of Nigerian Stock Prices Data, Greener Journal of Science Engineering and Technological Research 2(2):32-38.

Gregory C. Chow (1960). *Tests of Equality Between Sets of Coefficients in Two Linear Regressions.*Econometrica,Geisser, Seymour (1 June 1993) *Predictive Inference: An Introduction. Chapman & Hall, CRC.*

He, C. and R. Sandberg (2006) *Dickey-Fuller Type of Tests against Nonlinear Dynamic Models.* Oxford Bulletin of Economics and Statistics, 68, p. 835-861.

James D. Hamilton (1994). *Time Series Analysis. Princeton University Press.*

Kimberly D. (2008): *Fundamentals of National Income.* Mc Graw-Hill Inc. New York.

Li, M., Liu, G., & Zhao, Y. (2007, August). *Forecasting GDP growth using genetic programming.* In Natural Computation, 2007. ICNC 2007. Third International Conference on (Vol. 4, pp. 393-397). IEEE.

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in *time series models. Biometrika,* 65(2), 297-303.

Lu, Y. (2009). *Modelling and forecasting China's GDP data with time series models.* D- level Essay in Statistics Department of Economics and Society, Högskolan Dalarna, level Essay in Statistics Department of Economics and Society, Högskolan Dalarna, Sweden.

Paul J. (2012): *Time Series Models.* John Willey and Sons, New York.

Peter Brockwell and Richard Davis (2002). *Introduction to Time Series and Forecasting (2nd. Ed. ed.). Springer.*

Ruffin R.J. (1998): *Principles of Macroeconomics.* University of ChicagoPress,

S. E. Said and D. A. Dickey (1984). *Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order. Biometrika* 71.

Teräsvirta, T. (1994) *Specification, estimation, and evaluation of* *smooth transition autoregressive models.* Journal of the American Statistical Association.89, p 208-218.

T, Y. (2009). *Modelling and forecasting Australia's GDP data with time series models.* D- level Essay in Statistics Department of Economics and Society. 26-30.

Tse, R.Y.C. (1997). An application of the ARIMA model to real-estate prices in Hong Kong, Journal of Property Finance 8(2):152-163.

Ush, C. and R. dondro (2006) *Agumented Dickey-Fuller Type of Tests against Nonlinear Dynamic Models.* Oxford Bulletin of Economics and Statistics, 68, p. 835-866.

Watseen (2000). *residual autocorrelations in autoregressive-integrated moving average time series models.* Journal of the American statistical Association, 81(41), 117-119.