Journal of Multidisciplinary Engineering Science Studies (JMESS)
ISSN: 2458-925X
Vol. 2 Issue 11, November - 2016

A Box-Jenkins Model for Monthly Natural Gas Production in Nigeria

Ette Harrison Etuk
Department of Mathematics
Rivers State University of Science and Technology
Port Harcourt, Nigeria
ettetuk@yahoo.com,
ettehetuk@gmail.com
etuk.ette@ust.edu.ng

Alapuye Gbolu Eleki
Department of Statistics
Port Harcourt Polytechnic
Port Harcourt, Nigeria

Pius Sibeate
Rivers State Ministry of Education
Port Harcourt, Nigeria

***Abstract*— This is a study of monthly natural gas production in Nigeria as a time series. The realization sampled and analyzed spans from January 2008 to December 2015. Knowledge of the trend of the production and a model which adequately accounts for variability in the series could be helpful for planning and administrative purposes. The time plot shows a generally increasing trend. It is observed that the series is also seasonal of period 12 months. A non-seasonal differencing of the series is stationary and a first order moving average model is fitted to it. That is, fitted to the original data is the Box-Jenkins model, ARIMA(0,1,1). Taking advantage of the seasonality of the series, a seasonal differencing is done on the differenced series and a SARIMA(0,1,1)x(0,1,1)$_{12}$ model is fitted to the original series. This latter model is found to outdo the former in adequacy on all counts. In addition, using the January to May 2016 data, with this model, there is out-of-sample observation/forecast goodness-of-fit. Therefore the latter model is recommended as a basis for the forecasting or simulation of the series.**

Keywords—natural gas production; Box-Jenkins modelling; ARIMA modelling; SARIMA modelling; Nigeria

I. INTRODUCTION

Natural gas is a naturally occurring hydrocarbon gas comprising mostly methane and, in lesser proportions, the heavier hydrocarbons. Nigeria is ranked seventh in the world and first in Africa in natural gas production. The natural gas by-product of petroleum drilling called *associated gas* (AG) is flared due to inadequate infrastructure for its monetization. It is being speculated that Nigerian natural gas reserve is thrice as large as her crude oil reserve. Plans are being put in place to bring gas flaring to the barest minimum. By natural gas it is meant a combination of AG and non-associated gas. The largest national initiative in gas production is the National Liquefied Natural Gas (NLNG) project manned by some foreign oil companies and the Nigerian National Petroleum Company (NNPC).

Empirical study of natural gas production and utilization in Nigeria has engaged the attention of many researchers in recent times. For instance Audu [1] has shown empirically that the utilization of natural gas in Nigeria has positive impact on the economy. Gabriel *et al.* [2] also observed a positive impact of the utilization on the economy given a three-year time lag. Kareem *et al.* [3] demonstrated that gas flaring has a negative impact whereas gas production has a positive impact on the economy. Using multiple regression Diugwu *et al.* [4] showed that gas utilization impacted positively while gas production and flaring impacted negatively on the economy.

There is perhaps no time series analysis of the production of gas in Nigeria. This work involves a time series analysis of the monthly production. The methodology adopted is the Box-Jenkins approach which is an application of autoregressive integrated moving average (ARIMA) modeling, of which seasonal autoregressive integrated moving average (SARIMA) modeling is a special case.

Box-Jenkins modeling since its introduction in the 1970's has been extensively applied in time series modeling. For example,Yu *et al.* [5] fitted an ARIMA(2,2,1) to incidence of HIV infections in Korea. Lin *et al.* [6] modeled injury mortality in Xiamen, China as an ARIMA(0,1,1).

In the sequel an inspection shows that the realization of the series being analyzed herein shows annual seasonality. This has warranted the application of SARIMA modeling, which has been widely applied and successfully too, to many real-life time series. For instance, Kim *et al.* [7] demonstrated the comparative advantage of the use of SARIMA models over some other kinds of models in the description of the catch of anchovies in the Korean South Sea. Borhan and Arsad [8] described tourism to Malaysia from the US, Japan and South Korea by the use of SARIMA models. This is just to mention only a few.

II. MATERIALS AND METHODS

A. Data

The data for this work are monthly total natural gas production in billions of standard cubic feet (BSCF) from January 2008 to May 2016 from the NNPC website www.nnpcgroup.com/nnpcbusiness/businessinformation/oilgasinnigeri.aspx. The data are from the **Natural Gas Production and Utilization** subheading of **the MPI Figures** of the **Oil & Gas in Nigeria** section. The 2008 to 2015 data shall be analyzed and the five 2016 values shall be reserved for out-of-sample observation/forecast comparison.

B. Box-Jenkins Time Series Models

Box and Jenkins [9] defined an autoregressive moving average model of order p and q, denoted by ARMA(p, q) as

$$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-p} - \ldots - \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \ldots + \beta_q \varepsilon_{t-q} \quad (1)$$

where X_t is the value of a stationary time series $\{X_t\}$ at time t. Clearly p and q are integers and $\{\varepsilon_t\}$ is a white noise process. The α's and β's are constants such that model (1) is both stationary and invertible. If q=0 then (1) is an autoregressive model of order p, denoted by AR(p). If p=0 the model is a moving average model of order q, denoted by MA(q). Model (1) may be put as

$$A(L)X_t = B(L)\varepsilon_t$$

where $A(L) = 1 - \alpha_1`L - \alpha_2 L^2 - \ldots - \alpha_p L^p$ and $B(L) = 1 + \beta_1 L + \beta_2 L^2 + \ldots + \beta_q L^q$ and $L^k X_t = X_{t-k}$. A(L) and B(L) are respectively the autoregressive (AR) operator and the moving average (MA) operator.

If the time series $\{X_t\}$ is non-stationary, Box and Jenkins [9] proposed that differencing of the series to an appropriate degree might make it stationary. Suppose d is the least positive integer such that the d^{th} difference of $\{X_t\}$ denoted by $\{\nabla^d X_t\}$ is stationary where ∇=1-L. $\{X_t\}$ is said to be I(d), and replacement of X_t by its d^{th} difference in (1) yields an *autoregressive integrated moving average of order p, d, q,* denoted by ARIMA(p,d,q), in the original series.

If {Xt} is in addition seasonal, Box and Jenkins [9] further proposed that, in order to capture the seasonality, it might be modeled by

$$A(L)\Phi(L^s)\nabla^d\nabla^D_s X_t = B(L)\Theta(L^s)\varepsilon_t \quad (2)$$

where s is the seasonality period, $\Phi(L) = 1+\phi_1 L+\phi_2 L^2+\ldots+\phi_P L^P$ and $\Theta(L) = 1+\theta_1 L+\theta_2 L^2+\ldots+\theta_Q L^Q$ are the seasonal AR and MA operators respectively, the ϕ's and θ's being constants such that the entire model is stationary and invertible. ∇_s^D represents the D^{th} seasonal difference operator. Model (2) is called a *seasonal autoregressive integrated moving average model of order p, d, q, P, D, Q, s* denoted by SARIMA(p, d, q)x$(P,D,Q)_s$.

Box-Jenkins modelling involves first of all the determination of the orders in (1) and (2). The AR orders p and P are determined by the non-seasonal and the seasonal cut-off lags of the autocorrelation function (ACF). Their MA counterparts q and Q may be estimated by the non-seasonal and the seasonal cut-off lags of the partial autocorrelation function (PACF). It is sufficient to choose the differencing orders d and D such that their sum is at most for stationarity to be attained. The seasonality period often suggestive naturally or by an inspection of the series. Stationarity might be ascertained by the use of Augmented Dickey Fuller (ADF) test. Alternative models might be compared using Akaike's Information criterion (AIC). The model might be estimated by the least squares criterion.

C. Statistical Package

The statistical and econometric software Eviews 7 shall be used to do all the data analysis in this work. It uses the least squares criterion for model estimation.

III. RESULTS AND DISCUSSIONS

The time plot of the 2008-2015 realization, NNGP, in Figure 1 reveals an overall upward trend. That is, as time progresses the production of natural gas is on the increase. Its ACF and PACF in Figure 2 are as expected: autocorrelations from lags 1 to 6 are statistically significant. A non-seasonal differencing appears to have rid the series of the non-stationary behavior. The differenced series DNNGP shows no trend (See Figures 3 & 4). Besides the unit root test on NNGP is non-significant amounting to a corroboration of the null hypothesis of non-stationarity whereas the same test on DNNGP is highly significant suggesting that the

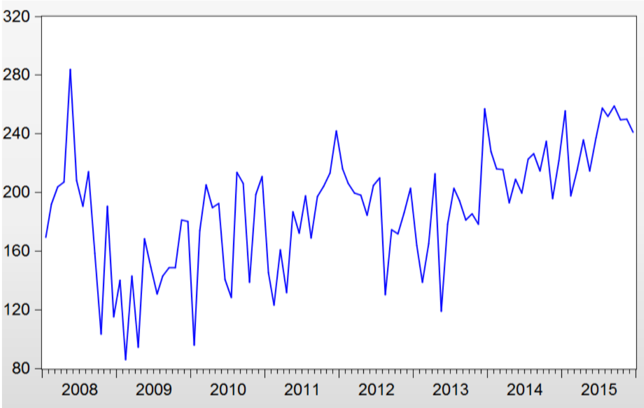


FIGURE 1: NNGP

Lag	AC	PAC	Q-Stat	Prob
1	0.538	0.538	28.615	0.000
2	0.467	0.250	50.401	0.000
3	0.502	0.270	75.909	0.000
4	0.407	0.045	92.869	0.000
5	0.319	-0.038	103.39	0.000
6	0.381	0.133	118.56	0.000
7	0.189	-0.207	122.32	0.000
8	0.152	-0.046	124.77	0.000
9	0.228	0.107	130.41	0.000
10	0.103	-0.085	131.58	0.000
11	0.065	-0.001	132.05	0.000
12	0.172	0.125	135.35	0.000
13	0.141	0.094	137.61	0.000
14	0.038	-0.114	137.77	0.000
15	0.131	0.037	139.78	0.000
16	0.077	-0.014	140.47	0.000
17	0.007	-0.092	140.48	0.000
18	0.135	0.116	142.69	0.000
19	0.116	0.055	144.35	0.000
20	0.030	-0.016	144.46	0.000
21	0.138	0.063	146.87	0.000
22	0.102	-0.027	148.19	0.000
23	0.021	-0.035	148.25	0.000
24	0.129	0.030	150.42	0.000
25	0.077	-0.067	151.20	0.000
26	-0.023	-0.065	151.28	0.000
27	0.035	-0.027	151.44	0.000
28	0.050	0.058	151.78	0.000
29	-0.133	-0.170	154.26	0.000
30	-0.068	-0.025	154.91	0.000
31	0.011	0.157	154.93	0.000
32	-0.096	-0.037	156.29	0.000
33	-0.158	-0.217	160.01	0.000
34	-0.133	-0.068	162.71	0.000
35	-0.139	0.187	165.70	0.000
36	-0.055	0.090	166.17	0.000

FIGURE 2: CORRELOGRAM OF NNGP

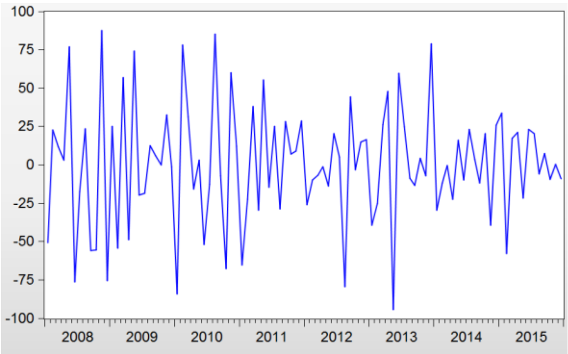


FIGURE 3: DNNGP

Lag	AC	PAC	Q-Stat	Prob
1	-0.433	-0.433	18.585	0.000
2	-0.118	-0.376	19.982	0.000
3	0.137	-0.138	21.893	0.000
4	-0.021	-0.067	21.940	0.000
5	-0.149	-0.215	24.226	0.000
6	0.287	0.150	32.841	0.000
7	-0.165	0.014	35.732	0.000
8	-0.119	-0.130	37.255	0.000
9	0.239	0.089	43.417	0.000
10	-0.107	-0.003	44.677	0.000
11	-0.162	-0.159	47.576	0.000
12	0.156	-0.138	50.291	0.000
13	0.100	0.116	51.435	0.000
14	-0.241	-0.070	58.115	0.000
15	0.179	-0.012	61.834	0.000
16	-0.005	0.063	61.836	0.000
17	-0.209	-0.130	67.030	0.000
18	0.172	-0.070	70.597	0.000
19	0.067	0.000	71.144	0.000
20	-0.207	-0.043	76.454	0.000
21	0.149	0.000	79.245	0.000
22	0.041	0.001	79.455	0.000
23	-0.208	-0.052	85.057	0.000
24	0.181	0.040	89.341	0.000
25	0.063	0.077	89.875	0.000
26	-0.181	0.023	94.297	0.000
27	0.052	-0.050	94.672	0.000
28	0.208	0.163	100.67	0.000
29	-0.260	0.029	110.19	0.000
30	-0.003	-0.150	110.20	0.000
31	0.204	0.032	116.23	0.000
32	-0.082	0.183	117.22	0.000
33	-0.061	0.026	117.78	0.000
34	0.030	-0.223	117.92	0.000
35	-0.110	-0.099	119.79	0.000
36	0.033	-0.025	119.96	0.000

FIGURE 4: CORRELOGRAM OF DNNGP

series DNNGP is stationary. That is, NNGP is I(1). Moreover the correlogram of Figure 4 suggests an MA(1) for DNNGP; that is, an ARIMA(0,1,1) for NNGP. The model as summarized in Table 1 is given by

$$DNNGP_t = -0.6847\varepsilon_{t-1} \qquad (3)$$

An inspection suggests that NNGP is seasonal of period 12 months. Yearly minimums are October, April, January, February, August, May, April and February respectively and the respective yearly maximums are May, November, August, December, January, December, October and September. It may be observed that 6 out of 8 of the minimums are in the first half of the year and 6 out of 8 of the maximums are in the second half of the year. This is an indication of a seasonal tendency of period 12 months. A 12-monthly differencing of DNNGP yields the series SDDNNGP whose time plot of Figure 5 suggests stationarity and whose ACF in Figure 6 suggests a SARIMA$(0,0,1)x(0,0,1)_{12}$ model for SDDNNGP; that is, a SARIMA$(0,1,1)x(0,1,1)_{12}$ model for NNGP.

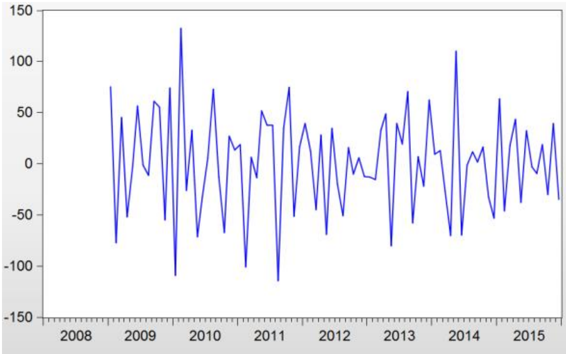


FIGURE 5: SDDNNGP

Lag	AC	PAC	Q-Stat	Prob
1	-0.463	-0.463	18.680	0.000
2	0.058	-0.199	18.980	0.000
3	-0.134	-0.257	20.581	0.000
4	0.085	-0.135	21.230	0.000
5	0.087	0.067	21.926	0.001
6	0.042	0.178	22.089	0.001
7	-0.145	0.004	24.060	0.001
8	-0.077	-0.173	24.620	0.002
9	0.222	0.106	29.348	0.001
10	-0.130	-0.041	30.999	0.001
11	0.177	0.141	34.100	0.000
12	-0.449	-0.364	54.315	0.000
13	0.247	-0.182	60.524	0.000
14	-0.008	-0.077	60.530	0.000
15	0.094	-0.008	61.447	0.000
16	-0.159	-0.078	64.146	0.000
17	-0.015	-0.055	64.171	0.000
18	0.084	0.133	64.939	0.000
19	-0.017	-0.031	64.972	0.000
20	-0.018	-0.222	65.007	0.000
21	-0.042	0.009	65.207	0.000
22	0.139	0.201	67.459	0.000
23	-0.045	0.211	67.696	0.000
24	0.031	-0.199	67.809	0.000
25	-0.005	0.095	67.812	0.000
26	-0.045	0.127	68.063	0.000
27	0.046	0.009	68.333	0.000
28	0.100	0.006	69.620	0.000
29	-0.125	-0.032	71.681	0.000
30	-0.101	-0.129	73.054	0.000
31	0.213	0.015	79.267	0.000
32	-0.072	-0.071	79.996	0.000
33	-0.005	0.020	80.000	0.000
34	-0.111	0.012	81.768	0.000
35	-0.027	-0.045	81.874	0.000
36	0.109	-0.170	83.668	0.000

FIGURE 6: CORRELOGRAM OF SDDNNGP

TABLE 1: ESTIMATION OF THE ARIMA(0,1,1) MODEL

Dependent Variable: DNNGP
Method: Least Squares
Date: 10/16/16 Time: 19:59
Sample: 2008M01 2015M12
Included observations: 96
Convergence achieved after 7 iterations
MA Backcast: 2007M12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.684738	0.074514	-9.189330	0.0000

R-squared	0.321776	Mean dependent var	0.219271
Adjusted R-squared	0.321776	S.D. dependent var	39.15240
S.E. of regression	32.24369	Akaike info criterion	9.794884
Sum squared resid	98767.31	Schwarz criterion	9.821596
Log likelihood	-469.1544	Hannan-Quinn criter.	9.805682
Durbin-Watson stat	1.977050		

Inverted MA Roots .68

TABLE 2: ESTIMATION OF THE SARIMA(0,1,1)X(0,1,1)12 MODEL

Dependent Variable: SDDNNGP
Method: Least Squares
Date: 10/23/16 Time: 16:06
Sample (adjusted): 2009M01 2015M12
Included observations: 84 after adjustments
Convergence achieved after 14 iterations
MA Backcast: 2007M12 2008M12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.713440	0.069101	-10.32463	0.0000
MA(12)	-0.921963	0.022958	-40.15795	0.0000
MA(13)	0.646179	0.069622	9.281211	0.0000

R-squared	0.652505	Mean dependent var	1.479048
Adjusted R-squared	0.643925	S.D. dependent var	49.50904
S.E. of regression	29.54305	Akaike info criterion	9.644635
Sum squared resid	70696.15	Schwarz criterion	9.731450
Log likelihood	-402.0747	Hannan-Quinn criter.	9.679534
Durbin-Watson stat	2.051247		

Inverted MA Roots	1.00	.86+.50i	.86-.50i	.70
	.50+.86i	.50-.86i	.00+.99i	.00-.99i
	-.50-.86i	-.50+.86i	-.86-.50i	-.86+.50i
	-.99			

TABLE 3: OUT-OF-SAMPLE OBSERVATION/FORECAST COMPARISON

Time	Observation	Forecast
January 2016	241.92	243.62
February2016	202.64	220.47
March 2016	222.55	240.94
April 2016	250.03	242.72
May 2016	230.05	245.99

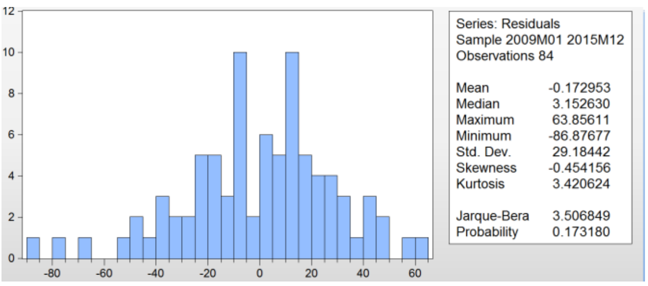
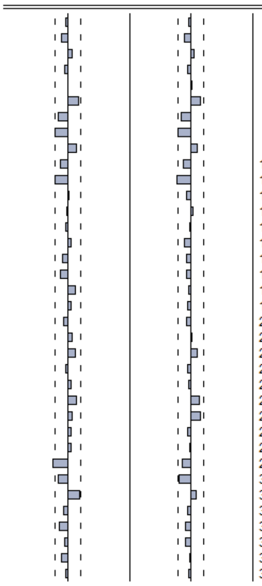


FIGURE 7: HISTOGRAM OF THE SARIMA(0,1,1)X(0,1,1)12 MODEL RESIDUALS



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.032	-0.032	0.0900	
		2	-0.104	-0.105	1.0476	
		3	0.066	0.060	1.4387	
		4	-0.055	-0.063	1.7145	0.190
		5	0.007	0.018	1.7194	0.423
		6	0.179	0.166	4.6762	0.197
		7	-0.170	-0.159	7.3896	0.117
		8	-0.229	-0.218	12.367	0.030
		9	0.150	0.104	14.548	0.024
		10	-0.116	-0.129	15.860	0.026
		11	-0.215	-0.231	20.443	0.009
		12	0.016	-0.070	20.469	0.015
		13	-0.023	0.028	20.522	0.025
		14	-0.042	-0.020	20.700	0.037
		15	0.041	-0.110	20.876	0.052
		16	-0.089	-0.074	21.712	0.060
		17	-0.137	-0.078	23.729	0.049
		18	0.124	-0.040	25.416	0.045
		19	0.057	-0.053	25.778	0.057
		20	-0.074	-0.069	26.394	0.068
		21	0.081	0.019	27.139	0.076
		22	0.130	0.102	29.108	0.064
		23	-0.042	-0.065	29.316	0.082
		24	0.054	-0.045	29.668	0.099
		25	0.132	0.137	31.814	0.081
		26	0.071	0.171	32.449	0.091
		27	0.049	-0.049	32.749	0.109
		28	0.049	-0.021	33.055	0.130
		29	-0.249	-0.142	41.179	0.030
		30	-0.169	-0.199	45.001	0.016
		31	0.204	0.091	50.687	0.005
		32	-0.081	-0.049	51.602	0.006
		33	-0.145	-0.095	54.599	0.004
		34	-0.062	-0.095	55.160	0.005
		35	-0.101	-0.013	56.676	0.005
		36	-0.045	-0.045	56.978	0.006

The estimation of the model as summarized in Table 2 yields

$$SDDNNGP_t = -0.7134\varepsilon_{t-1} - 0.9220\varepsilon_{t-12} + 0.6462\varepsilon_{t-13} + \varepsilon_t \quad (4)$$

Clearly model (4) shows supremacy over (3). AIC is lower with (4) than with (3). Moreover, R^2 is more with (4) than with (3), R^2 being interpreted as the proportion of variability in the dependent variable explained by the model. Adequacy of model (4) is not in doubt as its residuals are normally distributed (See Figure 7) and are not correlated (See Figure 8).

Moreover in Table 3, there is very close agreement between the out-of-sample observed values of January to May 2016 and their corresponding forecasts based on model (4). The chi-square goodness-of-fit test is not significant with calculated test statistic of 4.1105 less than the tabulated 5% significance critical value of 9.488 at 4 degrees of freedom. In fact, the p-value is more than 0.3.

III. CONCLUSION

It may be concluded that monthly gas production in Nigeria follows a SARIMA(0,1,1)x(0,1,1)12 model (4). Even though Nigeria is known to be in the gas zone with huge reserves of gas its production is limited because some constraints of constraints like infrastructural inadequacies and economic depression. A model for predicting the future prospects would help in strategic planning. Forecasting or simulation of the production could therefore be done on the basis of the model (4).

REFERENCES

[1] N. P. Audu, "The dynamics of gas utilization and the performance of Nigeria's economy: an error correction analysis." Journal of Research in National Development, vol. 8(1), www.transcampus.org/JORINDV8Jun2010/JournalsV8NO11Jun201011.html

[2] O. M. Gabriel, S. D. Kareem, F. Kari, G. M. Alam and O. D. Matuin, "Impact of gas industry on sustainable economy in Nigeria: further estimations through Eview", Journal of Applied Sciences, Vol. 12, 2012, pp.2244-2251.

[3] S. D. Kareem, D. V. Enoho and O.K. Sakiru, "Gas production, gas flaring and economic growth nexus: the Nigerian experience," IOSR Journal of Humanities and Social Science, vol. 8(2), 2013, pp. 41-44.

[4] I. A. Diugwu, M. A. Ijaiya, M. Mohammed and A. E. Egila, "The effect of gas production, utilization and flaring on the economic growth in Nigeria", Natural Resource, vol. 4, 2013, pp. 341-348.

[5] H. Yu, N. Kim, S. S. Kim, C. Chu and M. Kee, "Forecasting the number of human immunodeficiency virus infections in the Korean population using autoregressive integrated moving average model," Osong Public Health Research Perspective, vol. 4(6), 2013, pp. 358-362.

[6] Y. Lin, M. Chen, G. Chen,X. Wu and T. Lin, "Application of an autoregressive integrated moving average model for predicting injury mortality in Xiamen, China", BUJ Open,2015;5:e008491.doi:10.1136/bmjopen-2015-008491.

[7] J. Y. Kim, H. C. Jeong, H. Kim and S. Kang, "Forecasting the monthly abundance of achovies in the South sea of Korea using a univariate approach,"Fisheries Research, 161, 293-302.

[8] N. Borhan and Z. Arsad, "Forecasting international tourism demand from the US, Japan and South Korea to Malaysia: a sarima approach", Proceedings of the 21st National Symposium on Mathematical Sciences (SKSM21) AIP Conf. Proc. 1605, 2014, pp. 955-960: doi:10.1063/1.4887719©2014 AIP Publishing LLC 978-0-7354-1241-5/$30

[9] G. E. P. Box and G. M. Jenkins, Time Series Analysis, Forecasting and Control, San Francisco: Holden-Day, 1976.