Journal of Physical Science and Innovation **Volume 5, No. 2, 2013**
ISSN: 2277-0119

THE FITTING OF ARIMA MODEL IN FORECASTING NIGERIA GROSS DOMESTIC PRODUCT (GDP)

[1]Amadi, Innocent Uchenna and [2]Aboko, Simon Igboye
[1]Department of Mathematics/Computer Science, Rivers State University of Science and Technology, Oroworukwo, Nigeria
[2]Department of Mathematics/Statistics, Rivers State College of Arts and Science Rumuola, Nigeria.
E-mail: amadiuchennainno@gmail.com, abokoigboye@gmail.com

ABSTRACT

This research focused on Arima modeling technique to the forecasting of Nigerian Gross Domestic Product between the period of 1980 – 2011. For statistical analysis we have used graphical methods to display data distributions, Autocorrelation Function (ACF), Partial Autocorrelation Functions (PACF), Residuals and forecast, and differencing to check for stationarity. The ARIMA (2,1,2) model was proposed for the data from the first differences which shows stability and invertibility. Forecasting were made for future observations up to fifteen (15) years which shows an increasing trend over time, and the Akaike Information Criteria (AIC) and the adjusted multiple correlation coefficient (adjusted R-square) provide a good summary of the total variability explained by the chosen fitted model.

Keywords: Gross Domestic Product, Arima Models, Forecasting, Nigeria.

INTRODUCTION

One of the main objectives of statistics is to forecast future levels of economic activities by studying the behaviour of data in the past. In government, short-term and long-term planning is carried out on the basis of scientific analysis of past data of the various economic variables of the national economy. In business and industry, management would wish to forecasts the levels of customers' demands in order to plan towards commitment of resources towards increasing supply. Making such forecasts entails setting up statistical model showing how the economy works and indicating forces which determine demand and supply in the past. In all other areas of human endeavour, we use our past experience of previous years to forecast what is likely to happen in the future.

However, the analysis of time series involves description, control and production of the underlying processes. Time series data when analysed may enhance the understanding of past and current pattern of change. This research is based in forecasting, using Box-Jenkins method, concentrating first on the theoretical analysis and then applying the method to GDP in Nigerian economic data.

Using Arima model as a statistical tool has recently been attracting attentions of many research workers as it is expected that this time domain approach will give answers to many problems. For instance, Fatoki *et al.*, (2010), Wei *et al.*, (2010), Aboko (2013), Box and Jenkins (1976) and Box, Pierce (1970). This is just to mention but few.

In time series analysis, ARIMA models are flexible and widely used. The time series model can provide short run forecast for sufficiently large amount of data on the concerned variables very precisely. The abbreviation ARIMA (p,d,q) stands for autoregressive integrated moving average with three parameters, p, the order of autoregressive, d, the degree of differencing and q, the order of moving average. The ARIMA model, commonly known as Box-Jenkins model, is due to Box and Jenkins work for forecasting of a large variety of time series data. The underlying assumption is that the time series to be forecast has been generated by a stochastic process.

Models

Models for time series data have many forms and represent different stochastic processes when model variations in the level of a process, three broad classes of practical importance are the autoregressive (AR) models, integrated (I) and moving average (MA) models. These three classes depend linearly on previous data points. Combinations of these ideas produce autoregressive moving averages (ARMA) and autoregressive integrated moving averages (ARIMA) models

MATERIALS AND METHODS

The study made use of secondary data collected between 1980 – 2011 from Central Bank of Nigeria (CBN). Retrievable from the data and statistics publication of CBN website www.cenbank.org.

Model Formulation

The Box-Jenkins methodology refers to the set of procedures for identifying, fitting and checking ARIMA model with time series data. Forecasts follow directly from the form of the fitted model Box-Jenkins consists of four steps, such as:

1. Identification
2. Estimation
3. Diagnostic checking and
4. Forecasting

Box and Jenkins (1976) formalized the ARIMA modeling framework by defining four steps to be carried out in the analysis: identification of the model, estimated the coefficients and verify the model; identification of the model (i.e. how many terms to be included) is based on the autocorrelation function (ACF) and partial auto correlated function (PACF) of the differences; log transformed time series (Box and Jenkins, 1976). Estimation of the coefficients of the model is carried out by means of the maximum likelihood method.

Verification of the model is done through diagnostic checks of the residual (histogram and normal probability plot of residuals, standardized residuals and ACF on PACF of the residuals). The performance of ARIMA model was tested through comparisons of predictions with observations not used in the fitted model. The accuracy of ARIMA forecast model was compared to the average quarterly means over the previous years by examining the variance accounted for (δ^2) by the model.

The BJ ARIMA method applies only to stationary data series. A stationary time series has a mean, variance, and auto-correlation function (ACF) that are essentially constant through time. The stationary assumption simplifies the theory underlying (BJ) models and helps to ensure that we can get useful estimates of parameters from a moderate number of observations. If a time series is stationary, then the mean and variance of any other subset of the series will be constant. A series is said to be non-stationary if it does not vary about a constant mean, that is, the series does not exhibit homogenous behaviour of a kind. A model that represents homogenous non-stationary behaviour using Box-Jenkins and Reinsel (1994) is of the form:

$$\phi(\beta)(1-\beta)^d Y_t = \theta\beta a_t \quad (1)$$

Where

Y	=	is the response variable at time t.
$\Phi\beta$	=	represent the AR proceeds operator.
$\theta\beta$	=	the MA process operator.
a_t	=	represent the white noise.
β	=	is the backward shift operator and
d	=	is the number of time the data series must differenced to induce a stationary mean.

(1) A pth –order autoregressive model: AR(p), which has the general form

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} \quad \ldots + \Phi_p Y_{t-p} + \varepsilon_t \quad (2)$$

Where;

Y_t = Response (dependent) variable at time t

$Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$ = Response variable at time lag t-1, t-2,….t-p respectively.

$\Phi_0, \Phi_1, \Phi_2, \ldots, \Phi_p$ = coefficient to be estimated

ε_t = Error term at time t.

2. A qth –order moving average model: MA(q), which has the general form.

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \varepsilon_t - \theta_2 \varepsilon_{t-2} - \ldots - \theta q\, \varepsilon_{t-q} \quad (3)$$

Y_t = Response (dependent) variable at time t.

μ = constant mean of the process

$\theta_1, \theta_2, \ldots, \theta_q$ = coefficient to be estimated

ε_t = Error term at time t.

$\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-q}$ = Error in previous time periods that are incorporated in the response Y_t.

(3) Autoregressive Moving Average mode: ARMA (p,q), which has the general form

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_q Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \ldots - \theta_q \varepsilon_{t-q} \quad (4)$$

We can use the graph of the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) to determine the model which processes can be summarized as follows:

Table 1: How to Determine the Model by Using ACF and PACF Patterns

MODEL	ACF	PACF
AR(p)	Dies down	Cutoff after lag q
MA(q)	Cut off after lag p	Dies down
ARMA(p,q)	Dies down	Dies down

The Steps in the Arima Model- Building

Step 1: Model Identification (Model) Selection and Initial
Determine whether the series is stationary or not by considering the graph of ACF. If a graph of ACF of the series value either *cut off fairly quickly or dies down fairly quickly*, then the time series values should be considered *stationary*. If a graph of ACF dies down extremely slowly, then the time series values should be considered non-stationary.

If the series is not stationary, it can often be converted to a stationary series by differencing. That is the original series is replaced by a series of differencing. An ARIMA model is then specified for the differenced series. Differencing is done until a plot of the data indicates the series caries about a fixed level, and the graph of ACF either cuts off fairly quickly or dies down fairly quickly.

The theory of time-series analysis has developed a specific language and a set of linear operators. According to equation (2), a highly useful operator in time series theory is the lag or backward linear operator (B) defined by $BY_t = Y_{t-1.}$

Model for non-seasonal series are called **Autoregressive Integrated Moving Average Model**, denote by ARIMA (p,d,q). Here p indicates the order of the autoregressive part, d indicates the amount of differencing, and q indicates the order of the moving average part, If the original series is stationary, d =0 and the ARIMA models reduce to the ARMA models.
The difference linear operator (Δ), defined by

$$\Delta Y_t = Y_{t-1} = Y_t - BY_t = (1-B)Y_t \tag{5}$$

The stationary series W_t obtained as the dth difference (Δ^d) of Y_t,

$$W_t = \Delta^d Y_t = (1-B)^d Y_t \tag{6}$$

ARIMA (p,d,q) has the general form

$$\begin{aligned} &\phi_p(B)(1-B)^d Y_t = \mu + \theta_q\ (B)\varepsilon_t \\ &or\ \ \phi_p(B)W_t = \mu + \theta q\ (B)\varepsilon_t \end{aligned} \tag{7}$$

Once a stationary series has been obtained; then the identify form of the model to be used by using the theory in TABLE 1.

Step 2: Model Estimation
Here, the selection models are estimated and also computed for the value of AIC and BIC. According to the smallest value of AIC and BIC, ARIMA model are secondarily selected and it is done with the help of computer package.

Having tentatively selected a model, the next step is the estimate of the parameters of the model. In this work the maximum likelihood approach which has been proved to reflect all useful information about the parameters contained in the data is used.

Step 3: Model Checking Diagnostic
In this step, model must be checked for adequacy by considering the properties of the residuals whether the residuals from an ARIMA model must have the normal distribution or

should be random. An overall check of model adequacy is provided by the Ljuan-Box Q statistic. The test statistic Q is

$$Q_m = n(n+2)\sum_{k-1}^{m}\frac{r^2k(e)}{n-k} \sim X^2_{m-r;} \qquad (8)$$

Where

$r_k(e)$ - residual autocorrelation at lag k
n = the number of residuals
m = the number of time lags includes in the test.

If the p -value associated with the Q statistics is small (P-value < α), the model is considered inadequate. The analyst should consider a new or modified model and continue the analysis until a satisfactory model has been determined.

Moreover, we can check the properties of the residual with the following graph.

1. We can check about the normality by considering the normal probability plot or the p -value from the one-sample Kolmogorov Srnirov Test.

1. We can check about the randomness of the residuals by considering the graph of ACF and PACF of the residuals. The individual residual autocorrelation should be small and generally written ±2/Vn of zero.

Step 4: Forecasting with the Model

Forecasting for one period or several periods into the future with the parameters for a tentative model, we use E-view package for regression and forecasting. To choose one, two or three best approximations we use criteria R^2 an Akaike information criteria (AIC), which is based on the sum of squared residuals. The best specification is choosing with the lowest value of the A1C: the criteria are calculated as;

$$AIC = \log\frac{(RSS)}{T} + \frac{2k}{T}$$

Where RSS is sum of square residuals and k is the number of estimated parameters. E-views calculates both criteria R^2 and AIC. Therefore it is easy to use selection criteria.

In seasonal adjustment of GDP it was possible to do additional forecasting using ARIMA models. Two methods for forecasting were used: direct model for GDP and indirect model for component of GDP by institutional sectors. There forecast can be obtained directly from ARIMA (p,d,q) (P,D,Q). Model can be written as:

$$(1-\varphi_1 B - \varphi_2 B^2 - \varphi_p B^p)(1-\varphi_s B^s - \varphi_{2s} B^{2s} - \ldots - \varphi^{ps} B^{ps})(1-B)^d(1-B^s)^D Y_t$$

$$= \delta + (1-\theta_1 B - \theta_2 B^2 - \theta_q B^9)(1-\theta_s B^s - \theta_{2s} - B^{2s} - \ldots - \theta_{Qs} B_{Qs})\mu_t \qquad (9)$$

Where;

B = is the backshift operator
S = 4 for quarterly data

P and p are orders of autoregressive and seasonal autoregressive part
q and Q are orders of moving averages and seasonal moving average part
d and D are orders of difference and seasonal difference

For choosing the best forecast we made trails using four econometric criteria: RMSE (Root Mean Square Error), MAE (Mean Absolute Error) MAPE (Means Absolute Percent Error) and TIC (Theil Inequality Coefficient), These criteria are calculated using (E-view package) by using the following formula

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=T=1}^{T=h}(Y_t - Y_t)^2} \quad (10)$$

$$MAE = \frac{1}{h}\sum_{t=T=1}^{T=h} / Y_t - Y_t / \quad (11)$$

$$MAPE = \frac{1}{h}\sum_{t=T=1}^{T=h}[(Y_t - Y_t)/Y / \quad (12)$$

$$TIC = \frac{\sqrt{\sum_{t=T=1}^{T=h}(Y_t - Y_t)^2}}{\sqrt{\frac{1}{h}\sum_{t=T=1}^{T=h}Y_t^2} + \sqrt{\frac{1}{h}\sum_{t=T=1}^{T=h}y_t^2}} \quad (13)$$

The first two forecast error statistics depend on the scale of the dependent variable. These should be used relative measures to compare forecasts for the same series across different models. The smaller the error, the better the forecasting ability of that model according to that criteria. The other two statistics are scale invariant. The Theil inequality coefficient always lies between zero and one, while zero indicates perfect fit.

RESULTS AND DISCUSSION

A visual inspection of the time plot in figure 1 reveals that the time series is non-stationary. The sample autocorrelation functions for the sample period were computed to check whether the GDP series is stationary. As shown in figure 2, the ACF's are decreasing very slowly, indicating that the series is non stationary.

The first differences in figure 3 are stationary in mean and the data fluctuates around the constant mean, indicating that the series is stationary. The sample autocorrelation function in figure 6 shows that the population ACF (r^2) is significantly difference from zero and cults off after leg 2. Hence, the model is the second order moving average type or MA(2). Figure 7 shows the partial autocorrelation (PACF's) of the first difference from zero. The first differenced series is also of the second order autoregressive type or AR(2). The model identified as ARIMA (2,1,2). The model estimation**,** becomes

$$Y_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} = \varepsilon_t + \theta_1 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \quad (14)$$

The equation above can be written as

$(1-\theta_1 B- \theta_2 B^2)\, Y_t = (1- \theta_1 B - \theta_2 B^2)\varepsilon_t$

$(1-0.208939B-0.51852B^2)Y_t = (1-1.455189B – 0.81831B^2)\, \varepsilon_t$

The identification process having led to a tentative formulation for GDP model, we then obtain efficient estimate of the parameter μ, Φ_1, Φ_2, θ_1, θ_2, using JMP software packages, which employs the maximum likelihood approach involving nonlinear iterative techniques before the parameters are μ = 1032.23, Φ_1 = -0.208939, Φ_2 = -0.51852 and θ_1 = 1.455189, θ_2 = -0.81831.

Twenty three (23) iterations where performed before obtaining these estimate μ, Φ_1, Φ_2,. Figure 9 shows the plots of residuals. It is observed that the randomness of the residue plots man be considered as a signed or indication that there is no model lack of fit. It also shows that the data (Y_t) agrees with the fitted model.

The ARIMA (2,1,2) model is used to make forecast for future observations that the optimal in a minimum mean square error sense. The forecast in figure 10 shows that the width of the prediction limits increase then the lead time also increase.

Furthermore, for assessing the adequacy of the fitted model, the Akaike information criteria (AIC) and the adjusted multiple correlation coefficient. Adjusted R-Square provides a good summary of total variability explained by the fitted model equation. The results one 2221.83 and 0.808 respectively.

CONCLUSION

The ARIMA (2,1,2) model is adequate for the data. The model is effective and reliable.

REFERENCES

Aboko, I.S (2013). An Application of ARIMA Model in Forecasting Gross Domestic Product (GDP) in Nigeria. An Unpublished M.Sc. Dissertation, Department of Mathematics and Computer Science, Rivers State Science and Technology Port Harcourt. Nigeria.

Akaike, H. (1969). Fitting Autoregressive Models for Prediction. *An. Inst. Statist. Math*; 21, 243-247.

Beamish, N. & Priestly, M.B. (1981). A Study of Autoregressive and Window Spectral Estimation. *Applied Statistics*, 30, 41-58.

Bhuiyam, M.N.A., Kazi, S.A. & Roushan, J. (2007). *Modelling and Forecasting of the GDP of Manufacturing Industries in Bangladesh*, Jahangir Agar University Press, Dhaka, Bangladesh.

Box, G.E.P. & Jenkins, (1976). *Time Series Analysis: Forecasting by Control*, Rev. Ed., Oakums, California: Holden Day.

Box, G.E.P. & Jenkins, G.M. (1970). *Time Series, Forecasting and Control*, San, Francisco: Holden-day (Revised Edition Published (1976)).

Chartfield, C. (1985). *The Analysis of Time Series: An introduction* (3rd ed.), Chapman & Hall, New York.

E.H; Etuk, (2013). The Fitting of a Sarima Model to Monthly Naira-Euro Exchange Rates. *Mathematical Theory and Modeling, Vol. 3. No. 1.*

Fatoki, O., Mbata, U.A., Olalude, G.A, and Abass, O., *Journal of Statistics & Systems* Vol. 5 Issue 1, P63.

Wei Ning, Bian Kuan Jiang and Tuan Zhifa (2010) Analysis and Forest of Shaanxi GDP Based on the Arima Model. *Asian Agricultural Research* Vol. 2, Issue 01.

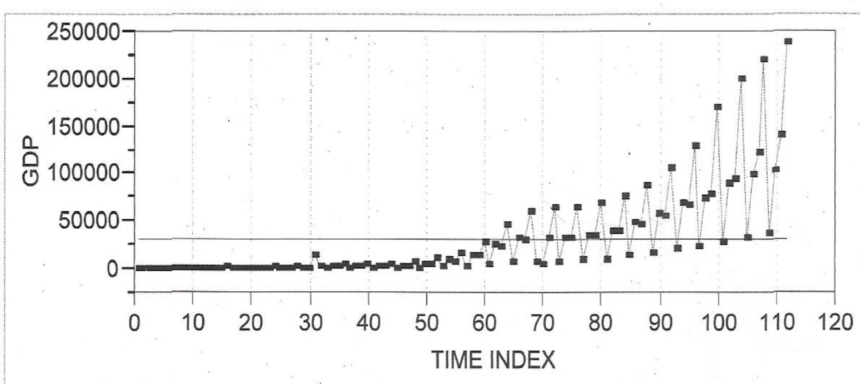
Model Comparison

Model	DF	Variance	AIC	SBC	RSquare	-2LogLH
AR(1)	110	1.57411e9	2375.8189	2381.2559	0.308	2370.2551
MA(1)	110	1.93306e9	2398.8255	2404.2625	0.15	2392.8982
I(1)	110	1.85463e9	2370.8458	2373.5553	0.181	2890.5991
ARI(1, 1)	109	1.24007e9	2328.1659	2333.585	0.458	2322.5738
IMA(1, 1)	109	909239126	2293.7212	2299.1402	0.602	2288.9871
ARIMA(1, 1, 1)	108	832888865	2285.9856	2294.1142	0.639	2278.5699
ARI(2, 1)	108	1.20947e9	2327.3926	2335.5212	0.476	2318.8487
IMA(1, 2)	108	521558418	2234.0288	2242.1574	0.774	2229.6133
ARIMA(2, 1, 2)	106	450763496	2221.8363	2235.384	0.808	2211.8855
I(2)	109	5.78187e9	2474.5793	2477.2798	-1.54	2988.6276
ARI(1, 2)	108	2.47249e9	2383.1342	2388.5351	-0.08	2377.9586
IMA(2, 1)	108	1.88111e9	2353.0642	2358.4651	0.173	2351.7553
ARIMA(2, 2, 2)	105	1.24315e9	2313.5007	2327.0031	0.47	2304.4878
ARI(2, 2)	107	2.45918e9	2384.5403	2392.6417	-0.06	2376.3264
IMA(2, 2)	107	2.09316e9	2366.8137	2374.9152	0.098	2359.4174
AR(2)	109	1.58855e9	2378.8418	2386.9973	0.308	2370.2551
MA(2)	109	1.03928e9	2331.3214	2339.4769	0.545	2326.2988

Model Comparison

Model	DF	Variance	AIC	SBC	RSquare	-2LogLH
ARIMA(1, 1, 2)	107	526342905	2237.0424	2247.8806	0.774	2229.6021
ARIMA(2, 1, 1)	107	791883900	2282.3817	2293.2198	0.66	2272.2881
ARIMA(1, 2, 1)	107	1.25033e9	2310.1337	2318.2351	0.455	2307.1434
ARIMA(1, 2, 1)	107	1.25033e9	2310.1337	2318.2351	0.455	2307.1434
ARIMA(2, 2, 1)	106	1.21703e9	2309.1651	2319.967	0.475	2303.5499
ARIMA(1, 2, 2)	106	1.2709e+9	2313.9292	2324.7312	0.452	2307.7049

Time Series GDP By TIME INDEX



Mean	31634.41
Std	47260.724
N	112

Figure 1: Original Data Time Series Plot of (GDP)

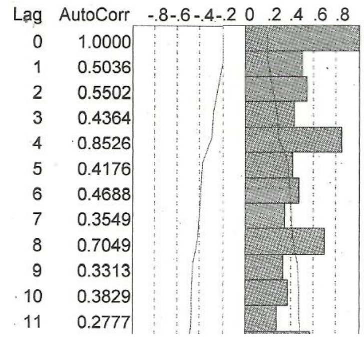
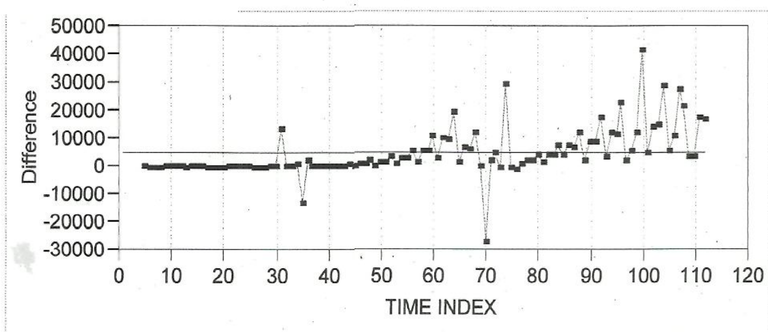


Figure 2: Autocorrelation Function of Original Data Lime Series Plot of (GDP)



Mean	4775.1815
Std	8389.0061
N	108

Figure 3: Plot first difference: (1-B)^1

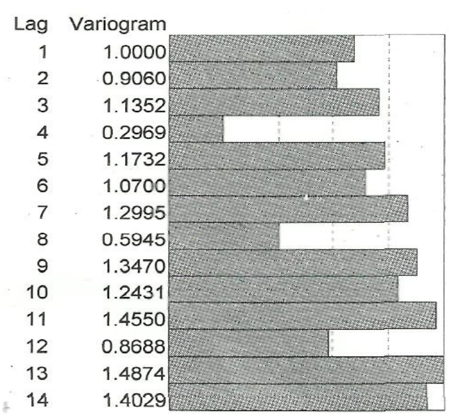


Figure 4: Variogram of Original Data Time Series Plot of (GDP)

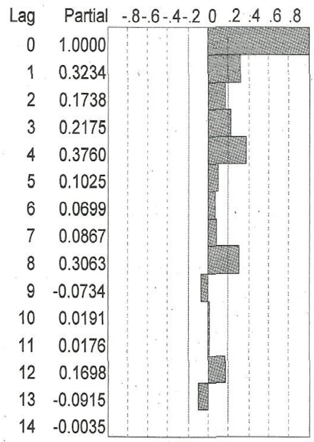


Figure 5: Partial Autocorrelation Function of Original Data Time Series Plot (GDP)

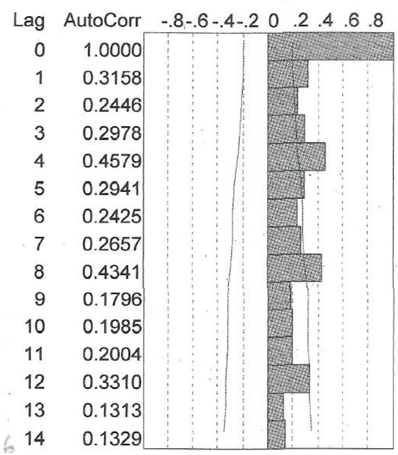


Figure 6: Autocorrelation Function of the First Differences

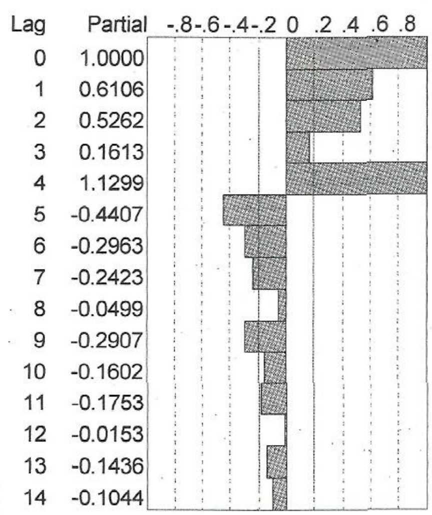


Figure 7: Partial Autocorrelation Function of the First Differences

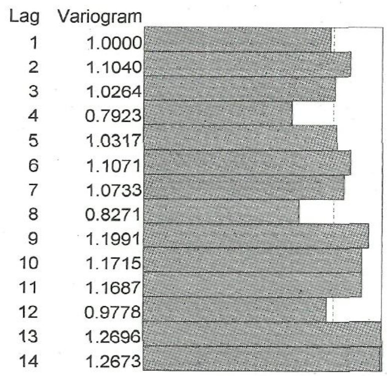


Figure 8: Variogram of the First Differences

Difference: (1-B^4)^1

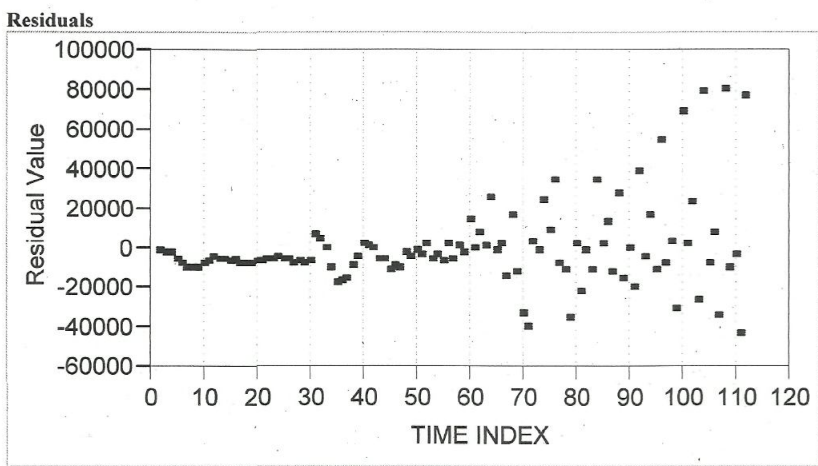


Figure 9: Plot of the Residual

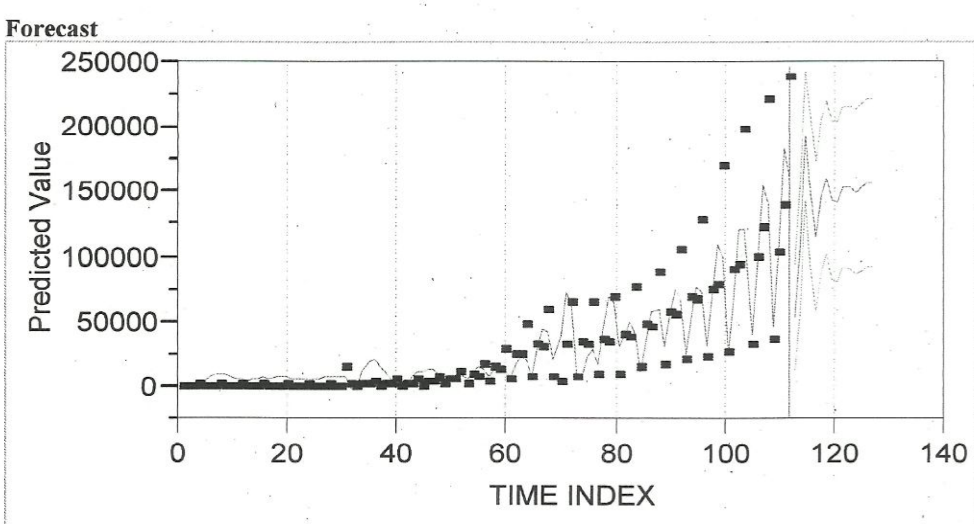


Figure 10: Forecast

Table 2: Model Summary

Model	DF	Variance	AIC	SBC	RSquare	-2LogLH
ARIMA(2, 1, 2)	106	450763496	2221.8363	2235.384	0.808	2211.8855

Model: ARIMA(2, 1, 2)

Model Summary

DF	106
Sum of Squared Errors	4.77809e10
Variance Estimate	450763496
Standard Deviation	21231.1916
Akaike's 'A' Information Criterion	2221.83632
Schwarz's Bayesian Criterion	2235.38397
RSquare	0.80798812
RSquare Adj	0.80074239
-2LogLikelihood	2211.88552
Stable	Yes
Invertible	Yes

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|
AR1	1	-0.208939	0.0990887	-2.11	0.0373
AR2	2	-0.5185254	0.1025485	-5.06	<.0001
MA1	1	1.45518952	0.0710789	20.47	<.0001
MA2	2	-0.8183135	0.0470977	-17.37	<.0001
Intercept	0	1234.21764	415.04184	2.97	0.0036
Constant Estimate		2132.06709			

Reference to this paper should be made as follows: Amadi, Innocent Uchenna and Aboko, Simon Igboye (2013), The Fitting of Arima Model in Forecasting Nigeria Gross Domestic Product (GDP). *J. of Physical Science and Innovation,* Vol. 5, No. 2, Pp. 79 – 90.

Biographical Note: Innocent Uchenna Amadi was born on July 8th, 1975. He had his HND in Computer Science from Rivers State Polytechnic, Bori, Nigeria in 2001, PGD in Statistics from Imo State University Owerri, Nigeria in 2008, M.Sc. in Applied Statistics from Rivers State University of Science and Technology, Nigeria in 2013. He is a member of Science Teachers Association of Nigeria (STAN). His research interests are in Time Series Analysis, Multivariate Methods and Stochastic Processes.

Biographical Note: Simon Igboye Aboko was born on January 6th, 1970. He had his B.Sc. in Mathematics, M.Sc. in Applied Statistics from Rivers State University of Science of Technology, Nigeria in 1998 & 2013 respectively. He is a member of Nigerian Statistical Association (NSA). His research interests are in Time Series Analysis, Demographic Measures Sampling and Survey Methods.