

Time Series Analysis Model for Production and Utilization of Gas (A Case Study of Nigeria National Petroleum Corporation “Nnpcc”)

H. R. Bakari¹, H. A. Chamalwa¹ and A. D. Mohammed²

¹(University of Maiduguri, Department of Mathematics and Statistics, Nigeria)

²(Federal Polytechnic Damaturu, Department of Mathematics and Statistics, Nigeria)

Abstract: *Time series analysis and forecasting has become a major tool in different applications for the private sector. Among the most effective approaches for analyzing time series data is the model introduced by Box and Jenkins, ARIMA (Autoregressive Integrated Moving Average). In this study we used Box-Jenkins methodology to build ARIMA model for annual production and utilization of gas from Nigeria National Petroleum Company (N. P. C.) for the period from 1970-2004. After the model specification; the best model for production was ARIMA (1, 1, 1) and for utilization was ARIMA (0, 1, 1). These models were used to forecasting the production and utilization of gas for the upcoming 4 years to help decision makers establish priorities in terms of gas demand management. An intervention time series analysis could be used to forecast the peak values of production and utilization data.*

Key words: *ARIMA, Box-Jenkins, forecasting, production and utilization model, Time series analysis,*

I. Introduction

Development and use of statistical and other quantitative methods in the environmental sciences have been a major communication between environmental scientists and statisticians [1]. This approach is called top-down approach which starts with statistical analysis of collected air pollution data [2]. Many methods and approaches for formulating forecasting models are available in the literature. This research exclusively deals with time series forecasting model, in particular, the Auto Regressive Integrated Moving Average (ARIMA). These models were described by Box and Jenkins [3] and further discussed in some other resources [4] such as: The Box-Jenkins approach which possesses many appealing features. It allows the manager who has only data on past years' quantities, rainfall as an example, to forecast future ones without having to search for other related time series data, for example temperature. Box- Jenkins approach also allows for the use of several time series, for example temperature, to explain the behavior of another series, for example rainfall, if these other time series data are correlated with a variable of interest and if there appears to be some cause for this correlation; Box-Jenkins (ARIMA) modeling has been successfully applied in various water and environmental management applications. The followings are examples where time series analysis and forecasting are effective: A manager of a private company would need forecast of an hourly volume and type of gas produced and utilized in order to schedule staff and equipment efficiently. Forecasting can also be an important part of a process control system through monitoring key processes. It may be possible to determine the optimal time and extent of control action; for example, a chemical processing unit may become less efficient as hours of continuous operation increase. Forecasting the performance of the unit will be useful in planning the shutdown time and overhaul schedule.

II. Materials And Methods

The main stages in setting up a forecasting ARIMA model includes model identification, model parameters estimation and diagnostic checking for the identified model appropriateness for modeling and forecasting. Model Identification is the first step of this process. The data was examined to check for the most appropriate class of ARIMA processes through selecting the order of the consecutive and seasonal differencing required to make series stationary, as well as specifying the order of the regular and seasonal auto regressive and moving average polynomials necessary to adequately represent the time series model. The Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are the most important elements of time series analysis and forecasting. The ACF measures the amount of linear dependence between observations in a time series that are separated by a lag k . The PACF plot helps to determine how many auto regressive terms are necessary to reveal one or more of the following characteristics: time lags where high correlations appear, seasonality of the series, trend either in the mean level or in the variance of the series. The general model introduced by Box and Jenkins includes autoregressive and moving average parameters as well as differencing in the formulation of the model. The three types of parameters in the model are: the autoregressive parameters

(p), the number of differencing passes (d) and moving average parameters (q). Box-Jenkins model are summarized as ARIMA (p, d, q). For example, a model described as ARIMA (1,1,1) means that this contains 1 autoregressive (p) parameter and 1 moving average (q) parameter for the time series data after it was differenced once to attain stationary. In addition to the non-seasonal ARIMA (p, d, q) model, introduced above, we could identify seasonal ARIMA (P, D, Q) parameters for our data. These parameters are: Seasonal autoregressive (P), seasonal Differencing (D) and seasonal moving average (Q). For example, ARIMA (1,1,1)(1,1,1) 12 describes a model that includes 1 autoregressive parameter, 1 moving average parameter, 1 seasonal autoregressive parameter and 1 seasonal moving average parameter.

These parameters were computed after the series was differenced once at lag 1 and differenced once at lag 12. The general form of the above model describing the current value X_t of a time series by its own past is:

$$(1-\phi_1B)(1-\alpha_1 B^{12})(1-B)(1-B^{12}) X_t = (1-\theta_1B)(1-\gamma_1B^{12})e_t \quad (1)$$

Where:

$1-\phi_1B$ = Non seasonal autoregressive of order 1

$1-\alpha_1 B^{12}$ = Seasonal autoregressive of order 1

X_t = The current value of the time series examined

B = The backward shift operator $BX_t = X_{t-1}$ and $B_{12}X_t = X_{t-12}$

$1-B$ = 1st order nonseasonal difference

$1-B^{12}$ = Seasonal difference of order 1

$1-\theta_1B$ = Non seasonal moving average of order 1

$1-\gamma_1B^{12}$ = Seasonal moving average of order 1

This model can be multiplied out and used for forecasting after the model parameters were estimated, as we discussed below. After choosing the most appropriate model (step 1 above) the model parameters are estimated (step 2) by using the least square method. In this step, values of the parameters are chosen to make the Sum of the Squared Residuals (SSR) between the real data and the estimated values as small as possible. In general, nonlinear estimation method is used to estimate the above identified parameters to maximize the likelihood (probability) of the observed series given the parameter values. The methodology uses the following criteria in parameter estimation:

- The estimation procedure stops when the change in all parameters estimate between iterations reaches a minimal change of 0.001
- The parameters estimation procedure stops when the SSR between iterations reaches a minimal change of 0.0001

In diagnose checking step (step three), the residuals from the fitted model shall be examined against adequacy. This is usually done by correlation analysis through the residual ACF plots and the goodness-of-fit test by means of Chi-square statistics (χ^2). If the residuals are correlated, then the model should be refined as in step one above. Otherwise, the autocorrelations are white noise and the model is adequate to represent our time series. After the application of the previous procedure for a given time series, a calibrated model will be developed which has enclosed the basic statistical properties of the time series into its parameters (step four). For example, the developed model, as shown in (1) above can be multiplied out and the general model is written in terms of X_t .

III. Results And Discussion

Since the data is a monthly production and utilization, Fig. 1, shows that there is a seasonal cycle of the series and the series is not stationary. The ACF and PACF of the original data, as shown in Fig. 2a and 2b, shows that the production and utilization data is not stationary. In order to fit an ARIMA model stationary data in both variance and mean are needed. Stationarity in the variance could be attained by having log transformation and differencing of the original data to attain stationary in the mean. For our data, we need to have seasonal first difference, $d = 1$, of the original data in order to have stationary series. After that, we need to test the ACF and PACF for the differenced series to check stationary.

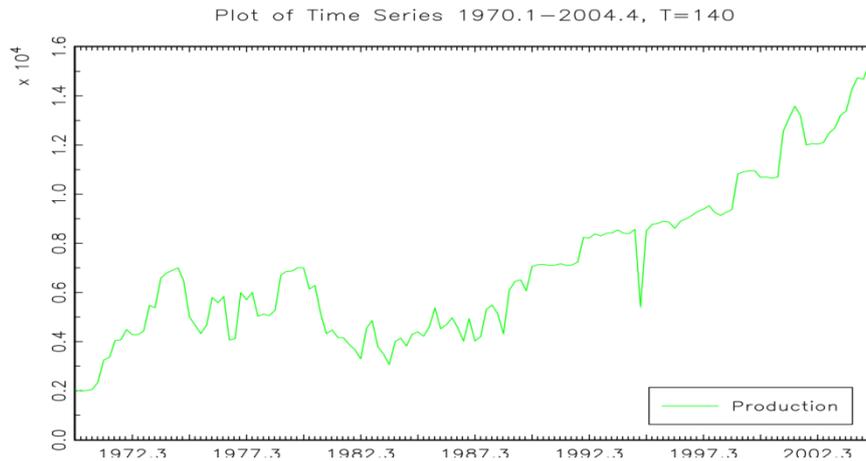


Figure 1a: Time plot for production before differencing

The plot shows the variability of the series appears to be changing with time. Therefore the mean and variance are not constant, suggesting that the series is not stationary

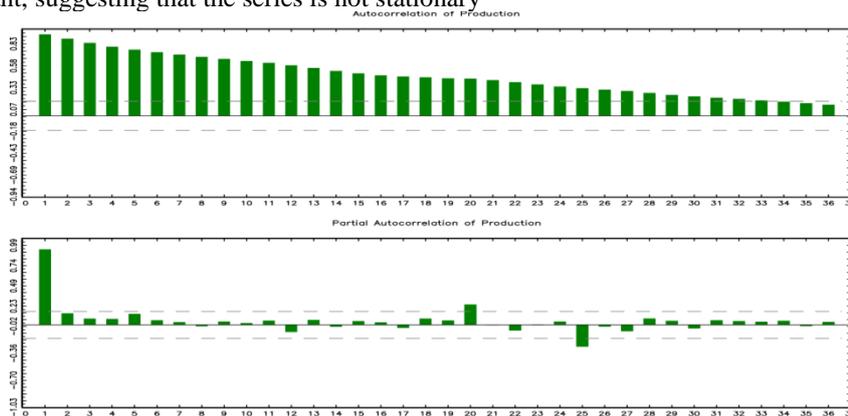


Figure 2a: The Autocorrelation function (ACF) and partial Autocorrelation function (PACF) before differencing

The Autocorrelation function (ACF) before differencing shows that it has no cyclical pattern but observed that the highest spike is in lag 1. While the partial Autocorrelation function (PACF) shows that it has a cyclical pattern and cut off after lag 1 but the highest spike is in lag 1, after lag 20 everything tends to zero

Plot of Time Series 1970.2–2004.4, T=139

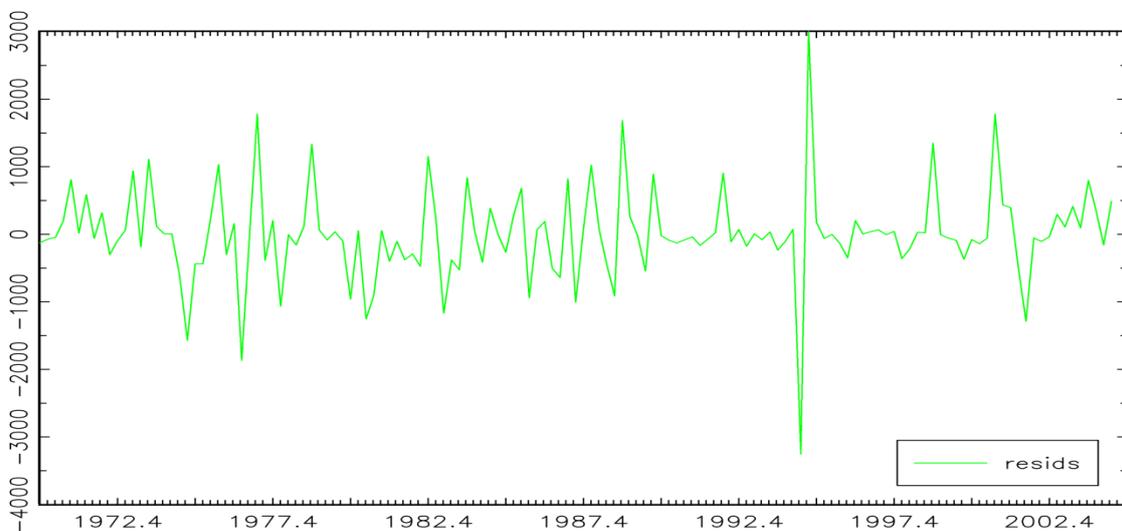


Figure 3a: Time plot for quarterly production after first differencing shows

The mean and variance appeared to be constant. Therefore, the series is stationary since the mean and variance are constant

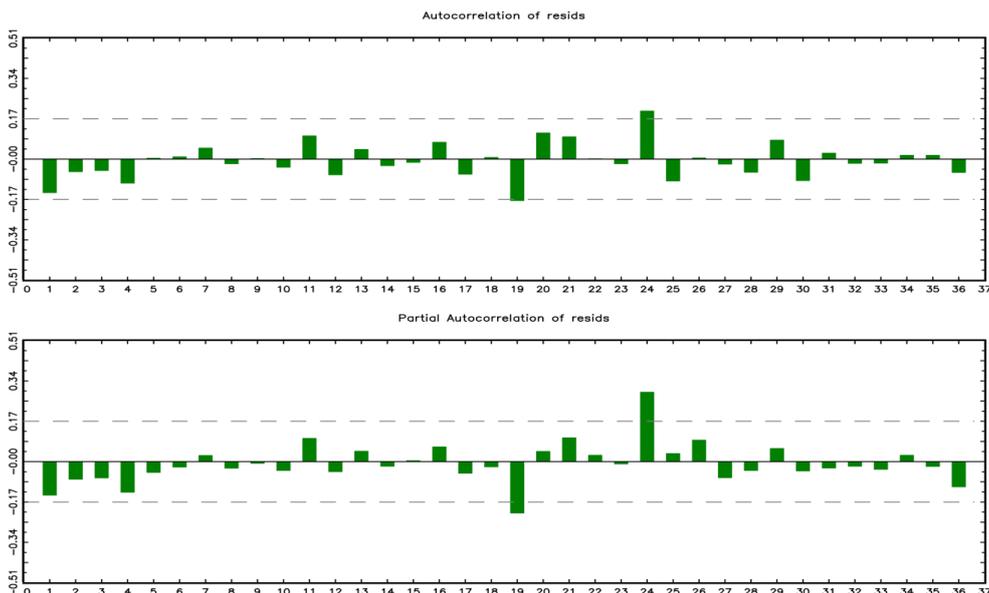
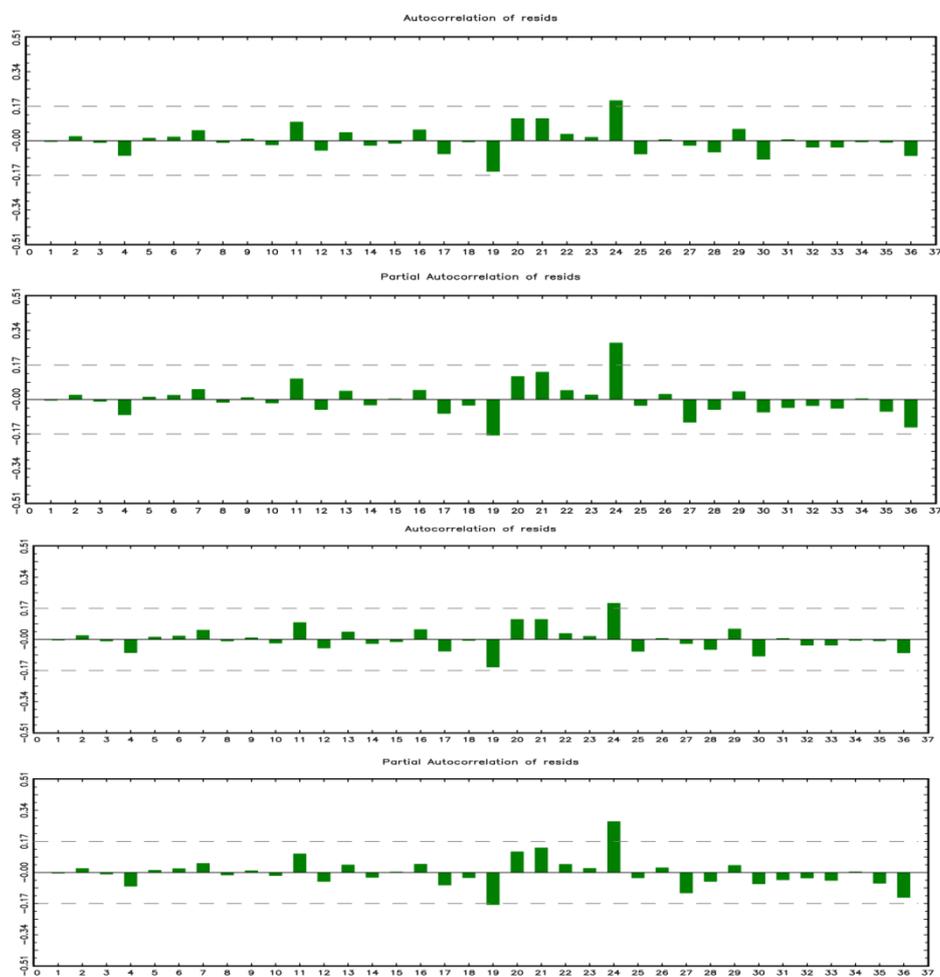


Figure 4a: The Autocorrelation function (ACF) and Partial autocorrelation function (PACF) after first differencing

ACF shows there is no correlation function after the first differencing. Therefore the series is stationary. The PACF shows that there is no correlation in the series after first difference. However the series indicate the presence of AR and MA process models



The correlation shows the following

All spikes for both ACF and PACF are within the confidence interval (upper and Lower), indicating that the selected model (ARIMA) can fit the data.

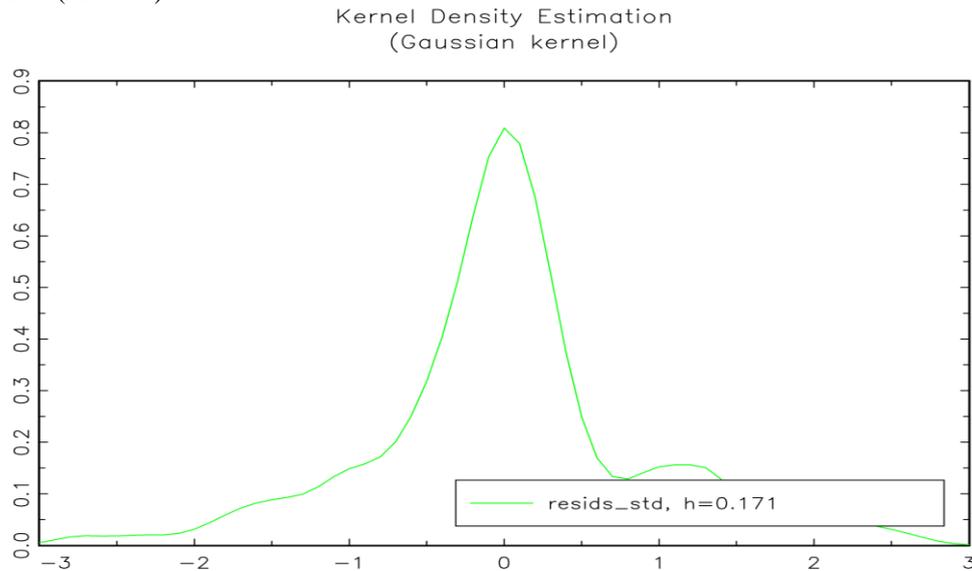


Figure 5a: The normal curve of production

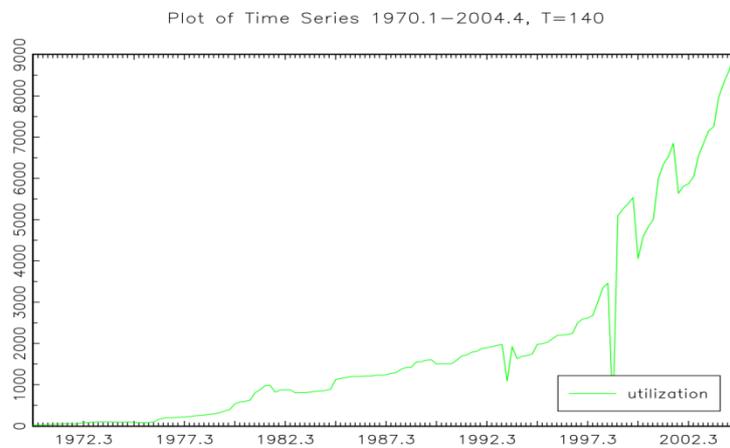


Figure 1b: Time plot utilization before differencing

The variability of the series appears to be changing with time. The mean and variance are not constant, suggesting that the series is not stationary. The Autocorrelation function (ACF) and partial Autocorrelation function (PACF) before differencing

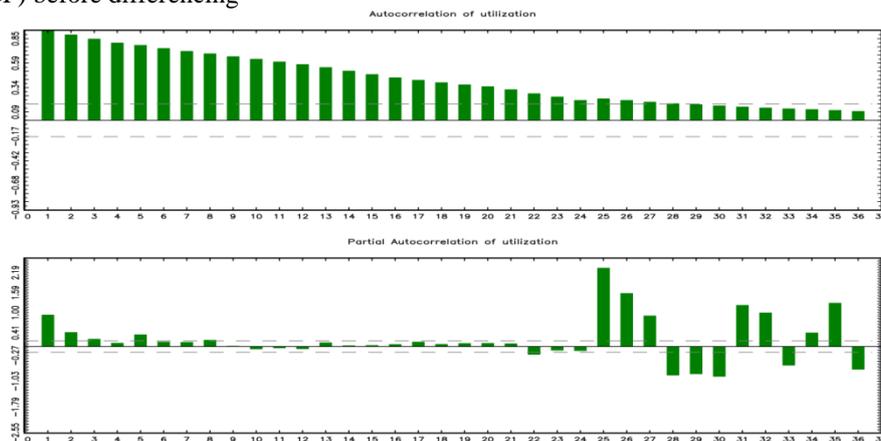


Figure 2b: the Autocorrelation function (ACF) and partial Autocorrelation function (PACF) before differencing.

The ACF shows no cyclical pattern and it was observed that the highest spike is in lag 1. The PACF shows that it has a cyclical pattern and cut off after lag 1 but the highest spike is in lag 1, after lag 35 everything tends to zero

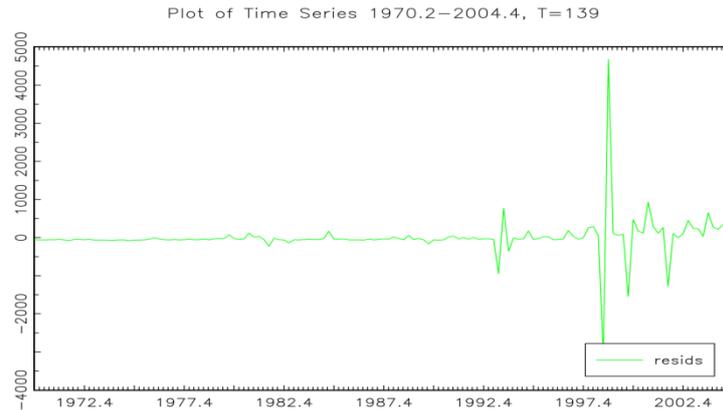


Figure 3b: Time plot for utilization after first differencing

The plot shows that the mean and variance appeared to be constant. Therefore the series is stationary since the mean and variance are constant. The Autocorrelation function (ACF) and Partial autocorrelation function (PACF) after first differencing.

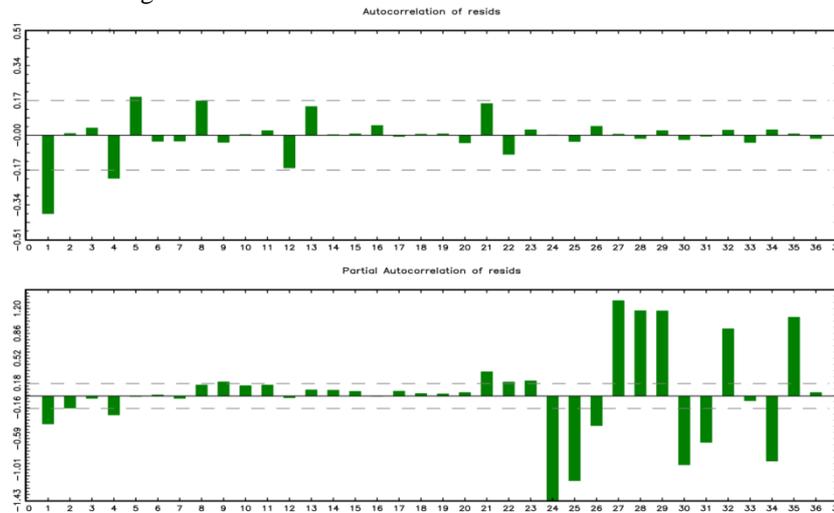


Figure 4b: ACF and PACF plots

There is no correlation function after the first differencing. Therefore the series is stationary. The PACF shows that there is no correlation in the series after first difference. Therefore the series indicate the presence of AR and MA process models

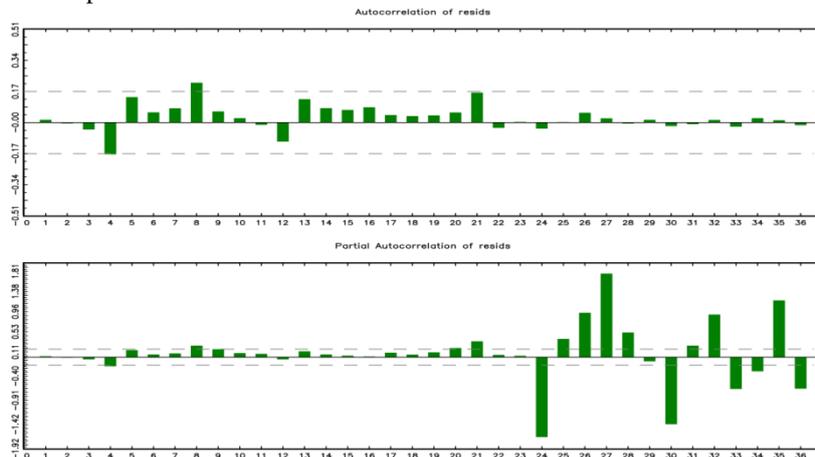


Figure 5b: The correlation shows the following

All spikes for both ACF and PACF are within the confidence interval (upper and Lower), indicating that the selected model (ARIMA) can fit the data.

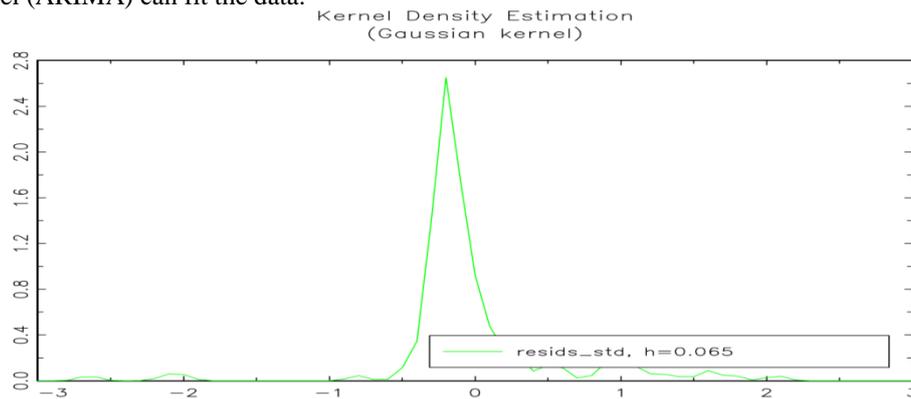


Figure 6b: The Graph of normal curve for utilization

As we discussed before ARIMA (1, 0, 0) (0, 1, 1) model could be written in the following form, as shown in (3):

$$(1-\phi_1B)(1-B^{12}) X_t = (1-\gamma_1B^{12}) e_t \tag{3}$$

This equation can be multiplied out and written in a form that is used in forecasting as shown in (4):

$$X_t = (1+\phi_1)X_{t-12} + \phi_1X_{t-1} + e_t - \gamma_1 e_{t-12} \tag{4}$$

In (4), the value of X_t could be estimated by substitution the parameter values as we estimated above. Fig. 6a and 6b shows a comparison between the real values and the ones resulted from the developed ARIMA model for the period between 1970 and 2004. It is clear that the model was not able to represent the peak values. In addition, it is clear that in both the production and utilization the pattern continues for the upcoming years and there are no indications that the amount of gas decreases with time.

IV. Conclusion

The research is aimed at finding the amount of gas produce and utilized by NNPC within the period under study. This is with a view to find whether there is an increase or decrease in production and usage of gas. Time plot analysis was used in this research work to analyze the pattern of gas produce and utilized from 1970 to 2004. The study observed that the series have irregular pattern, after taken the first differencing the series became stationary. Nevertheless in modeling ARIMA (p, d, q) the best model is ARIMA (1, 1, 1) for production and ARIMA (0, 1, 1) for utilization. A four years forecast have also been made to know the expected amount of Gas production and utilization. The time plot reveals seasonal variation.

From the result, trend is used to find out whether gas is increasing or decreasing and which shows that there is increase in production and utilization of gas with time. From the forecast for production it shows that gas production will be increasing by each quarter and forecast for utilization shows that usage of gas will also be increasing by each quarter. Hence these models can be adopted as far as gas production and utilization is concerned as well as demand and management in Nigeria.

References

- [1] A. M. Herzberg and L. Frew, *Can public policy be influenced* (Environmetrics, 2003) (14)1-10.
- [2] C. K. Lee, *Multifractal characteristics in air pollutant concentration time series*. (Water, Air, and Soil Pollution., 2002) (135), 389-409.
- [3] G.E.P. Box, and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*. Revised Edn. (Hoden-Day, San Francisco, 1976). <http://adsabs.harvard.edu/abs/1976tsaf.conf.....B>
- [4] V. Walter, *Applied time series and Box-Jenkins Models*. (Academic Press Inc., Orlando, Florida, 1983)
- [4] C. Chatfield, *The Analysis of Time Series-an introduction*, 5th Edn., (Chapman and Hall, UK., 1996.)
- [4] D.C.Montgomery, and L.A. Johnson, *Forecasting and Time Series Analysis*. (McGraw-Hill Book Company, 1967), <http://www.abebooks.com/Forecasting-Time->