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

MODELLING AND FORECASTING OF NIGERIAN CRUDE OIL PRICES USING BOX–JENKINS TECHNIQUE

Yakubu .M. Yeldu[1], Shehu .L.[2] and [3]Mukhtar .G.
[1, 2,3]*Department of Statistics, College of Science and Technology*
Waziri Umaru Federal Polytechnic, Birnin Kebbi, Kebbi State, Nigeria.
E-mail: yeldouyakubu@gmail.com, yeldouykb@yahoo.com

ABSTRACT

Box – Jenkins modelling approach has been applied for the time series analysis of Weekly (Forcados, Nigeria) Spot Price FOB (Dollars per Barrel) from August 07, 2000 to September 02, 2013. Basic statistical properties of these series were investigated. After taking the first order difference the time series seems to be stationarity. Autocorrelation and partial autocorrelation plots were used to make tentative identification of the form and order of Box – Jenkins Autoregressive Integrated Moving Average (ARIMA) models. Initially several non – seasonal ARIMA models were postulated for further analysis. These models were then estimated for their adequacy based on the significance of the parameter estimates, mean square and Box – Pierce (Ljung – Box) statistics. Based on these criterion ARIMA (2, 1, 2) model fitted adequately and was also used for short term forecasting. The estimated model is: y_t = 0.2541y_{t-1} + 0.5031y_{t-2} + 0.0179y_{t-1} + 0.1004y_{t-2} + 0.5242 + ε_t.

Keywords: **Time Series, ARIMA Models, Forecasting, Crude Oil, Box-Jenkins Technique.**

INTRODUCTION

The Crude oil industry in Nigeria is the largest industry and main generator of Gross Domestic Product (GDP) in the West African nation which is also the continent's most populous. Oil is a major source of energy in Nigeria and the world in general. Oil being the mainstay of the Nigerian economy plays a vital role in shaping the economic and political destiny of the country. Although Nigeria's oil industry was founded at the beginning of the century, it was not until the end of the Nigeria civil war (1967 - 1970) that the oil industry began to play a prominent role in the economic life of the country.

Nigeria can be categorized as a country that is primarily rural, which depends on primary product exports (especially oil products). Since the attainment of independence in 1960 it has experienced ethnic, regional and religious tensions, magnified by the significant disparities in economic, educational and environmental development in the south and the north. These could be partly attributed to the major discovery of oil in the country which affects and is affected by economic and social components (Odularu, 2007).

According to Oil and Gas Journal (OGJ), Nigeria had an estimated 36.2 billion barrels of proven oil reserves as of January 2009. Before the construction of oil refinery at Port-Harcourt in 1964, oil produced was exported in crude form. Since the oil refinery in Port-hacourt could not meet the needs of both internal consumption and export, oil refineries were constructed in Warri (1967) and Kaduna (1976) and with this, export rose from 245,000 tonnes in 1958 to about 108 million tons in 1974.

A survey conducted in 1997 by the Environmental Resources Managers Ltd, revealed that Nigeria has 159 total oil fields and 1481wells in operation. The most productive region of the nation is the coastal Niger Delta Basin in the Niger Delta or "South-South" region which encompasees78 of the 159 oil fields. Most of Nigeria's oil fields are small and scattered, and as of 1990, these small fields accounted for 62.1% of all Nigerian production. This contrasts with the sixteen largest fields which produced 37.9% of Nigeria's petroleum at that time. As a result of the numerous small fields, an extensive and well-developed pipeline network has been engineered to transport the crude.

STATEMENT OF THE PROBLEM

Since December 2005, Nigeria has experienced increased pipeline vandalism, kidnappings and militant takeovers of oil facilities in the Niger Delta. The Movement for the Emancipation of the Niger Delta (MEND) is the main militant organization attacking oil infrastructure for political objectives, claiming to seek a redistribution of oil wealth and greater local control of the sector. Additionally, kidnappings of oil workers for ransom are also common and security concerns have led some oil services firms to pull out of the country and oil workers unions threatening to strike over security issues for their members.

The instability in the Niger Delta has caused significant amounts of shut-in production and several companies declaring *force majeure* on oil shipments. EIA estimates Nigeria's effective oil production capacity to be around 2.7 million barrels per day (bbl/d) but as a result of attacks on oil infrastructure, 2008 monthly oil production ranged between 1.8 million bbl/d and 2.1 million bbl/d. Additional supply disruptions for the year were the result of worker strikes carried out by the Petroleum and Natural Gas Senior Staff Association of Nigeria (PENGASSAN) that shut-in 800,000 bbl/d of Exxon Mobil's production for about 10 days in late April/early May.

With all these and many more, the crude oil industry is under serious threat to meet, maintain and even surpass its daily production needs in order to remain afloat and be able to compete favourably with other leading producers in the world market. Subsequent government policies has seen series of increase in the pump prices of refined crude oil products such as Petrol, kerosene, diesel, e.t.c. which has caused a lot of unrest and strike actions by pressure and labour groups over the years. To this end, the paper seeks to base on available past data study crude oil price using a time series model.

OBJECTIVES OF THE STUDY

The aim of this paper is to study the pattern and behaviour of the weekly prices of Nigerian crude oil. To achieve this, the following objectives are formulated:

i. To develop Autoregressive integrated moving average (ARIMA) models for the weekly crude oil prices using Box- Jenkins procedure.
ii. To identify among the models the one that best fit the data under study.
iii. Use the model to forecast future crude oil prices.

THEORETICAL FRAME WORK

This study is focused on forecasting of crude oil price using a time series model. There are many studies that addressed the accuracy of crude oil prices modelling and time series forecasting. These include Autoregressive Conditional Hetroscedasticity, ARCH type models (Fong and See, 2002) Neural Networks ANN (Kulkami and Haidar, 2009), Generalized

Autoregressive Hetroscedascity models (Siti *et al.,* 2011), Autoregressive Moving Average (Cabedo and Moya, 2003) and etc in Chang *et al.,* (2009). However, the complexity of the model specification does not guarantee high performance on the out – performed out – of sample forecasts.

Yeldu and Mukhtar (2010) found that if the current trend continues, the production, export, and domestic consumption of Nigeria's crude oil will all be on the increase with 1006402, 838476 and 167926 thousand barrels expected by the year 2015 for production, export and domestic consumption respectively

The Box – Jenkins approach of forecasting crude oil prices are highlighted since these are the focus of the current study. Primarily there are two main approaches to forecasting. In the first approach the forecasts are made in a cause and effect framework where the forecasting variable is assumed to be affected by one or more other variables called the covariates. This is also sometimes called fundamental analysis. This approach seems intuitively more appealing because we are putting forward logical reasons for the ups and downs of our forecasts. There have been several studies e.g. (Michael and John, 2005) used this approach to investigate the effect of inventories and the factors that might have contributed to the oil price increase in addition to demand and supply for crude oil, by expanding a model for crude oil prices to include refinery utilization rates, a nonlinear effect of OPEC capacity utilization, and conditions in futures markets as explanatory variables. This approach has several limitations i.e. we may not be very sure that what variables exactly are causing the changes in the crude oil prices.

Even if we know theoretically the covariates affecting the crude oil prices for example, would be very difficult to identify the exact functional form of these covariates with the crude oil prices. More over it would almost be impossible to make measurements for the future values of these covariates. Therefore the forecasts made on the estimated values of the covariates would have excessively increased forecast errors. The second main forecasting approach is the Time Series modeling. In this approach we no longer predict the future movements of the time series by relating it to a set of other variables. Instead we base our predictions solely on the past behavior of the variable and that variable alone. We may or not be able to explain the up and down movements in the crude oil price time series based on the economic theory or the inventory levels or by intuitive reasoning that why this series behaved the way it did. It may have moved up and down partly in response to socioeconomic or political reasons but much of its movements could have been influenced by the factors that we may simply not be able to explain. In the present study we intend to use this approach to develop a time series modeling approach to forecast weekly average price of Nigerian crude oil.

MODELS AND METHODOLOGY

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. In theory, the most general class of models for forecasting a time series which can be stationary by transformations such as differencing and logging. ARIMA models form an important part of the Box-Jenkins approach to time-series modeling. A non – seasonal ARIMA model is classified as an ARIMA (p, d, q) model, with:

p is the number of autoregressive terms

d is the number of non-seasonal differences

q is the number of moving average lags
$\{y_t\}$ is said to be ARIMA (p,d.q) if

$$(1-L)^d \emptyset'(L) y_t = c + \theta(L)\varepsilon_t$$

Where $\emptyset'(L)$ is defined in $\emptyset(L) = (1-L)\emptyset'(L), \emptyset'(z) \neq 0$ for all $|z| \leq 1$. And $\theta(L)$ is defined in $\theta(z) \neq 0$ for$|z| \leq 1$.

When the process $\{y_t\}$ is stationary if and only if d = 0, in such case it reduces to ARMA (p, q) process: $\emptyset(L)y_t = c + \theta(L)\varepsilon_{t,}\varepsilon_t \sim WN(0, \sigma^2)$

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: Determination of the Model Using ACF and PACF Patterns

Model	**ACF**	**PACF**
AR(p)	Dies down	Cut off 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 (Selecting an Initial Model)

Determine whether the series is stationary or not by considering the graph of ACF. If a graph of ACF of the time series values either cuts 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 differences. An ARMA model is then specified for the differenced series. Differencing is done until a plot of the data indicates the series varies 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 (2.2), 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, denoted 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 $\Delta Y_t = Y_t - Y_{t-1} = Y_t - BY_t = (1 - B)\,Yt$

The stationary series Wt obtained as the dth difference (Δ^d) of Y_t,
$Wt = \Delta^d Y_t = (1 - B)^d Y_t$

ARIMA (p, d, q) has the general form:
$\emptyset_p(B)\,(1 - B)^d Y_t = \mu + \theta_q(B)\varepsilon_t$or $\emptyset_p(B)Wt = \mu + \theta_q(B)\varepsilon_t$

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

STEP 2: Model Estimation

Estimate the parameters for a tentative model has been selected. The unknown parameters of the preliminary ARMA (p,q) model should be estimated. The estimation is easily done by package programs such that E-Views, Minitab, Gretel, etc. Finally, diagnostic checks of the residuals are conducted. If the estimated model is correct, the residuals should behave like a white noise process. Alternatively, the structure of the ARMA process can be determined by using model selection criteria. The most famous ones are the Akaike Information Criterion (AIC) and Schwartz Information Criterion (SIC):

AIC = T ln (residual sum of squares) + 2n,

SIC = T ln (residual sum of squares) + n ln(T),

Where; T is the number of usable observation, and n is the number of parameters to be estimated.

In practice, several ARMA models are estimated, and the one with the smallest AIC or SIC is selected as the best model (Enders, 1995).

STEP 3: Model Checking

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 and should be random. An overall check of model adequacy is provided by the Ljung-Box Q statistic. The test statistic Q is $Q_m = n(n+2)\sum_{k=1}^{m}\frac{r^2k(e)}{n-k} \sim X^2_{m-r}$;

Where rk (e) = the 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 statistic 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 –Smirnov Test.
2. We can check about the randomness of the residuals by considering the graph of ACF and PACF of the residual. The individual residual autocorrelation should be small and generally within ±2/$\sqrt{n}$ of zero.

DATA AND ANALYSIS

The data employed in this paper is Weekly (Forcados, Nigeria) Spot Price FOB (Dollars per Barrel) from August 07, 2000 to September 02, 2013. . The data is obtained from New York Mercantile Exchange (NYMEX) prices on the internet. Firstly, the weekly data of the said period are disposed and analyzed. We denote weekly data of Nigerian Crude prices by p_t. Figure1 and figure 2 presents the line graph of log (Pt), and log (P_t) - log (P_{t-1}).

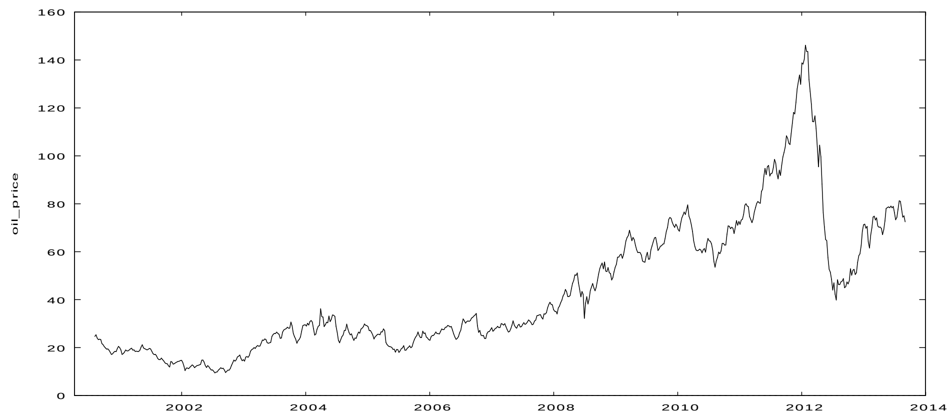

Figure 1: Plot of the Weekly Oil Prices

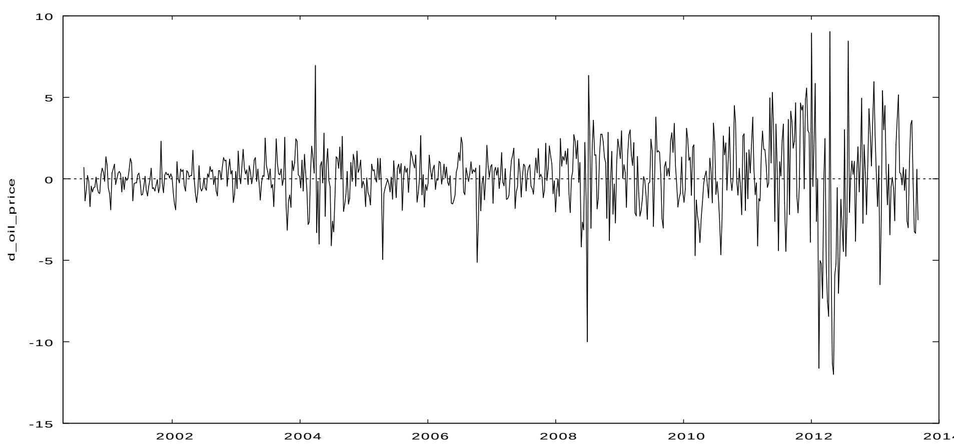

Figure 2: Plot of the Price Returns (First Difference)

From Fig.1, obviously, the graph log (pt) is not stationery. After taking the first order difference the time series seems stationery. We should ensure that the time series being analyzed is stationery before specifying a model. The ADF statistic tests the hypothesis of presence of unit root against the alternative of no unit root and the decision rule is to reject the null hypothesis when the value of test statistic is less than the critical value. The KPSS statistic tests the null hypothesis of stationarity against the alternative of non stationarity and the decision rule is to accept the null hypothesis when the value of test statistic is less than the critical value. The results of the ADF and KPSS test are in Table 4.01.

Table 2 Unit Root Test for the Crude Oil Price Returns

Critical Values	**ADF Test Statistics:-38.0111**	**KPSS Test Statistics: 0.05305**
1%	-3.48	0.216
5%	-2.89	0.146
10%	-2.57	0.119

According to table 2 the ADF test statistics is greater than all the critical values in absolute value so the hypothesis of non – stationarity is rejected. Also is shown that the KPSS test statistic is less than the critical the hypothesis is accepted.

That means the differenced series is stationery which is denoted as *I(0)*. the next is to examine autocorrelation function (ACF) to see the degree of correlation in the data points of the series. The one with higher degree of correlation will be the right candidate to model with.

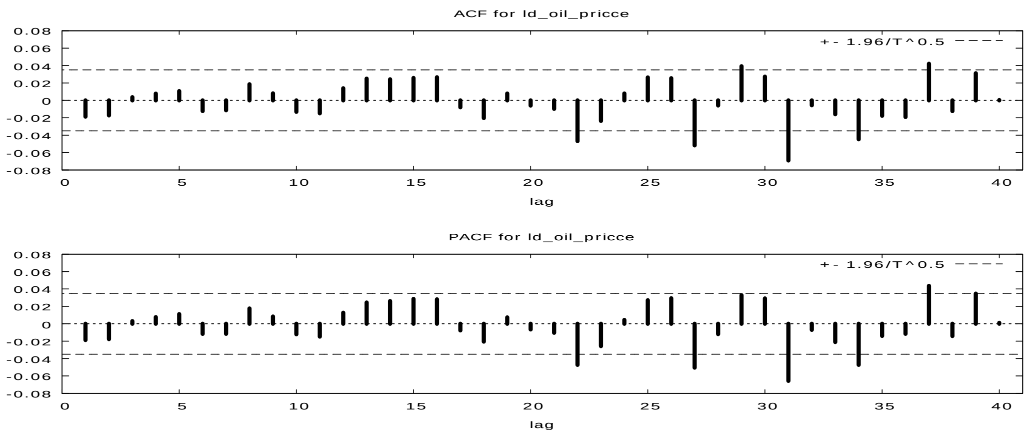

Fig. 3 ACF PACF of the Price Returns

We compare the Akaike Information Criterion (AIC) and Schwartz Information Criterion (SIC) of all the possible models and find out a model to fit the data better than the others, which is the one with the lowest AIC and SIC value. The fitted models are ARIMA (2,1,1), ARIMA (2,1,2) and Finally ARIMA(3,2,3).

Table 5: Comparison of Different Models

Model	**AIC**	**SIC**
ARIMA(2,1,1)	378.993	381.584
ARIMA(2,1,2)	378.992	381.583
ARIMA(3,2,3)	381.066	384.952

From the table above, the final model is ARIMA (2,1,2) we denote ,
then

ESTIMATION

The final model i.e. ARIMA (2, 1, 2) is estimated using maximum likelihood estimation (MLE). The estimated model is:

LJUNG – BOX STATISTIC

Ljung and Box statistic is computed as the weighted sum of squares of a residual of a sequence of autocorrelations, which is used to determine whether a time series consists simply of random values (white noise). Residual Sum is the squares of residuals. It is measure of the discrepancy between the data and the estimation model: RSS = $\sum_{j=1}^{n} \ddot{\varepsilon}_j^2$

H_o: No residual autocorrelation at lag 1 to m

$$LM(m) = n(n+2)\sum_{j=1}^{m}(n-k)^{-1}r_j^2\,(\ddot{\varepsilon})^d \to x^2(m-n-p), where\ r_j(\ddot{\varepsilon}) = \frac{\sum_{j=k+1}^{n}\varepsilon_\iota \ddot{\varepsilon}_{t-k}}{\sum_{i=+1}^{n}\ddot{\varepsilon}^2}$$

Where n is the sample size, k is the lag autocorrelation, and m is the number of lags being tested. Thus, for ARIMA (2, 1, 2) model, we can calculate the X – squared = 0.6345, p – value 0.472 > 0.05. It means we can accept null hypothesis that the residuals are random, and they are independent and identically distributed. The residual plot is shown in fig.4. Based on these criterion a non seasonal model of the form ARIMA (2, 1,2) was chosen for short term forecasting. The forecasts along with 95% prediction limits are given in table 3 which are plotted in figure 5 along with the actual values.

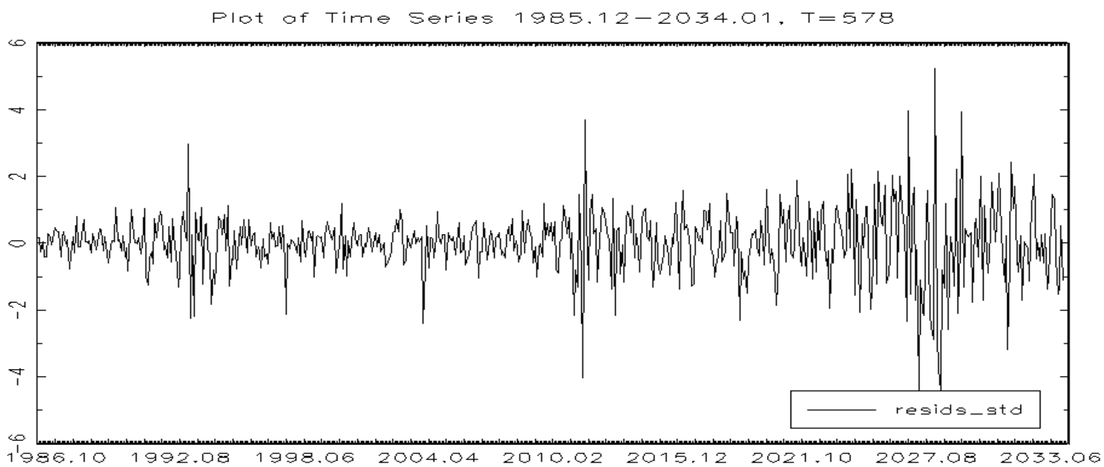

Figure 4: The Residual Plots

Period Forecasted Variable: Weekly Oil Price (In Levels)

Observation	Oil Price	Prediction	Std. Error	95% Interval
2013/09/09	undefined	71.5549	2.10683	(67.4256, 75.6842)
2013/09/16	undefined	71.4200	3.35141	(64.8513, 77.9886)
2013/09/23	undefined	70.9428	3.70457	(63.6820, 78.2037)
2013/09/30	undefined	70.7660	4.01953	(62.8879, 78.6442)
2013/10/07	undefined	70.4967	4.15227	(62.3584, 78.6350)
2013/10/14	undefined	70.3532	4.24801	(62.0272, 78.6791)
2013/10/21	undefined	70.1963	4.29020	(61.7877, 78.6049)
2013/10/28	undefined	70.0988	4.31378	(61.6439, 78.5537)
2013/11/04	undefined	70.0101	4.32180	(61.5395, 78.4807)
2013/11/11	undefined	69.9534	4.32406	(61.4784, 78.4284)
2013/11/18	undefined	69.9094	4.32408	(61.4344, 78.3844)
2013/11/25	undefined	69.8847	4.32483	(61.4082, 78.3612)
2013/12/02	undefined	69.8714	4.32811	(61.3885, 78.3544)
2013/12/09	undefined	69.8708	4.33467	(61.3749, 78.3666)
2013/12/16	undefined	69.8791	4.34507	(61.3629, 78.3952)

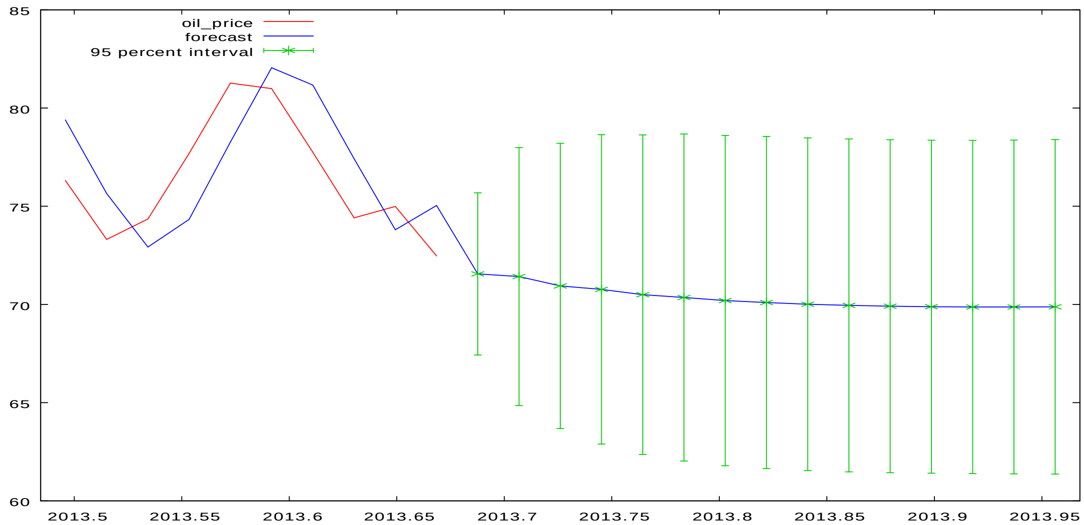

Figure 5: Plot for Weekly Average Crude Oil Price Forecast

CONCLUSION

Weekly (Forcados, Nigeria) Spot Price FOB (Dollars per Barrel) from August 07, 2000 to September 02, 2013 are studied using Box – Jenkins techniques of time series analysis. The prices have a steady increasing trend over some period and a sharp decline and then again a steady slowly increasing series indicating a complex non stationery series. However, the series became stationery just after first difference. The sample autocorrelation and partial sample autocorrelation plots of the first differenced series were used to identify the type and the order of the ARIMA models.

It seems the other models used do not fully capture the variation in prices. The model initially identified by KPSS and ADF test and other diagnostics checking ARIMA (2,1,2) adequately fitted the data. Using the model the 15 periods after the last date of the period under study were used for forecasting.

RECOMMENDATIONS

The findings are relevant for policymakers and industry analysts. They establish the nature of the stochastic process underlying oil prices and the importance of components driving this process. An explanation of the process parameter estimates in terms of the underlying fundamentals for the oil markets are offered in order to comprehend the economics underpinning the observed oil prices dynamics. A change in the process parameters would require a change in the underlying fundamentals. Our alternative modeling approaches are highly relevant for forecasting, risk management, derivatives pricing, and gauging market's sentiment. Our findings could be helpful for monitoring oil markets and developing policies for stabilizing oil markets. With the knowledge of other factors the forecast results can be used to achieve viable and workable frame work on future price patterns.

REFERENCES

Bowerman, B.L. and O'Connel, R.T. Forecasting and Time Series An Applied Approach, 3rd Edition, Duxbury Press, 1993.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. Time Series Analysis, Forecasting and Control, 3rd ed., Prentice Hall, Englewood Clifs, 1994.

Brock Well, P.J. and Davis, R.A. Introduction to Time Series and Forecasting**,** 2nd. ed., Springer, Verlag, 2002.

Chang, C.L., M. McAleer and R. Tansuchat (2009). Modeling Conditional Correlations for Risk Diversification in Crude Oil Markets, to Appear in *Journal of Energy Markets*, Energy Information Administration (EIA). The Official Energy Statistics from the U.S. Government Accessed on 27th May, 2010 @ www.eia.doe.gov/emeu/cabs/Nigeria/background/html

Engle, R.F. (1982) "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", Econometrical, 50, 98, 1OO8.

Environmental Resources Managers limited (September, 1997), *Niger Delta Environmental Survey Final Report Phase I*: Environmental and Socio-Economic Characteristics. Lagos.

Gbadebo, Olusegun Odularu (2008). *Crude Oil and the Nigerian Economic Performance* Oil and Gas Business. Accessed on 5th April, 2010 @ http://www.ogbus.ru/eng/

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

Kuan, C. (2003). Lecture on Basic Time Series Models. *Institute of Economics, Academia Sinica, Lecture Notes.*

Kwiatkowski, D., Philips, P.C.B., Schmidt, P. and Shin, Y. (1992) Testing the Null of Stationary Against the Alternative of Unit Root: How Sure Are We That The Economic Time Series Have A Unit Root, *Journal of Econometrics*, 54, 159 – 178.

Ljung, G.M. and Box, G.E.P. (1978). On a Measure of Lack of Fit in Time Series Models. Biometrika 65, p. 553–564.

Michael, Y., John Z, and Joanne. A Monthly Crude Oil Spot Price Forecasting Model Using Relative Inventories. *International Journal of Forecasting,* Volume 21, Issue 3, July-September 2005, Pages 491-501

NYMEX Sites on net http://quotes.ino.com/exchanges/?e=NYMEX. Accessed 05/01/2010

Oil and Gas Journal, Volume 107, Issue 1 January, 2009. Accessed on the 17th of April, 2010 at www.ogj.com/index.html

Yeldu Y.M and Mukhtar G. (2011). Forecasting the Production, Domestic Consumption and Export in Nigeria's Crude Oil Industry. Accepted for Publication in *Nigerian Journal of Science and Technology. NJST Vol.7 No.1*.

Reference to this paper should be made as follows: Yakubu .M. Yeldu *et al.,* (2013), Modelling and Forecasting of Nigerian Crude Oil Prices Using Box – Jenkins Technique. *J. of Physical Science and Innovation*, Vol. 5, No. 2, Pp. 50 – 59.