International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Study of Autoregressive and Nonlinear Models for Stationarity and Invertibility of Time Series Analysis

Bature, Tajudeen Atanda[1], Yusuff, Nurudeen Ayinla[2], Hamisu, Salisu[3], Lawal, Adekunle Yusuf[4], Ishola, Kabiru Babatunde[5], Ahmed Yusuf Shero[6]

[1]Department of Statistics, Ahmadu Bello University, Zaria, Nigeria

[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria

[3]Department of Statistics, Ahmadu Bello University, Zaria, Nigeria

[4]Nigerian Army College of Education, Ilorin, Nigeria

[5]Department of Computer Science, Federal University Lafia, Lafia, Nigeria

[6]Department of Statistics, University of Ilorin, Ilorin, Nigeria

Corresponding Email: tajudeenatanda56@yahoo.com

ABSTRACT

In this research we attempt to examine empirically the best model for forecasting. The data employed in this study comprised 264 monthly observations of average monthly crude oil price in Nigeria spanning from January, 1997 to December, 2018. Analysis was carried out, at first the stationarity condition of the data series was observed by autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, then checked using Kwiatkowski-philips-schmidt-shin (KPSS) and Augmented Dickey Fuller (ADF) test statistic. It has been found that Average monthly crude oil price is non-stationary. After we have taken first difference of the series, the same types of plots and the same types of statistic show that the data is stationary. The best ARIMA models have been selected by using the criteria such as AIC, BIC and SIC. The model for which the values of criteria are smallest is considered as the best model and by comparing the models selected, one with minimum mean squared error, root mean square error and mean absolute error is the best. Hence ARIMA (0, 1, 2) is found as the best model for forecasting the average monthly crude oil price data.

KEYWORDS: Autocorrelation Function,Stationarity, Autoregressive Model, Moving Average Model, Autoregressive Moving Average Model, White Noise, Forecasting and Autoregressive Integrated Moving Average Model.

1.0 INTRODUCTION

Time, in terms of years, months, days or hours is a device that enables one to relate phenomena to a set of common, stable and reference points. In making conscious decisions under uncertainty, we all make forecasts. Almost all managerial decisions are based on some form of forecast. Essentially the concept of time series is based on the historical observations. It involves explaining past observations in order to try to predict those in the future (Ahiati, 2007). A time series is a collection of observations measured sequentially through time. These measurements may be made continuously through time or be taken at a discrete set of time points (Chatfield, 2000).

In recent years, crude oil price has become one of the major economic challenges facing most countries in the world especially those in Africa including Nigeria. Crude oil is a major focus of economic policy worldwide as described by Ayadi, O.F. (2005).Crude oil price dynamics and evolution can be studied using a stochastic modeling approach that captures the time dependent structure embedded in the time series crude oil price data. The Autoregressive Integrated Moving Average (ARIMA) popularly known as Box-Jenkins Methodology (G. P. E. Box and G. M. Jenkins (1978)) and the autoregressive conditional heteroscedasticity (ARCH) models, with its extension to generalized autoregressive conditional heteroscedasticity (GARCH) models as introduced by Engle (1982) and Bollerslev (1986) respectively accommodates the dynamics of conditional heteroscedasticity (the changing variance nature of the data). Heteroscedasticity affects the accuracy of forecast confidence limits and thus has to be handled properly by constructing appropriate non-constant variance models (Amos, 2010).

2.0 LITERATURE REVIEW

Another definition of a time series is that of a collection of quantitative observations that are evenly spaced in time and measured successively. Examples of time series include the continuous monitoring of a person’s heart rate, hourly readings of air temperature, daily closing price of a company stock, monthly rainfall data, and yearly sales figures. Time series analysis is generally used when there are 50 or more data points in a series. If the time series exhibits seasonality, there should be 4 to 5 cycles of observations in order to fit a seasonal model to the data, Brockwell, P.J. and Davis, R.A. (2002).

Models for time series data can have many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are the autoregressive (AR) models, the integrated (I) models, and the moving average (MA) models. These three classes depend linearly on previous data points. Combinations of these ideas produce autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models. The autoregressive fractionally integrated moving average (ARFIMA) model generalizes the former three, Gershenfeld, N.(1999).

Forecasting is an important part of econometric analysis, for some people probably the most important. How do we forecast economic variables, such as GDP, inflation, exchange rates, stock prices, unemployment rates, and myriad other economic variables? The method of forecasting that have become quite popular: Autoregressive Integrated Moving Average (ARIMA) popularly known as Box-Jenkins Methodology (G. P. E. Box and G. M. Jenkins (1978)), and the special problems involved in forecasting prices of financial assets, such as stock prices, exchange rates, etc. These assets prices are characterized by the phenomenon known as Volatility clustering, that is, periods in which they exhibit wide swings for an extended time period followed by a period of comparative tranquility. One only has to look at the Dow-Jones index in the recent past. The so-called Autoregressive Conditional Heteroskedasticity (ARCH) models can capture such volatility clustering. Philip Franses noted; "since such (financial time series) data reflect the result of trading among buyers and sellers at, for example, stock markets, various sources of news and other exogenous economic events may have an impact on the time series pattern of asset prices.Given that news can lead to various interpretations, and also given that specific economic events like an oil crisis cases can last for some time, we often observe that large positive and large negative observations in financial time series tend to appear in clusters." (Philip Hans Franses, (1998))

In his seminal 1982 paper, Robert F. Engle described a time series model with a time-varying volatility. Engle showed that this model, which he called ARCH (autoregressive conditionally heteroscedastic), is well-suited for the description of economic and financial price. Nowadays ARCH has been replaced by more general and more sophisticated models, such as GARCH (generalized autoregressive heteroscedastic). (Engle, (1982))

In econometrics, Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests are used for testing a null hypothesis that an observable time series is stationary around a deterministic trend. Such models were proposed in 1982 by Alok Bhargava in his Ph.D. thesis where several John von Neumann or Durbin–Watson type finite sample tests for unit roots were developed (see Bhargava, 1986). Later, Denis Kwiatkowski, Peter C.B. Phillips, Peter Schmidt and Yongcheol Shin (1992) proposed a test of the null hypothesis that an observable series is trend stationary (stationary around a deterministic trend). The series is expressed as the sum of deterministic trend, random walk, and stationary error, and the test is the Lagrange multiplier test of the hypothesis that the random walk has zero variance. KPSS type tests are intended to complement unit root tests, such as the Dickey–Fuller tests. By testing both the unit root hypothesis and the stationarity hypothesis, one can distinguish series that appear to be stationary, series that appear to have a unit root, and series for which the data (or the tests) are not sufficiently informative to be sure whether they are stationary or integrated.

It is evident from this chapter that substantial efforts have been put into the development, specification and estimation of volatility models. Naturally, the issue of forecasting is also important, since volatility figures prominently in a variety of applications in investment, portfolio management, asset pricing, risk management and monetary policy. A large literature has appeared over recent years on investigating which model is superior in terms of predictive power and why. Poon and Granger (2003) review some ninety-three papers that appeared over two decades on this subject, and conclude that implied volatility estimated from options data appears to provide the most reliable forecasts as a wider information set is used. GARCH models generally rank second, often having comparable performance to that of simple volatility forecasts based on smoothing filters, especially for series that are likely to contain non-stationarities.

3.0 RESEARCH METHODOLOGY

This chapter is designed to explain the basic methodology of the research work. Hence the various methods applied for the purpose of achieving the objectives of the study are going to be explained herewith. In particular, the basic procedures of the Box-Jenkins methodology is explained.

Forecasting is an important part of econometric analysis, for some people probably the most important. How do we forecast economic variables, such as GDP, inflation, exchange rates, stock prices, unemployment rates, and myriad other economic variables? In this chapter, we discuss the method of forecasting that have become quite popular: Autoregressive Integrated Moving Average (ARIMA) popularly known as Box-Jenkins Methodology (G. P. E. Box and G. M. Jenkins (1978)), and the special problems involved in forecasting prices of financial assets, such as stock prices, exchange rates, etc. These assets prices are characterized by the phenomenon known as Volatility clustering, that is, periods in which they exhibit wide swings for an extended time period followed by a period of comparative tranquility. One only has to look at the Dow-Jones index in the recent past.

3.1 AR(p), MA(q), AND ARIMA MODELING OF TIME SERIES DATA

To introduce several ideas, some old and some new, weillustrated these with the Oil Price time series data for Nigeria (see APPENDIX). A plot of this time series is given in chapter 4, figures 4.1.1a, 4.1.1b and 4.1.2 (undifferenced logged) and 4.1.3 and 4.1.4 (first differenced logged). Note: that the data in level form is non-stationary but in the (first) difference form it is stationary.

If a time series is stationary, we can model it in a variety of ways.

3.1.1 An Autoregressive (AR) process

Let y_t represent the logged data at time t . if we model y_t as

$$y_t = \alpha_1 y_{t-1} + e_t \qquad (3.1)$$

Where e_t is an uncorrelated random error term with zero mean and constant variance σ^2 (i.e. it is white noise), then we say that y_t follows a first order autoregressive, or AR(1), stochastic process. Here the value of y at time t depends on its value in the previous time period and a randomterm. In other words, this model says that the forecast value of y at time t is simply some proportion $(=\alpha_1)$ of its value at time $(t-1)$ plus a random shock or disturbance e at time t.

But if we consider this model

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + e_t \qquad (3.2)$$

Then we say that y_t follows a second-order autoregressive, or AR(2), process. That is, the value the value of y at time t on its value in the previous two time period.

In general, we can have

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + e_t$$
$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + e_t \qquad (3.3)$$

In which case y_t is a p[th]-order autoregressive, or AR(P), process.

3.1.2 A Moving Average (MA) Process

The AR process just discussed is not the only mechanism that may have generated y.

Suppose we model y as follows:

$$y_t = \theta + \beta_0 e_t + \beta_1 e_{t-1} \qquad (3.4)$$

Where θ is a constant and e, as before, is the white noise stochastic error term, and $\beta_0 e_t = e_t$

Here we say that y follows a first-order moving average, or MA(1), process.

But if we y follows the expression

$$y_t = \theta + \beta_0 e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} \qquad (3.5)$$

Then it is an MA (2) process.

More generally,

$$y_t = \theta + \beta_0 e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \ldots + \beta_q e_{t-q}$$
$$y_t = \theta + \sum_{i=0}^{q} \beta_i e_{t-i} + e_t \qquad (3.6)$$

is an MA (q) process. In short, a moving average process is simply a linear combination of white noise error terms.

4.0 DATA ANALYSIS

Having explored the general theory of ARIMA model in the preceding chapter, this chapter is dedicated to fitting the ARIMA model to the Average Crude Oil Price data courtesy of the YCHART (ychart.com) ,with a view to achieve the pre-stated aim and objectives of the research work. The data employed in this study comprise of 264 monthly observations of the Average Crude Oil Price in Nigeria spanning from January, 1997 to December, 2018.

4.1 GRAPHICAL ANALYSIS

4.1.1a Time plot of the series before differencing

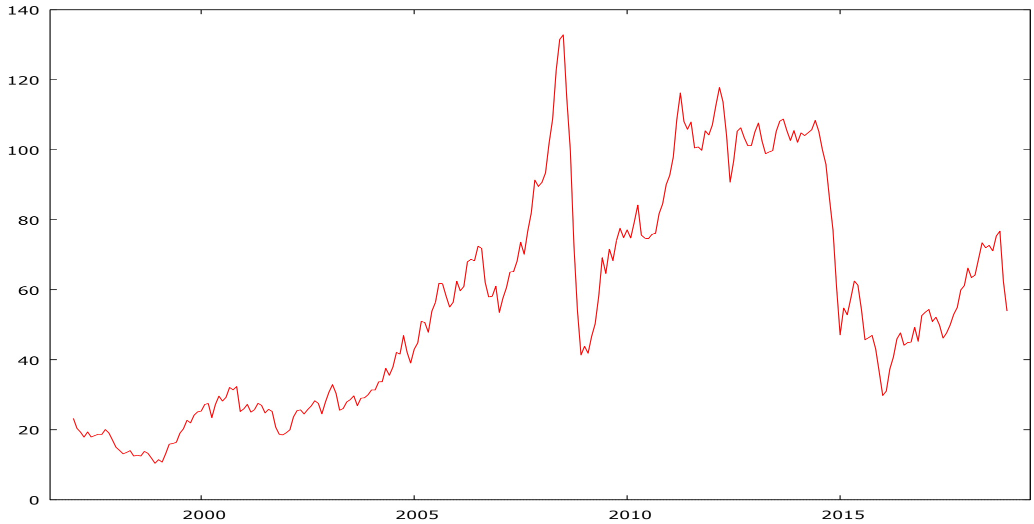

Figure 4.1.1a Time plot of the series before differencing

The time plot of the series above gives an initial step about the likely nature of the crude oil price time series. The series plot exhibit the phenomenon of volatility clustering, that is, periods (in which the oil price) shows wide swings for an extended time period followed by periods in which there is relative calm.Over the period of study, the crude oil price has been increasing, that is, showing an upward trend, in a fluctuationnal pattern, suggesting perhaps that the mean and the variance of the log of crude oil price has been changing with time or over time in Nigeria.

4.1.1b Time plot for the log of the series before differencing

In the case of strongly trending series such as the Oil prices, and so on, higher levels of the variable in question are likelyto be associated with higher variability in absolute terms. The obvious "fix", employed in manymacro-econometric studies, is to use the logs of such series rather than the raw levels. Provided the*proportional* variability of such series remains roughly constant over time, the log transformationis effective. As a result we take the log transformation of the data, and henceforth, the log of this series will be used throughout.

Over the period of study, the log of crude oil price has been increasing, that is, showing an upward trend, in a fluctuationnal pattern, suggesting perhaps that the mean and the variance of the log of crude oil price has been changing with time or over time in Nigeria. The plot of the log below confirms that series is still trending upward. As such there is need to difference the series.

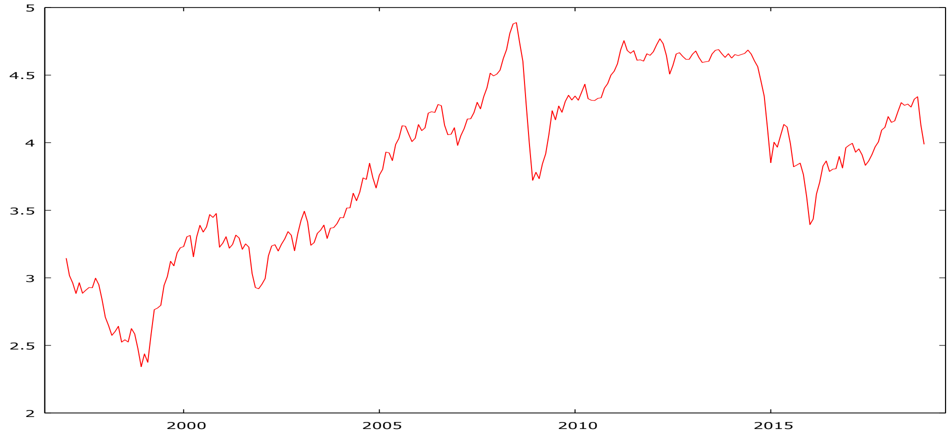

Figure 4.1.1b Time plot of the series' log before differencing

4.1.2 The ACF and PACF of the log before differencing

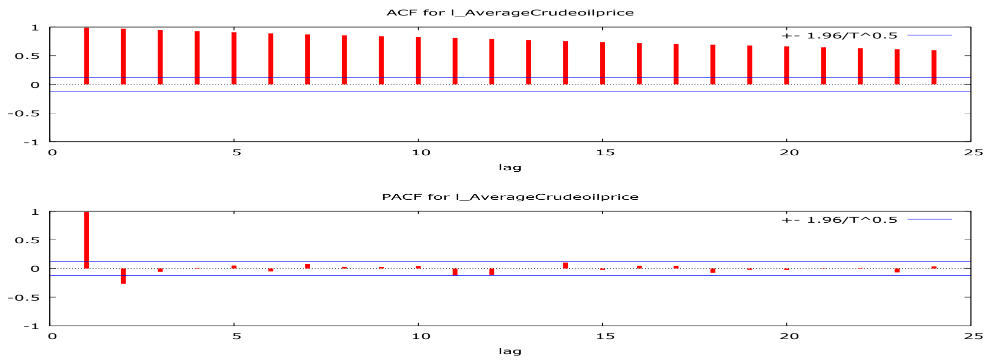

Figure 4.1.2: The ACF and PACF of the log before differencing

The figure above is the correlogram (ACF and PACF) of the crude oil price's log of the data series before differencing. The most striking feature of this correlogram is that the autocorrelation coefficients at various lags are very high, (at lag 1 = 0.9883) up to a lag of 24 months (at lag 24 = 0.5968); these are individually statistically significantly different from zero, out of the 95% confidence bounds. This is the typical correlogram of a non-stationary time series. The autocorrelation starts at a very high value and a decline (spikes down) very slowly toward zero as the lags lengthens, showing a purely MA series. After the first two lags, the PACF drops dramatically, and most PACFs after lag 2 are statistically insignificant, showing an AR of order 2.

4.1.3 UNIT ROOT TESTS BEFORE DIFFERENCING

Table 4.1.1: ADF and KPSS tests

TEST	**TEST STATISTIC**	**CRITICAL/P-VALUES**
ADF without constant	0.220186	0.7502
ADF with constant	-1.79022	0.3858
ADF with constant and trend	-1.8273	0.6916

A formal application of the Augmented Dickey-Fuller is shown in the table above, confirming the log of the average crude oil price series is not stationary, and since the tests statistics are greater than the level of significance. $(\alpha = 0.05)$

Since the time series is not stationary, we have to make it stationary before we can apply the Box-Jenkins methodology. This can be done by differencing the series once.

4.1.4 Time plot of the series after differencing

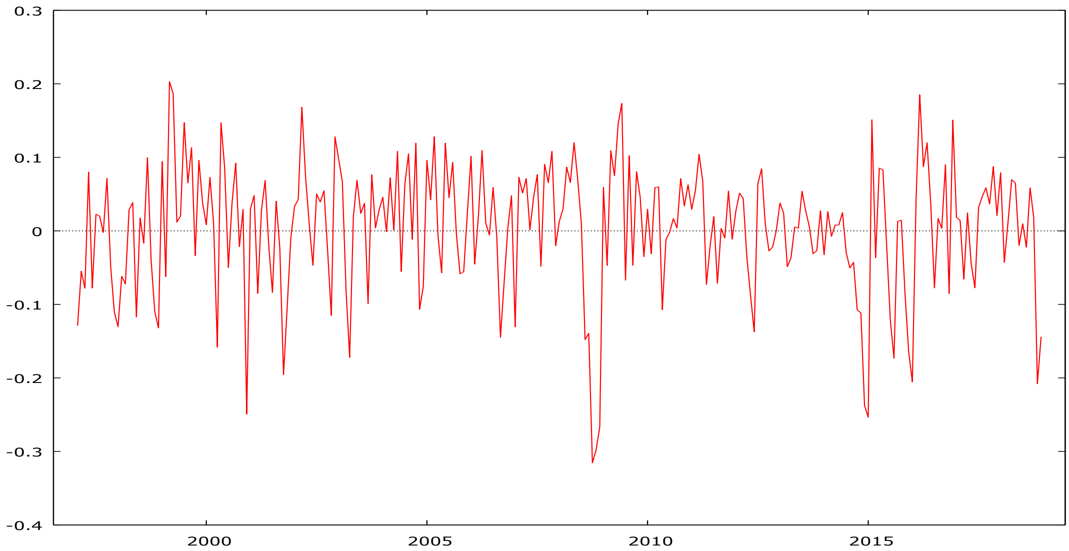

Figure 4.1.3: Time plot of the series after differencing

The figure above plotted the first differences for the log of the average crude oil price series. Unlike Figure 4.1.1b, we do not observe any trend in this plot.

We can also see this visually from the ACF and PACF correlogram given below.

4.1.5 ACF and PACF after differencing

The ACFs at lag 1 and 11, PACFs at 1, 9 and 12 seem statistically different from zero (at the 95% confidence limit, those lags are asymptotic and so can be considered approximate), but at all other lags, they are not statistically different from zero. We therefore conclude that the data series is now stationary. A formal application of the Augmented Dickey-Fuller and KPSS unit root tests below may show that this is indeed the case.

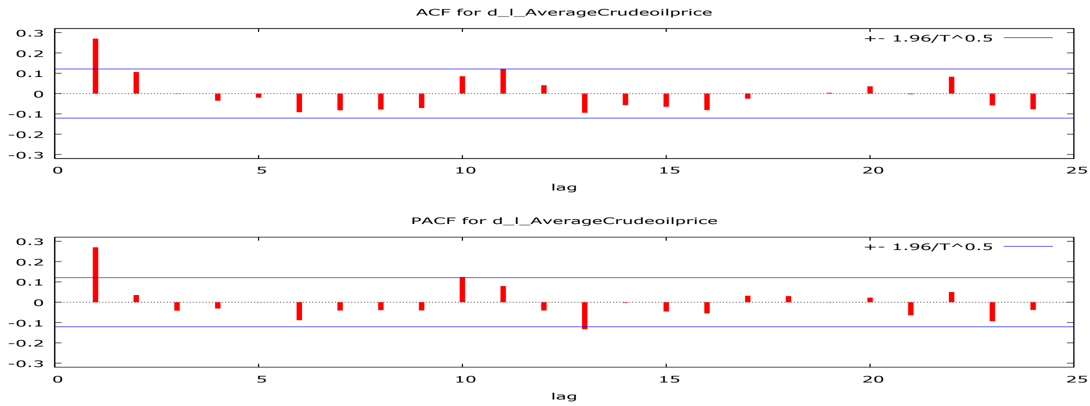

Figure 4.1.4: ACF and PACF after differencing

Since the entire test statistics of the ADFs are less than the critical regions, we reject the null hypothesis and therefore conclude that there is no unit root or the time series is stationary. Likewise, the KPSS tests, the test statistics are all less than the p-values; we therefore accept the null hypothesis and conclude that the data series is stationary around a deterministic trend.

4.1.6 UNIT ROOT TESTS AFTER DIFFERENCING

Table 4.1.2: ADF and KPSS tests

TEST	TEST STATISTIC	CRITICAL/P-VALUES
ADF without constant	-12.1716	5.782e-024
ADF with constant	-12.1656	1.016e-021
ADF with constant and trend	-12.1742	1.381e-021
KPSS without trend	0.118298	1% 5% 10% 0.7400.4620.348

4.2 MODEL IDENTIFICATION AND SELECTION

Table 4.2.1: Result of ARIMA model identification and selection

MODEL	AIC	HQC	SIC
ARIMA(0,1,1)	-550.8718	-546.5652	-540.1554
ARIMA(0,1,2)	-552.5318	-546.7895	-538.2431
ARIMA(0,1,3)	-550.5953	-543.4174	-532.7345
ARIMA(1,1,0)	-553.6762	-549.3695	-542.9598
ARIMA(1,1,1)	-551.9945	-546.2523	-537.7059
ARIMA(1,1,2)	-550.5724	-543.3946	-532.7116
ARIMA(1,1,3)	-550.7320	-542.1187	-529.2991
ARIMA(2,1,0)	-552.1071	-546.3648	-537.8184
ARIMA(2,1,1)	-550.2585	-543.0807	-532.3978
ARIMA(2,1,2)	-548.6713	-540.0580	-527.2384
ARIMA(2,1,3)	-548.4873	-538.4383	-523.4822
ARIMA(3,1,0)	-550.5806	-543.4028	-532.7198
ARIMA(3,1,1)	-550.4501	-541.8367	-529.0171
ARIMA(3,1,2)	-548.4709	-538.4220	-523.4659
ARIMA(3,1,3)	-546.4926	-535.0081	-517.9154

The table above tested fifteen (15) models with low AIC, HQC and SIC which is common in ARIMA modeling and find the best models among them. ARIMA (0,1,2), ARIMA (1,1,0) and ARIMA (2,1,0) models are selected because they have minimum AIC, HQC and SIC.

4.3 MODEL ESTIMATION

After the best models have been chosen, the parameters of the models are estimated. The results of these estimates are shown in the tables below: -

Table 4.3.1: Result of ARIMA (0,1,2) model estimation

PARAMETER	COEFFICIENT	STD. ERROR	T-RATIO	P-VALUE
Constant	0.00278800	0.00712400	0.3914	0.6955
theta_1	0.263072	0.0609127	4.319	1.57e-05
theta_2	0.124876	0.0642521	1.944	0.0520

The estimated ARIMA (0, 1, 2) model is given by

$$\boldsymbol{Y_t = 0.00278800 + e_t + 0.263072e_{t-1} + 0.124876e_{t-2}}$$

Table 4.3.2: Result of ARIMA (1,1,0) model estimation

PARAMETER	COEFFICIENT	STD. ERROR	T-RATIO	P-VALUE
Constant	0.00280302	0.00709005	0.3953	0.6926
phi_1	0.274883	0.0597816	4.598	4.26e-06

The estimated ARIMA (1, 1, 0) model is given by

$$\boldsymbol{Y_t = 0.00280302 + 0.274883Y_{t-1} + e_t}$$

Table 4.3.3: Result of ARIMA (2,1,0) model estimation

PARAMETER	COEFFICIENT	STD. ERROR	T-RATIO	P-VALUE
Constant	0.00267898	0.00739485	0.3623	0.7171
phi_1	0.264337	0.0618437	4.274	1.92e-05
phi_2	0.0410578	0.0625268	0.6566	0.5114

The estimated ARIMA (1, 1, 0) model is given by

$$\boldsymbol{Y_t = 0.00267898 + 0.264337Y_{t-1} + 0.0410578Y_{t-2} + e_t}$$

5.0 CONCLUSION

The study has presented us with an opportunity to have an extensive understanding of the theory of time series analysis and its application to real life situation. The stages in the model building (that is the identification, estimation and checking) strategy has been explored and utilized. Based on minimum AIC, SIC and HQC values, the best fit ARIMA modelstrend to be ARIMA(0,1,2), ARIMA(1,1,0), ARIMA(2,1,0). Afterestimation of the parameters of selected models, a series of diagnostic and forecast accuracy tests were performed. Having satisfied all the model assumptions, ARIMA(0,1,2)models (with smaller MSE, RMSE and MAE which is an indication that they explains the data better than the other models) is judged to be the best models for forecasting in the Box-Jenkins, and therefore these models is used for a six months forecast.

5.1 RECOMMENDATION

Although the application in this research study is based on oil price data, future research can also consider;

- The multivariate GARCH, where other macro-economic variables such as exchange rate and money supply will be involved in the model.
- Other areas of application for instance, environmental and pollution data, health researches in the context of longitudinal data, agriculture and geo-statistics to mention but few.

As a whole, the picture paints an unstable future for the Nigerian economy following oil price shocks. There is a strong need for policy makers to focus on policies that will strengthen/stabilize the macroeconomic structure of the Nigerian economy with specific focus on; alternative sources of government revenue (reduction of dependence on oil proceeds), reduction in monetization of crude oil receipts (fiscal discipline), aggressive saving of proceeds from oil booms in future in order to withstand vicissitudes of oil shocks in future.

REFRENCES

- Abramson B., A. Finizza, (2015). Using belief networks to forecast oil prices, *International Journal of Forecasting*, **7**(3): 299–315.
- Abramson B., A. Finizza, (2005). Probabilistic forecasts from probabilistic models: a case study in the oil market, *International Journal of Forecasting*, **11**(1): 63–72.
- Adeniyi, O. (2009). "OIL PRICE SHOCKS AND MACROECONOMIC PERFORMANCE IN NIGERIA", Department of Economics University of Ibadan, Nigeria.
- Barone-Adesi G., F. Bourgoin, K. Giannopoulos, (1998). Don't look back, *Risk August*, **8**: 100–103.
- Bhargava, A. (1986). "On the Theory of Testing for Unit Roots in Observed Time Series". *The Review of Economic Studies***53** (3): 369–384.
- Blanchard,O.J and J. Gali (2007). "The macroeconomic effects of Oil Price Shocks: Why are the 2000s different from the 1970s?", MIT department of Economics Working Paper No.07-21.
- Bollerslev, Tim (1986). "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31:307-327
- Campbell, J.Y, Lo, A.W. and Mackinlay, A.C (1997), Modelling daily value at risk using realized volatility and ARCH type models, Journal of empirical finance volume 11, no. 3 pages 379-398.
- CBN (2009). Annual Reports and statement of Accounts for the year ended 31st December.
- Denis Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Yongcheol Shin (1992): Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. Journal of Econometrics 54, 159–178.
- Dickey, D. A., and Fuller, W. A. (1979). 'Distribution of the Estimators for Autoregressive Time Series with a Unit Root', *Journal of the American Statistical Association*, 74, 427–31.
- Fan, J. and Yao, Q. (1998) Efficient estimation of conditional variance functions in stochastic regression. *Biometrika***85**, 645-660.
- Francq, C., Roussignol, M. & J. M. Zako¨ıan (2001): Conditional heteroskedasticity driven by hidden Markov chains. Journal of Time Series Analysis 22, 197-220.
- Gershenfeld, N. (1999). *The Nature of Mathematical Modeling*. New York: Cambridge University Press. pp. 205–208.
- Geweke, John; Horowitz, Joel; Pesaran, Hashem (2008). "Econometrics". In Durlauf, Steven N.; Blume, Lawrence E. *The New Palgrave Dictionary of Economics* (Palgrave Macmillan).
- Ghysels, E., Santa-Clara P., and Valkanov, R. (2006), 'Predicting Volatility: Getting the Mostout of Return Data Sampled at Different Frequencies', *Journal of Econometrics*, 131, 59–95.
- H. G. Huntington, (1994), Oil price forecasting in the 1980s: what went wrong? *The Energy Journal*, **15**(2): 1–22.
- H. P. Pesaran (1990), "Econometrics," *Econometrics: The New Palgrave*, p. 2, citing Ragnar Frisch (1936), "A Note on the Term 'Econometrics'," *Econometrica*, 4(1), p. 95.