Modeling and Forecasting Nigerian Crude Oil Exportation: Seasonal Autoregressive Integrated Moving Average Approach

Kayode Ayinde[1], Habib Abdulwahab [2]

[1]Department of Statistics, Ladoke Akintola University of Technology, P.M.B. 4000
Ogbomoso, Oyo State, Nigeria

[2]Department of Mathematics,
Usmanu Danfodio University, P.M.B. 2346, Sokoto State, Nigeria

Abstract: ***In the last few decades, crude oil (petroleum) has claimed the topmost position in Nigerian export list, constituting a very fundamental change in the structure of Nigerian international trade. This paper is intended to model and forecast Nigerian monthly crude oil exportation (in barrels) by applying seasonal autoregressive moving average (SARIMA) into the same data collected between January, 2002 and December, 2011. Result reveals an upward trend of the series which became stationary at 1st difference, a sharp drop between 2007 and 2009 and autocorrelation function with significant spikes at lag 1, 7 and 12 suggesting the presence of seasonality in the series. Based on Akaike Information Criterion (AIC), Schwartz Bayesian Information Criterion (SBIC) and Hannan-Quinn Information Criterion (HQC), the best model was SARIMA (1, 1, 1) x (0, 1, 1)$_{12}$. The diagnosis on such model was confirmed, the error was white noise, presence of no serial correlation and a forecast for current and future values within 24 months period was made which indicates that the crude oil exportation is fairly unstable.***

Keywords: *SARIMA Model, BIC, Forecasting, Crude Oil Exportation*.

1. Introduction

Nigeria is a mono product economy because oil accounts for over 95 percent of its export earnings. Moreover, about 70 percent of government revenue is derived from oil and over 90 percent of new investments are associated with oil and allied products [1]. As an oil-exporting Third World nation, Nigeria's economic development has witnessed trials and tribulations, as the nation's fortunes have risen and fallen in the stormy seas of the international oil market. Nigeria's vulnerability to oil price shocks stems from the nations over dependence on crude oil export. This is amply evident from the drastic decline in non-oil exports over the past three decades of petroleum production in Nigeria.

Crude oil accounted for 7.1 per cent of total exports in 1961, which was dominated at that time by cocoa, groundnut and rubber, in that order. In 1965, oil had climbed to 13.5 per cent of the nation's export earnings, and by 1970, it had become the leading source of foreign exchange, accounting for 63.9 per cent. The 1973 Arab oil embargo against the United States of America did not only earn Nigeria the windfall revenue of an oil boom. By 1979, petroleum sales had completely overshadowed non-oil exports, as it then contributed about 95 per cent of the country's export earnings. During the peak of the oil boom, Nigeria's premium crude, the Bonny Light (37° API), fetched the commanding price of 40 dollars a barrel. In 1990, following the Gulf War and the United Nations trade embargo on Iraq and Kuwait, not only did the Organization of Petroleum Exporting Countries (OPEC) re-allocated the production shares of both nations to other producers such as Nigeria: There was also a sharp momentary increase in crude oil prices. From the low spot price of 15.49 dollars in June 1990, the average spot price of the Bonny Light soared to 36.78 dollars a barrel in August 1990. In 1993, the spot price of the Bonny Light, on the average, was about 18 U.S. dollars per barrel [2]. As of 2000, oil and gas exports accounted for more than 98% of export earnings and about 83% of federal government revenue, as well as generating more than 14% of its GDP. It also provided 95% of foreign exchange earnings and about 65% of government budgetary revenues. Nigeria proving oil reserves were estimated by Unites States Energy Information Administration (EIA) at between 16 and 22 billion barrels ($3.5x10^9$ m^3), but other sources claims there could be as much as 35.3 billion barrels ($5.61x10^9$ m^3). Its reserves make Nigeria the 10th most petroleum rich nation, and by the far the most affluent in Africa. In mid 2001 its crude oil production was averaging around 2.2 million barrels (350,000 m^3) per day [3]. In 2012, National Bureau of Statistics (NBS) breakdown of Nigeria's export by commodities showed that mineral products was the highest exported commodity in the year under review and it accounted for 84.1% of the total export value to N18.87 trillion. the NBS further said that crude oil export is still determinant factor to the structure of Nigeria's export, which contributed a total of N15.5trillion representing 69.2% of the total export in 2012[4].

So strategic is the petroleum sector to the Nigerian economy that crucial aspects of this sector such as exploration, production, gas utilization, conservation, petroleum policy and legislation are sensitive economic issues. Consequently in this research, attempt is made to model and forecast future values for Nigerian monthly Crude oil Exportation data between January 2002 and December 2011 using Seasonal Autoregressive Integrated Moving Average (SARIMA) models.

2. Literature Review

Many economic time series are seasonal. Its volatility notwithstanding, Nigerian monthly crude oil export series tends to exhibit some seasonality. Box and Jenkins [5], Madsen [6] and Boubake [7] are a few of authors that have written extensively on seasonal ARIMA models which are specially articulated for seasonal time series (Ette 2013). However, literature on forecasting crude oil exportation is scarce.

A SARIMA model was proposed by Box and Jenkins [5] specifically for series that are seasonal in nature. A few other authors that have written extensively on such models are Madsen [6], Priestley [8] and Saz [9]. Ette [10] used a technical approach to model Nigerian Monthly Crude Oil Prices using seasonal ARIMA model for the period of 2004 to 2011. The study revealed that crude oil price has a slightly positive and not easily discernible seasonality exhibiting a peak in 2008 and a depression in early 2009. His good work expatiated that seasonal difference once produced a series with SARIMA (0, 1, 1)x(1, 1, 1)$_{12}$ as the optimum model.

Ette and Eberechi [11] studied the Nigerian Crude Oil Production using multiplicative SARIMA from January 2006 – August 2012. The time plot reveals a negative trend between 2006 and 2009 and a positive trend from 2009 to 2012. Twelve-month differencing yields a series with significant spikes of the autocorrelation function at lags 1 and 12 suggests an autocorrelation structure of a (0, 1, 1) x(0, 1, 1)$_{12}$ SARIMA model which he used and found to be adequate. Other studies concentrated on modelling crude oil market volatility. For example Vo [12] proposed a hybrid model based on combination of Markov Switching (MS) model and stochastic volatility model (SV), called MSSV to predict volatility of crude oil short-term price. The author concluded that there is strong evidence to support regime-switching in the oil market, and the MSSV model was superior to both SV and MS in forecasting the volatility of oil market for out-of-sample. It is noted that, in-sample forecasting, the MS model was superior to the SV and MSSV.

3. Methodology

Box and Jenkins [5] gave Seasonal Autoregressive Integrated Moving Average (SARIMA) of a time series model as:

$$\Phi_P(\mathrm{B}^s)\phi(\mathrm{B})\nabla_s^D\nabla^d Z_t = \alpha_t + \Theta_Q(\mathrm{B}^s)\theta(\mathrm{B})\,\mathrm{a}_t \qquad (1)$$

This denotes an ARIMA (p, d, q) $\times$(P, D, Q)$_s$.

The SARIMA is explained thus:-

p = denotes the number of autoregressive terms.

d = is an integer which denotes the number of times the series must be differenced to attain stationarity.

q = denotes the number of moving average terms.

P = denotes the number of Seasonal autoregressive terms.

D = denotes the number of Seasonal differences required to attain stationarity.

Q = denotes the number of Seasonal moving average terms.

s = denotes the seasonal period or the length of the season.

The time series SARIMA model building is a selection of the appropriate model for the data in an iterative procedure based on the three fundamental steps of Box and Jenkins (1976). These procedures are:

(i) Model Identification,
(ii) Model Estimation and
(iii) Model Validation or Diagnostic Checking

3.1 Model Identification

At the model identification stage, our goal is to detect seasonality, if it exists, check for stationarity and to identify the order for the seasonal & non-seasonal autoregressive and seasonal & non-seasonal moving average terms. For Box-Jenkins models, it isn't necessary to remove seasonality before fitting the model. Instead, it can include the order of the seasonal terms in the model specification to the ARIMA estimation software

3.1.1 Testing for Stationarity

A test of stationarity (or non – stationarity) that has become widely popular over the past several years is the unit root test. This is the test that is used to carry out or to know the order of integration. It is important to know the order of integration of non-stationary variables, so they may be differenced before being included in a regression equation. The most common unit root tests are Augmented Dickey – Fuller (ADF) test [13], Kwiatkowski Phillip Schmidt Shin (KPSS) test [14], Phillips – Perron (PP) test and DF – GLS [15] test.

3.1.2 Choice of Best Model

Once stationarity and seasonality have been addressed, the next step is to identify the order (the p and q) of the autoregressive and moving average terms. The primary tools for doing this are the autocorrelation plot and the partial autocorrelation plot. However, Box and Jenkins [5] cited that the model should be parsimonious and therefore, recommended the need to use as few model parameters as possible so that the model fulfills all the diagnostic checks. Akaike [16] suggested a mathematical formulation of the parsimony criteria of the model building as information criteria for the purpose of selecting an optional model fits to a given data. The model that gives the minimum AIC, Schwartz Information Criterion (SBIC) [17], and Hannan-Quinn Information Criterion (HQC) [18] is selected as a parsimonious model.

3.2 Model Estimation

After an optimum model has been identified, the model estimation methods make it possible to estimate simultaneously all the parameters of the process, the order of integration coefficient and the parameters of an ARMA structure. The estimator of the exact maximum likelihood proposed by Sowell [19] is the vector $\hat{\beta} = \left(\hat{d}, \hat{\phi}', \hat{\theta}'\right)'$ which maximizes the log – likelihood function $L(\beta)$

$$L(\beta) = -\left(\frac{n}{2}\right)\ln(2\pi) - \left(\frac{1}{2}\right)\ln(R) - \left(\frac{1}{2}\right)x'R^{-1}x \quad (2)$$

Where R is the variance – covariance matrix of the process. The matrix R is a complicated algebraic expression and it is difficult to calculate. We therefore use methods based on an approximation of the likelihood function. The two main techniques that are available for the spectral approximation are that of Fox and Taqqu [20] and the minimization of the conditional sum of squared residuals. Asymptotically, these two methods converge on the exact maximum likelihood estimator. The estimator suggested by Fox and Taqqu [20] is the vector $\hat{\beta}^{SA}$ which maximizes the following expression

$$L^{\alpha}(\beta) = \sum_{j=1}^{n-1}\left[\ln\left(2\pi f_x(\omega_j)\right) + \frac{I_T(\omega_j)}{f_x(\omega_j)}\right] \quad (3)$$

This expression is easier to use but can display a basis in small samples.

The statistical/econometric software Gretl which was used for this work utilised the Exact Maximum Likelihood Estimation.

3.3 Model Diagnosis

After estimation of the model, the Box – Jenkins model building strategy entails a diagnosis of the adequacy of the model. More specifically, it is necessary to ascertain in what way the model is adequate and in what way it is inadequate. This stage of the modelling strategy involves several steps [21].

A good way to check the model adequacy of an overall Box – Jenkins model is to analyze the residual obtained from the model. The statistics have suggested determining whether the first K sample autocorrelation indicate the adequacy of the model and they are the Box – Pierce statistics and the Ljung – Box statistics (Portmanteau test). In spite of this, we can also check the model adequacy by examining the sample autocorrelation function of the residual (ACF) and the sample partial autocorrelation function of the residual (PACF). We can conclude that the model is adequate if there are no spikes in the ACF and PACF. We can also employ the Jargue – Bera test for non-normality of residual.

4. Empirical Result

Figure (1) shows the time plot of the series which indicates some seasonal fluctuations as well as trend. Thus, the data have to be seasonally adjusted, using the moving average method.

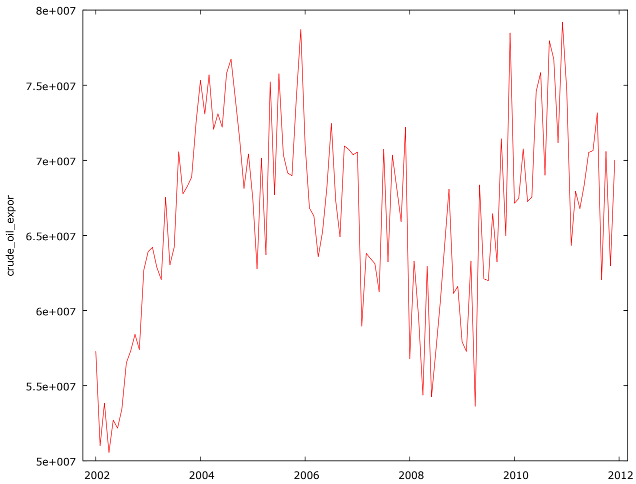

Figure 1: Time plot of monthly crude oil exportation during the period of 2002-2011

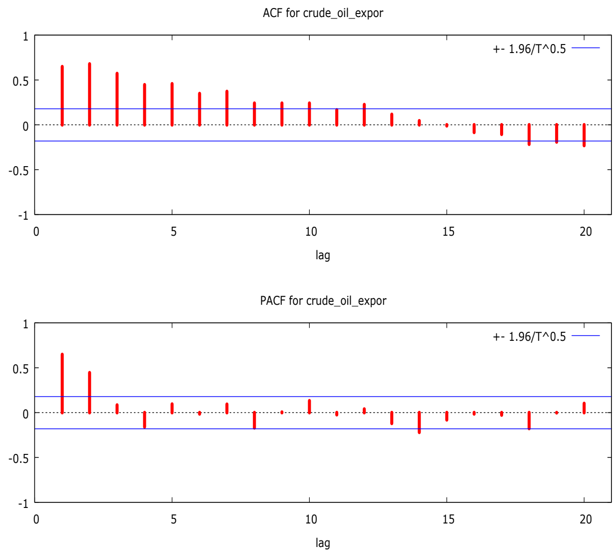

Figure 2: Correlogram of monthly crude oil Exportation

Table 1: Result of Stationarity Test Using Augmented Dickey-Fuller Statistics

	Original data		*First Difference*	
ADF	Test Statistics	P-Value	Test Statistics	P-Value
RWC	-1.53162	0.118	-2.47865	0.01277
RWCT	-2.20908	-2.93(5%), -3.46(1%)	-4.23746	-2.93(5%), -3.46(1%)

The first step in developing a Box-Jenkins model is to determine if the series is stationary. For this, we use the autocorrelation function (ACF) & partial autocorrelation function (PACF) in figure 2 and Augmented Dickey-Fuller test (ADF) in Table 1 above. Because the autocorrelation (ACF) start high and decline slowly, then series is nonstationary, and should be differenced. The ADF test statistics at each assumption respectively were greater than the critical values at level which reveals the fact that the null hypothesis is accepted. Thus, the original data (series) has unit root and it is non-stationary. However, it becomes stationary at first order differencing (see figure 3 and Table 1). Hence, the series is I (1).

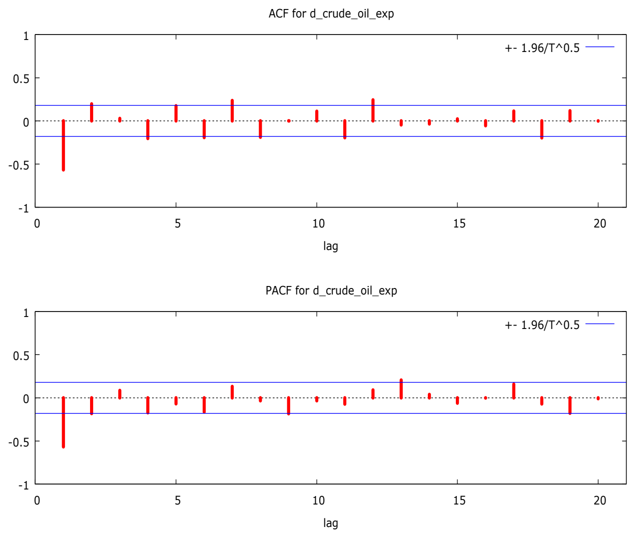

Figure 3: Correlogram of monthly crude oil exportation after

1st difference

The ACF and PACF show an ACF with significant spikes at lags 1, 7 and 12 (Figure 3). The spike at lag 12 shows that the series (monthly oil exportation) at first differencing is seasonal of period 12 and that a seasonal moving average & autoregressive term is present. Moreover, the PACF has significant spike at lag 1 and 13. Consequently, the seasonal difference on the stationary series was taken (with I value at order 1) to compute various SARIMA model. The best selected models were selected base on the smallest AIC, SBIC and HQC. Thus, ten (10) candidate models were chosen out of the 64 possible model combinations, using p = 1, 2; q = 1, 2; P = 1, 2; Q = 1, 2 and D=d=1. These are presented in Table 2. Out of these models, a parsimonious (best) model is obtained using the lowest information criteria as SARIMA (1, 1, 1) x (0, 1, 1)$_{12}$. The estimates of the model are presented in Table 3.

Table 2: Results of several models for model identification

SARIMA (p, d, q)x(P, D, Q)	*AIC*	*SBIC*	*HQC*
(0, 1, 1)x(1, 1, 1)$_{12}$	3586.192	3599.556	3591.609
(1, 1, 1)x(0, 1, 1)$_{12}$+	3584.827+	3598.191+	3590.245+
(1, 1, 1)x(1, 1, 1)$_{12}$	3584.863	3600.900	3591.364
(0, 1, 2)x(1, 1, 1)$_{12}$	3585.268	3601.305	3591.769
(1, 1, 1)x(1, 1, 2)$_{12}$	3586.804	3605.514	3594.389
(2, 1, 0)x(0, 1, 1)$_{12}$	3584.911	3598.275	3590.329
(2, 1, 0)x(1, 1, 1)$_{12}$	3584.953	3600.490	3590.954
(2, 1, 0)x(0, 1, 2)$_{12}$	3585.199	3601.236	3591.700
(2, 1, 0)x(1, 1, 2)$_{12}$	3586.417	3605.127	3594.002
(2, 1, 1)x(1, 1, 1)$_{12}$	3586.435	3605.145	3594.019

Note:
+ = Best model;
AIC = Alkaike information criteria;
SBIC = Schwarz Bayesian information criteria;
HQC = Hannan-Quinn information criteria

Base on the selection criteria AIC, BIC and HQC, the above table shows that SARIMA (1, 1, 1)x(0, 1, 1)$_{12}$ was selected to be the best model. Hence, Table 3 presents the model estimates.

Table 3: Results of Exact ML estimator with SARIMA

Model Estimator	*Coeff*	*Std. error*	*P-value*
Const	-31894.7	64224.7	0.6195
phi_1	-0.264311	0.15793	0.0942 *
theta_1	-0.389570	0.153842	0.0113 **
Theta_1	-0.908220	0.26828	0.0007 ***

Evaluations of gradient: 15
Estimated using Kalman filter (exact ML)
Dependent variable: (1-L) (1-Ls) crude_oil_export
Standard errors based on Hessian

Mean dependent variable -148007.4
S.D. dependent variable 6700914
Mean of innovations -401217.3
S.D. of innovations 3973497
Log-likelihood -1787.414
Akaike criterion 3584.827
Schwarz criterion 3598.191
Hannan-Quinn 3590.245

NOTE:
*, **, *** for 1%, 5% and 10% α-level respectively

Table 4: Residual Autocorrelation Function

LAG	*ACF*	*PACF*	*Q-stat*	*P-value*
1	0.0049	0.0049	0.0026	0.959
2	-0.0150	-0.0150	0.0277	0.986
3	-0.0541	-0.0540	0.3562	0.949
4	-0.1208	-0.1209	2.0077	0.734
5	-0.0395	-0.0417	2.1865	0.823
6	0.1058	0.1007	3.4799	0.747
7	0.1512	0.1420	6.1448	0.523
8	-0.0677	-0.0853	6.6847	0.571
9	-0.0408	-0.0429	6.8823	0.649
10	-0.0633	-0.0301	7.3643	0.691
11	0.0549	0.0977	7.7310	0.737
12	-0.0312	-0.0523	7.8509	0.797
13	0.1912	0.1499	12.3868	0.496
14	-0.0981	-0.1257	13.5946	0.480
15	-0.0158	0.0328	13.6263	0.554
16	0.1173	0.1492	15.3887	0.496
17	-0.0145	0.0109	15.4161	0.566
18	-0.0422	-0.0907	15.6494	0.617
19	-0.1196	-0.1516	17.5452	0.553
20	0.0260	0.0394	17.6356	0.611

Test for null hypothesis of normal distribution:
Chi-Square (2) = 0.054 with p-value 0.97344

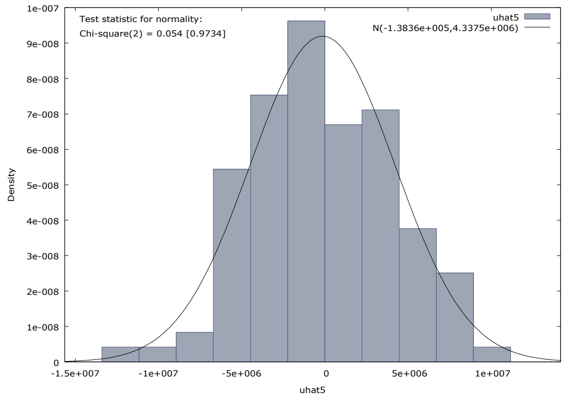

Figure 3: Histogram of Residual

From Table 3, the coefficient of SARIMA (1, 1, 1) x (0, 1, $1)_{12}$ model were valid and stationary condition was met and satisfied since the coefficients were all less than one (-0.264311, -0.389570,-0.908220) and are also significant since their P– value are less than 0.1, 0.05 and at 0.01 respectively. These means that the overall significance of the coefficients of SARIMA (1, 1, 1) x (0, 1, $1)_{12}$ are significant and hence AR (1), MA (1) and SAR (1) thus explain the series.

Further model accuracy was reported in Table 4, where the ACF and the PACF of the error were presented. These reports indicate that the errors are normal distributed (white noise, independent of time in essence they are random. From both ACF and the PACF, their values at lag 1 up to lag 20 hovers around the zero line, this makes the model valid and adequate. Also concentrating on Q-stat p –value from lag 1 up to lag 20, each P –value were greater than the exact p–value (0.05) which indicates that from lag 1 to lag 20 the hypothesis of (no serial correlation) was not rejected. Also, since the P-value (0.97344) of the Jacque-Berra Normality Test was greater than 0.05, we accept the null hypothesis and conclude that the residual is normally distributed.

5. Conclusion

This research aims to identify a time series model forecast for Crude Oil Export (barrels) in Nigeria for the period of January 2002 to December 2011 through the use of Box Jenkins fundamental approach. The modelling cycle was in three stages, the first stage was model identification stage, where the series was not non- stationary at level form base on the result provided by ADF test, correlogram and time plot. It was found out that the series was stationary at the 1^{st} difference. More so, seasonal difference was made due to significant spikes at lag 1, 7 and 12 of the stationary series I (1). Base on the selection criteria AIC, SBIC and HQC, reports show that SARIMA (1, 1, 1) x (0, 1, $1)_{12}$ was selected to be the best model to fit the data. The second stage was the model estimation, where the parameters conforms to the stationary conditions (less than one) and finally the third stage was model diagnosis where the errors derived from the model (1, 1, 1)x(0, 1, $1)_{12}$ was normally distributed, random (no time dependence) and no presence of error serial correlation. An out sample forecast for period of 24 months term was made, and this shows that the crude oil export will continue to be unstable for the period forecasted.

References

[1] Amnesty International Nigeria: Oil, Poverty and violence. Retrieved 9 May 2007, 2006.
[2] S.W. Petters “The Petroleum Sector in Nigeria”- An article on onlinenigeria.com, 2012.
[3] U.S EIA. “Nigeria Country Analysis Brief”, U.S Energy Information Administration (U.SEIA), 2007.
[4] Businessdayonline.com/2013/04/solid-mineral-crude- oil-grow-nigerias-2012-export-value-by-15-5
[5] G.E.P. Box and G.M. Jenkins. “Time series analysis, forecasting and control, holden-day, San Francisco”, 1976.
[6] H. Madsen. Time Series Analysis, Chapman & Hall/CRC, London, 2008.
[7] H.B.H Boubaker, “The forecasting performance of seasonal and nonlinear models” Asian Economic and Financial Review 1(1): 26-39, 2011.
[8] M. B. Priestley. “Spectral Analysis and Time Series”, Academic Press, London, 1981.
[9] G. Saz, “The Efficacy of SARIMA Models for Forecasting Inflation Rates in Developing Countries”: The Case for Turkey. International Research Journal of Finance and Economics, 62, 111 – 142, 2011.
[10] E.H Etuk,., (2013). Seasonal Arima Modelling Of Nigerian Monthly Crude Oil Prices. Asian Economic and Financial Review, 3(3):333-340, 2013.
[11] E.H. Ette and H.A. Eberechi.”Multiplicative Sarima Modelling Of Nigerian Monthly Crude Oil Domestic Production”. Journal of Applied Mathematics & Bioinformatics, vol.3, no.3, 2013, 103-112, 2013.
[12] M. T. VO, "Regime-switching stochastic volatility: Evidence from the crude oil Market." Energy Economics 31, 3131 31(5): 779-788, 2009.
[13] Dickey DA, FullerWA. “Distributions of the Estimators for Autoregressive Time Series with a Unit Root." Journal of the American Statistical Association” 74(366), 427-431, 1979.
[14] M.S. Pandit, and S. Wu, Time series and System analysis with applications. Wiley Series, New York, 1983.
[15] P. Perron, Z . Qu “A Simple Modication to improve the Finite Sample Properties of Ng and Perron's Unit Root Tests." Economics Letters, 94(1), 12-19, 2007.
[16] D.A. Dickey, W. A Fuller. \Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." Econometrica, 49(4), 1057-1072, 1981.
[17] G. Schwarz “Estimating the Dimension of a Model." The Annals of Statistics, 6(2), 461- 464, 1978.
[18] E. J. Hannan, B.G. Quinn. “The Determination of the Order of an Auto regression." Journal of the Royal Statistical Society B, 41(2), 190- 195, 1979.
[19] F.B. Sowell ‘Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models’, Journal of Econometrics, 53, pp. 165 – 188, 1992.
[20] R. Fox, and M.S. Taqqu, ‘Large sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series’, Annals of Statistics, 14, PP. 517 – 532, 1986.
[21] Kendall, M., and Ord, K. “Time Series Analysis”, 3rd Ed. New York: Oxford University Press, pp. 110, 1990.