Journal of Probability and Statistical Science
22(1), 1-15 Sep. 2024

Friedman Test Technique for Optimizing a Seasonal Box-Jenkins ARIMA Model Building

Elisha J. Inyang
Department of Statistics, University of Uyo

Imoh U. Moffat
Department of Statistics, University of Uyo

Ettebong P. Clement
Department of Statistics, University of Uyo

ABSTRACT

This study employs the Box-Jenkins methodology for time series modelling to analyze Nigerian crude oil production data. To enhance the model-building process, the Friedman rank test for seasonality was utilized. Data were sourced from the Nigerian Petroleum Corporation's Annual Statistical Bulletin, covering the period from January 1997 to September 2014. The findings indicate that the yearly and monthly mean distributions exhibit non-constant behavior over time, implying that both natural and governmental factors significantly impact crude oil production (COP) dynamics. Consequently, a seasonal $ARIMA(0,1,1)(0,1,1)_{12}$ model was fitted and deemed appropriate for the data.

***Keywords**:* Box-Jenkins Models, Time Series Modelling, Friedman Rank Test, Seasonality Test.

1. Introduction

Since crude oil and natural gas account for approximately two-thirds of global energy consumption, crude oil is becoming more and more important to the global economy. Crude oil is the most traded and largest commodity globally, making up around ten percent of global trade, and its global consumption exceeds five hundred billion US dollars, roughly ten percent of the US GDP. International trading in crude oil involves a wide range of participants, including enterprises, governments, private refineries, countries that export and import oil, and also oil speculators. The price of crude oil is mostly determined by the availability and demand for the commodity, while it is also greatly impacted by a wide range of unpredictable past, present, and future events, including weather, stock levels, GDP growth, political issues, and public expectations. Moreover, oil prices differ globally since moving crude oil from one nation to another takes a long time. Due to these factors, the market is very dynamic and fluctuates, and the underlying process causing this complexity is unknown. In a country that relies heavily on petroleum, where the sale of the enormous reserves of natural gas and oil provides almost all of the government's income. Nigeria is Africa's leading oil producer, producing more than 2.5 million barrels of crude oil per day, ranking among the top 10 in the world. However, the country's residents have been living through a continuous energy crisis due to a lack of petroleum products, unstable gas prices, and a shortage of gasoline, all of which have a detrimental impact on the national economy.

In actuality, fuel shortages have been a defining feature of Nigeria's history, particularly throughout the military regime of the 1908s and 1990s and even during the democratic administrations of 2003, 2005, 2008, 2012 – 2013, and most recently 2022 – 2023. Government mismanagement and corruption that have led to mistrust are blamed by the public on the government, while gasoline middlemen are blamed by the government for hoarding. Citizens have experienced a rolling energy crisis for several months between 2022 and 2023. Every gas station

- Received February 2024, in final form August 2024.
- Elisha J. Inyang (corresponding author), Imoh U. Moffat and Ettebong P. Clement are affiliated with the Department of Statistics, University of Uyo.
inyang.elisha@yahoo.com

in the country has seen variations of this scene. Nigerians continue to experience fuel shortages in the shadow of plenty, despite all of the government's efforts to find a solution.

2. Literature Review

Time series provides useful tools that help predict the future by approximating models that use past data, and these time series models have advantages over other statistical models in certain situations. They can be easily used for forecasting purposes because historical sequences of observations are readily available from published secondary sources, and these successive observations are statistically dependent. In such situations, time series models are a boon for forecasters. Many researchers have employed techniques developed by [3] in their work.

Nigerian crude oil production was studied by Omekara *et al.* [27] using the Box-Jenkins ARIMA model. Based on their findings, a SARIMA$(1,1,1)(0,1,1)_{12}$ model was determined to be appropriate for the dataset. The work of Fatoki *et al.* [10] concentrated on using ARIMA modelling tools to analyze Nigeria's crude oil production from 1980 to 2013. The series was well-fit using the ARIMA(1,2,2) model. To develop an appropriate time series model for the monthly crude oil output in Nigeria, Sadeeq and Ahmadu [28] fitted ARIMA$(2,1,0)(2,1,1)_{12}$ to the series, which covered the years 2002 – 2016. Etaga *et al.* [8] fitted a time series model to the production and export of crude oil from Nigeria between January 1999 and December 2015. The best models for producing and exporting crude oil were SARIMA$(1,0,1)(2,0,0)_{12}$ and SARIMA$(2,1,0)(1,0,1)_{12}$, respectively, according to their findings. Taiwo *et al.* [33] examined and discussed the volatility and swings in the monthly crude oil price in Nigeria using time series analysis. The fitted ARIMA(5,1,2) model was found to be adequate for the given data. Ijeoma [15] thought about applying the Box-Jenkins approach to examine Nigeria's monthly crude oil prices from January 2000 to December 2013. The results showed that the ARIMA(1,1,1) model was appropriate for the dataset.

Similarly, a time series model is fitted by Clement [5] to the chemical viscosity reading data. The initial models fitted to the same data points by Box-Jenkins are compared with the results using the normalized Bayesian Information Criterion (BIC). Consequently, the proposed model performed better than the Box-Jenkins models. Again, using data on the exchange rate between the US dollar and the Nigerian naira, Clement [6] constructs a statistical time series model. It was determined that an ARIMA(0,1,1) would suffice for the dataset. Shittu and Inyang [31] used the ARIMA-Intervention model to simulate Nigerian monthly crude oil prices to compare the outcome with the intervention model's utilizing a lag operator. Their findings showed that the ARIMA-Intervention model outperformed the alternative model. Etuk *et al.* [9] examined the effect of the declaration of cooperation (Doc) on the production of crude oil (COP) in Nigeria. The results showed that the DoC's 35% level production cut had a negative impact on COP. Moffat and Inyangs [23] investigated the impact of the Nigerian government amnesty programme on her crude oil production. The result disclosed that the Humanitarian Initiative's assistance had no effect on the output of crude oil.

Inyang *et al.* [18] used the Box-Tiao method to investigate how international oil politics affected the price of Nigerian crude. The Organization of Petroleum Exporting Countries' intervention in December 2016 had a notable and abrupt effect on the price of Nigerian oil upon its introduction, as seen by the corresponding 33.72% increase in price. Again, a time series intervention model based on the ESM and ARIMA Models was used by Inyang *et al.* [16] to model the daily exchange rates between the Pakistani rupee and the Nigerian naira. The comparison of the intervention model with ESM and the ARIMA-Intervention model showed that the latter performed better.

Inyang *et al*. [17] evaluated the response of the comparative value of the Bangladesh Taka to the Naira owing to the 2016 financial crisis using an intervention model based on ETS plus ARIMA models. Their results revealed that the intervention caused a 68.49% depreciation in the value of the Naira exchanged with the Bangladesh Taka in the exchange rate market, with a decay rate of 0.6. In light of this, this work seeks to investigate the trend pattern and provide a framework for determining the rate of crude oil production (COP) in Nigeria by fitting an appropriate time series model.

3. Materials and Method

3.1 Data Description

The monthly crude oil output for the period of January 1997 to September 2014 is the secondary data set utilized for this analysis. A total of 213 observations were extracted from the Nigeria National Petroleum Corporation (NNPC) Annual Statistical Bulletin [24]. The statistical package used for the analysis of this work is the R language (R-4.1.3-win) [32].

3.2 Model Specification

The Box-Jenkins ARIMA Procedure:
The modeling technique used, Box-Jenkins [2-4], is restricted to stationary series:
(a). Stationary: Making the series stationary is thus the initial move in the analysis.
Time Series Plot: A constant variance and mean suggest stationary behavior.
Trend Stationary: When trend and slope are eliminated, a stationary series is what remains.
Difference Stationarity: If the series is transformed, it ends up being stationary.
(b). Model Identification: ACF and PACF are the identification tools. After determining the stationary pattern of the series, one may identify the structure of the ARMIA(p,d,q) model for the process by examining the correlogram plots.
(c). Model Estimation: Calculate the models' estimated parameters. Here, the analysis is done using R software.
(d). Diagnostic Checking: This process is known as model validation. Used to select a better model among competing models.

3.2.1 Autoregressive Moving Average Process [ARMA(p,q)]

Give an autoregressive process of order p and a moving average process of order q, we have:

$$F_t = \partial_1 F_{t-1} + \cdots + \partial_p F_{t-p} + \varpi_t + \wp_1 \varpi_{t-1} + \cdots + \wp_q \varpi_{t-q} \qquad (1)$$

With shift operator Ω defined as:

$$\Omega^0 \equiv 1 \text{ }, \Omega F_t = F_{t-1}, \text{ } \Omega^k F_t = F_{t-k}, k = 1,2 \ldots \qquad (2)$$

Applying (2), (1) reduces to:

$$\left(1 - \partial_1 \Omega - \partial_2 \Omega^2 - \cdots - \partial_p \Omega^p\right) F_t = \left(1 + \wp_1 \Omega + \wp_2 \Omega^2 + \cdots + \wp_q \Omega^q\right) \varpi_t$$

$$\partial(\Omega) F_t = \wp(\Omega) \varpi_t \qquad (3)$$

Where:

$$\partial(\Omega) = 1 - \partial_1 \Omega - \partial_2 \Omega^2 - \cdots - \partial_p \Omega^p$$

$\wp(\Omega) = 1 + \wp_1\Omega + \wp_2\Omega^2 + \cdots + \wp_q\Omega^q$

F_t is the COP dataset at time t; $\partial's$ and $\wp's$ are the AR and MA parameters; ϖ_t is the white noise term and assumed to have these properties:

$E(\varpi_t) = 0, \text{Cov}(\varpi_t, \varpi_t) = Var(\varpi_t) = \sigma^2,$
$Cov(F_t, \varpi_{t+k}) = 0,\ k \neq 0,\ Cov(F_t, \varpi_{t+k-p}) = 0,\ k > p$

Nonstationary Models

Most time series that are seen in real life show non-stationary conduct, which can be attributed to seasonal fluctuation or trend, which is a shift in the local mean. The method to model a non-stationary time series is to reduce such a process to a stationary one and then model using what we know concerning the stationary process.

Differencing

Differencing is a form of adjustment in time series when the series has been found to contain components such as trend, the regular differencing operator is introduced. The regular (non-seasonal) difference of order one is given by

$\nabla F_t = F_t - F_{t-1} = (1 - \Omega)F_t$ (4)

First differencing may be stationary depending on the nature of the trend. Sometimes higher order differencing may be required.

In general,

$\nabla^{\mathrm{d}} F_t = \nabla\left(\nabla^{\mathrm{d}-1} F_t\right) = (1 - \Omega)^d F_t$ (5)

Where d is the order of differencing

At times non-stationarity may be due to seasonal variation; to reduce such series to a stationary one, seasonal differencing could be applied. The seasonal difference of order D with period S is given by

$\nabla^{\mathrm{D}}{}_{\mathrm{S}} F_t = \nabla_{\mathrm{S}}(\nabla^{\mathrm{D}-1} F_t)$ (6)

For instance, the first seasonal difference at period S is given by

$\nabla_s F_t = F_t - F_{t-s}$

For monthly seasonal data with period S=12

$\nabla_{12} F_t = F_t - F_{t-12}$

For quarterly data with period S=4

$\nabla_4 F_t = F_t - F_{t-4}$

When a polynomial trend series is differentiated, the degree of the polynomial is typically lowered by the level of the differencing. That is, level one differencing eliminates a linear trend, whereas level two differencing eliminates a quadratic trend.

3.2.2 Autoregressive Integrated Moving Average Process [ARIMA(p,d,q)]

Given an ARMA model with parameters ARMA(p,q) and the differencing operator in (5), the resulting model is given as ARIMA(p,d,q), written using (2) as

$$\partial(\Omega)\,\nabla^{d}F_t = \wp(\Omega)\epsilon_t \qquad (7)$$

3.2.3 Seasonal Autoregressive Integrated Moving Average Process [ARIMA(p,d,q)(P,D,Q)s]

The seasonal autoregressive integrated moving average process is an integrated series that is obtained after the seasonal components have been removed. It is written as

SARIMA(p,d,q). $(P,D,Q)_S$ (8)

whereas S denotes the order of seasonality, the uppercase letters stand for the model's seasonal components, and the lowercase letters stand for the non-seasonal components.

Generally, for the multiplicative seasonal model

$$\partial(\Omega)\,\text{Å}_{S}(\Omega^{S})\nabla^{D}{}_{S}\nabla^{d}F_t = \wp(\Omega)\mathbb{Q}_S(\Omega^{S})\varpi_t \qquad (9)$$

Where:

$\partial(\Omega) = 1 - \partial_1\Omega - \partial_2\Omega^2 - \cdots - V_p\Omega^p \equiv 1 - \sum_{i=1}^{p}\partial_i\Omega^i$

$\text{Å}_{S}(\Omega^{S}) = 1 - \text{Å}_{S,1}\Omega^{S} - \text{Å}_{S,2}\Omega^{2S} - \cdots - \text{Å}_{S,P}\Omega^{PS} \equiv 1 - \sum_{i=1}^{P}\text{Å}_{S,i}\Omega^{iS}$

$\wp(\Omega) = 1 + \wp_1\Omega + \wp_2\Omega^2 + \cdots + \wp_q\Omega^q \equiv \sum_{j=0}^{Q}\wp_j\Omega^q\,, \wp_0 = 1$

$\mathbb{Q}_S(\Omega^{S}) = 1 + \mathbb{Q}_{S,1}\Omega^{S} + \mathbb{Q}_{S,2}\Omega^{2S} + \cdots + \mathbb{Q}\Omega^{QS} \equiv \sum_{j=0}^{Q}\mathbb{Q}_{S,Q}\Omega^{jS}$

∇^d and $\nabla^D{}_S$ remained as defined in (5) and (6) respectively.

$\text{Å}_S's$ and $\mathbb{Q}_S's$ are parameters of the seasonal autoregressive and moving average process.

3.3 Unit Root Test

As a prerequisite for any further analysis in time series modeling, it is pertinent to formally diagnose the characteristics of the series that are used in the study [19].

The regression equation in (10) serves as the foundation for the Augmented Dickey-Fuller (ADF) [7] test.

$F_t = \partial F_{t-1} + \sum_{j=1}^{p-1}F_j\Delta F_{t-j} + \varpi_t$ (10)

Where F_t is the series being tested and the quantity of lag-differenced terms (p) is included to capture any autocorrelation.

Hypothesis:

$H_0: \beta = 0$ (series contains a unit root)

Against

$H_1: \beta \neq 0$

Test statistics are:

$\daleth_\rho = \frac{\hat{\partial}-1}{S.E(\hat{\partial})} \sim t_\propto(n)$ (11)

3.4 Seasonality Test

3.4.1 Friedman Test (Stable Seasonality Test)

The test statistic is constructed as follows. Considering the matrix of data $\{f_{ij}\}_{n\times k}$ with number of years in the sample (i.e. n rows) and the frequency of the data (i.e. k columns).

The data matrix is substituted by a new matrix $\{\ell_{ij}\}_{n\times k}$, such that ℓ_{ij} is the rank of f_{ij} within block i [11 – 14, 21, 26].

$$\mathcal{H} = \frac{\exists_t}{\mathrm{e}_\varpi} \qquad (12)$$

Where;

$$\exists_t = n\sum_{j=1}^{k}(\overline{\ell_{.J}} - \bar{\ell}) \text{ and } \mathrm{e}_\varpi = \frac{\sum_{i=1}^{n}\sum_{j=1}^{k}(\overline{\ell_{.J}} - \bar{\ell})^2}{n(k-1)} \qquad (13)$$

For large n $(n > 1`5)$ or k $(k > 4)$, the probability distribution of $\mathcal{H}$ can be approximated by that of a chi-square distribution. Therefore, the p-value is given by $P({f^2}_{k-1} > \mathcal{H})$

Hypothesis:
H_0: $\text{Ⅎ} = 0$ (no stable seasonality)
Vs
H_1: $\text{Ⅎ} \neq 0$

3.5 Model Validation

Diagnostic testing is a crucial stage in building time series models, and this consists of scrutinizing a variety of diagnostics to determine whether the selected model is healthy and hence ready to forecast. We consider here;

3.5.1 Plot of the residual ACF

Plotting the correlogram of the fitted model's residuals allows one to assess the goodness of fit soon after a suitable ARIMA model has been established. The residuals are white noise, suggesting that the model fits well, provided a majority of the autocorrelation coefficients are inside the bound of $\pm\frac{2}{\sqrt{T}}$, T is the total duration of the data points [20].

3.5.2 Akaike Information Criterion (AIC)

The AIC [1, 19-20, 29], is formulated as

$$AIC = M_\varsigma\left[1 + \frac{2P}{\varsigma - P}\right] \qquad (16)$$

Where:
M_ς = Index related to production error (known as residual sum of squares)
p = No. of parameters in the model, ς = No. of data points.

3.5.3 Bayesian Information Criterion (BIC)

Given a limited number of models, the BIC is a criterion for choosing a model. The model with the smallest BIC value among two or more estimated models should be chosen [5,19-20, 29]. It is given by:

$$\text{BIC} = nln\hat{\sigma}_\varpi^2 + kln(n) \qquad (17)$$

Where $\hat{\sigma}_\varpi^2$ is the estimated error variance defined by

$\hat{\sigma}^2_{\varpi} = \frac{1}{T}\sum_i^{\varsigma}(f_i - \bar{f})^2$

f = Observed data, ς = observations length, k = quantity of estimated parameters.

3.5.4 Ljung Box Test

Up to lag k, the Ljung Box Test can be used to determine if serial autocorrelation is present or absent [19-20]. We compute the statistic $\mho$ to perform the Ljung Box test [22]. Given a series f_t of length ς:

$$\mho(\mathcal{M}) = \varsigma(\varsigma+2)\sum_{j=1}^{\aleph}\frac{r_j^2}{\varsigma-j} \quad (18)$$

Where: r_j = accumulated sample autocorrelations, $\aleph$ = the time lag.

Hypothesis:
H_0: (residuals do not show any autocorrelation)
Against
H_1: (H_0 is false)

4. Results and Discussion

The dataset reflects the monthly Nigerian crude oil output from January 1997 to September 2014. Figure 1 displays the sequence plotted against time. The graph of the series does not show any unique pattern as it rises and falls at random, perhaps due to the mechanism that generated the dataset.

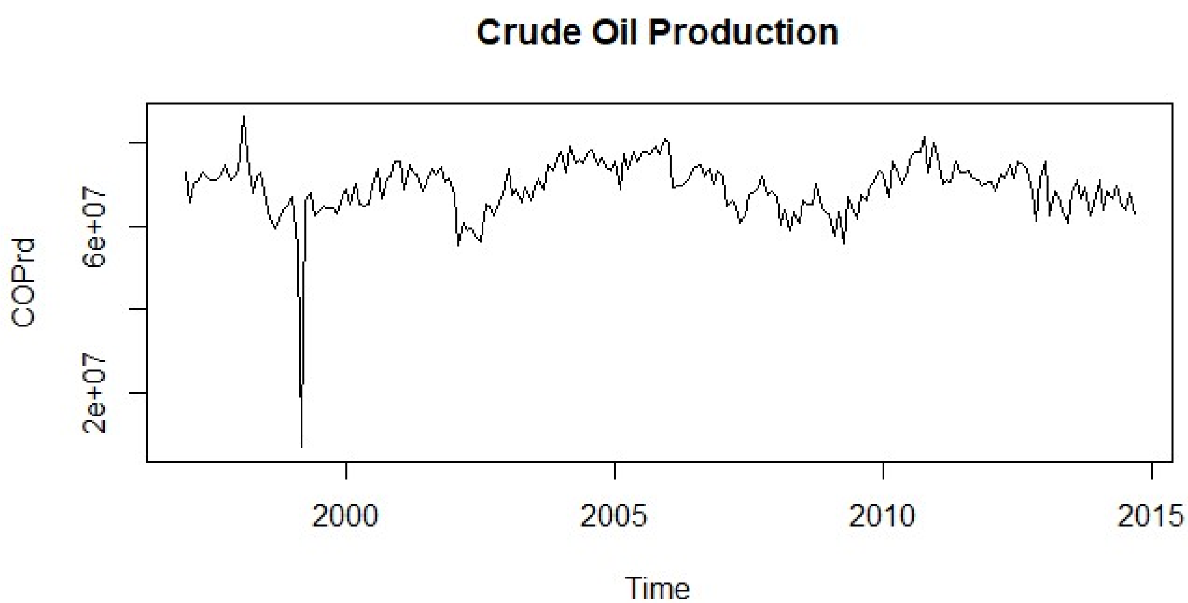

Figure 1. Time Plot of Crude Oil Production (COP)

Table 1. Monthly Mean of Crude Oil Production

S/N	Month	Monthly Total	Monthly Mean
1	January	1300265866	72236993
2	February	1188137444	66007636
3	March	1265356009	70297556
4	April	1212953595	67386311
5	May	1248069240	69337180
6	June	1215050295	67502794
7	July	1248316869	69350937
8	August	1277414026	70967446
9	September	1246098394	69227689
10	October	1213671337	71392432
11	November	1170111387	68830082
12	December	1213821003	71401235

Table 2. Yearly Mean of Crude Oil Production

S/N	Year	Yearly Total	Yearly Mean
1	1997	855721134	71310095
2	1998	826443999	68870333
3	1999	773677520	64473127
4	2000	828547638	69045637
5	2001	865173583	72097799
6	2002	740687180	61723932
7	2003	844150929	70345911
8	2004	910156486	75846374
9	2005	918660619	76555052
10	2006	869196506	72433042
11	2007	803000708	66916726
12	2008	768745932	64062161
13	2009	780347940	65028995
14	2010	896043406	74670284
15	2011	866245232	72187103
16	2012	852776655	71064721
17	2013	800488096	66707341
18	2014	599201906	66577990

To explore the effect of the mean and to check if the mean is a constant function of time or not, the monthly and yearly means of the series were computed, as are respectively shown in Tables 1 and 2. The highest monthly mean of 72236993 is observed in January, while the lowest monthly mean of 66007636 is observed in February. Similarly, the highest yearly mean of 76555052 was observed in 2005, while the least yearly mean of 61723932 was observed in 2002. Therefore, it is seen that both the monthly and yearly means are not constant functions of time and that government policies and programmes in the oil sector of the Nigerian economy, coupled with oil theft and natural phenomena, affect the distribution of crude oil production.

Model Identification

The graph of the original series in Figure 1 exhibits the characteristics of a non-stationary series, with peaks at intervals of equal length indicating the presence of a seasonal trend (also see Figure 4). These attributes are confirmed by both the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the series in Figures 2 and 3, respectively. The inability of both the ACF and the PACF to die out quickly at high lags shows that the series is not stationary, as confirmed by the unit root test at level in Table 3.

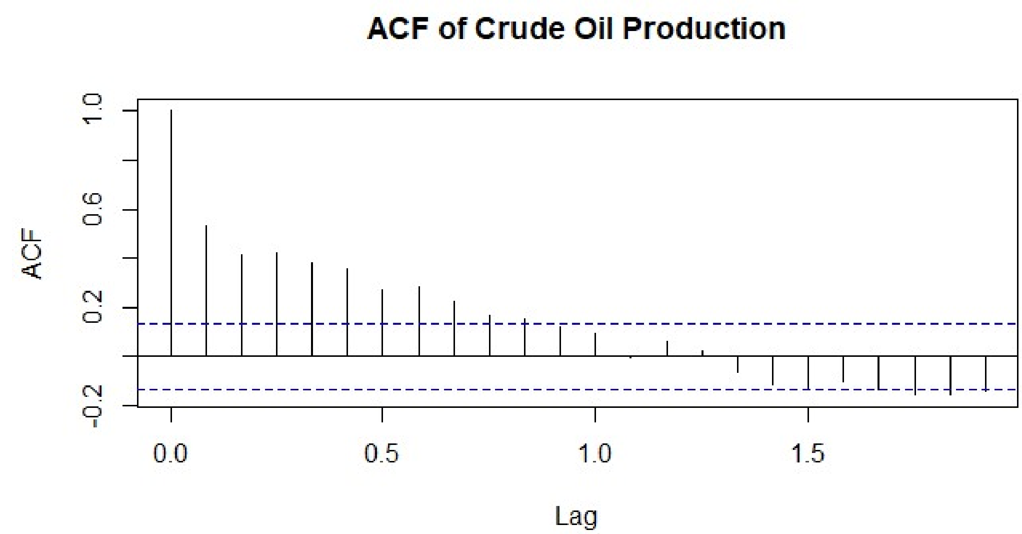

Figure 2. ACF of Crude Oil Production

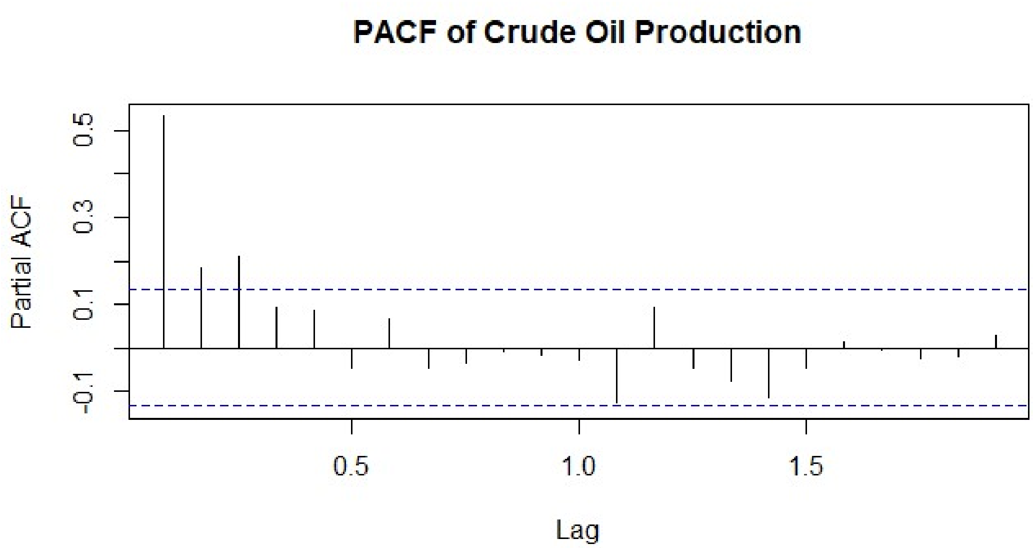

Figure 3. PACF of Crude Oil Production

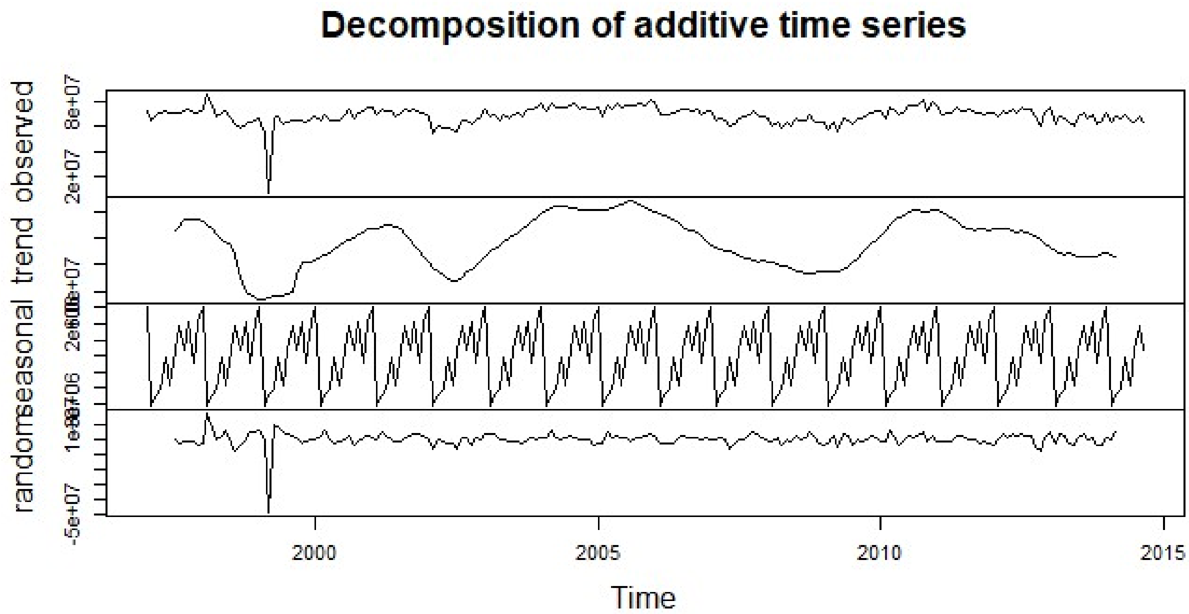

Figure 4. Time Series Decomposition of the COP

Table 3. Unit Root Test at Level

Test	Augmented Dickey-Fuller
Data	COP
Dickey-Fuller	-3.3075
Lag order	5
P-value	0.07127
Alternative hypothesis	Stationary

Table 4. Unit Root Test at First Difference

Test	Augmented Dickey-Fuller
Data	COP
Dickey-Fuller	-8.3079
Lag order	5
P-value	0.01
Alternative hypothesis	Stationary

Table 5. Seasonality Test before Seasonal Differencing

Test	Friedman rank
Data	COP
Test Statistic	82.96
P-value	3.93241e-13
Alternative hypothesis	Data is Seasonal

Table 6. Seasonality Test after Seasonal Differencing

Test	Friedman rank
Data	COP
Test Statistic	1.11
P-value	0.9999162
Alternative hypothesis	Data is Seasonal

To attain stationarity, the series was transformed. The series transformation difference method was used, and both regular and seasonal differencing were carried out. The graph of the differenced series in Figures 5 and 8 indicates that the series is stationary. The plots of the ACF and the PACF of the differenced series in Figures 6, 7, 9, and 10 confirmed that the series is stationary. Following the initial difference, a unit root test was performed in Table 4, and the results showed that the series was stationary.

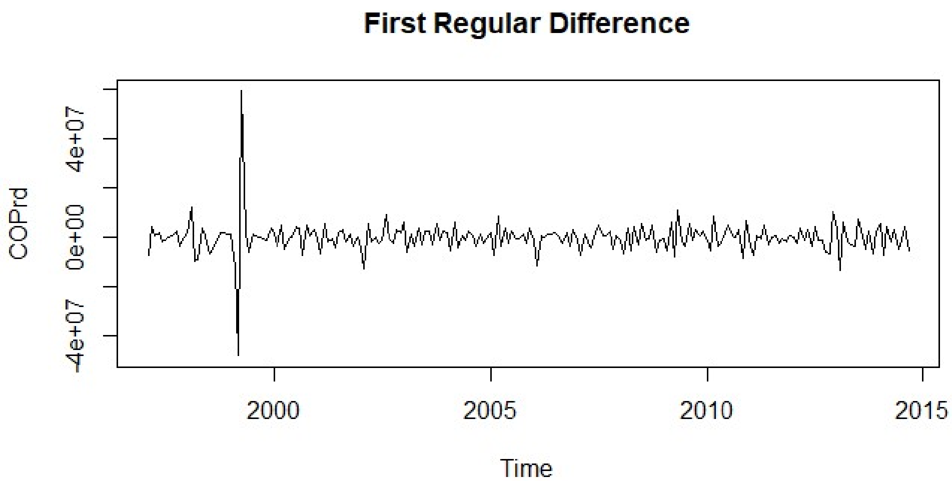

Figure 5. First Order Regular Differencing

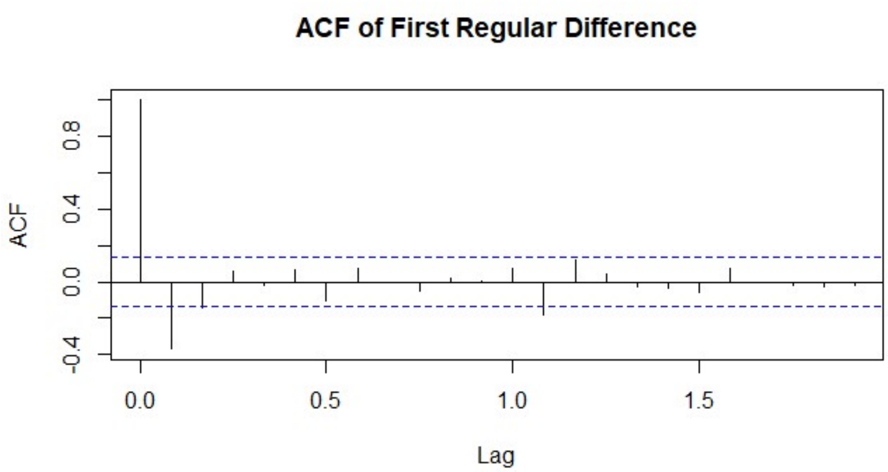

Figure 6. ACF of First Order Regular Differencing

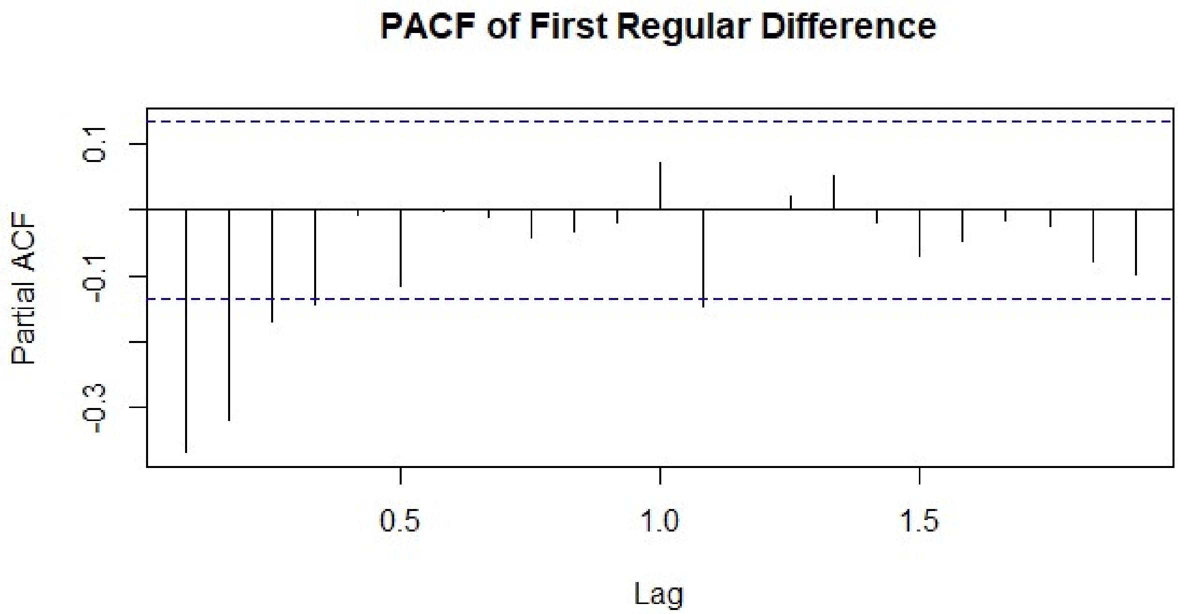

Figure 7. PACF of First Order Regular Differencing

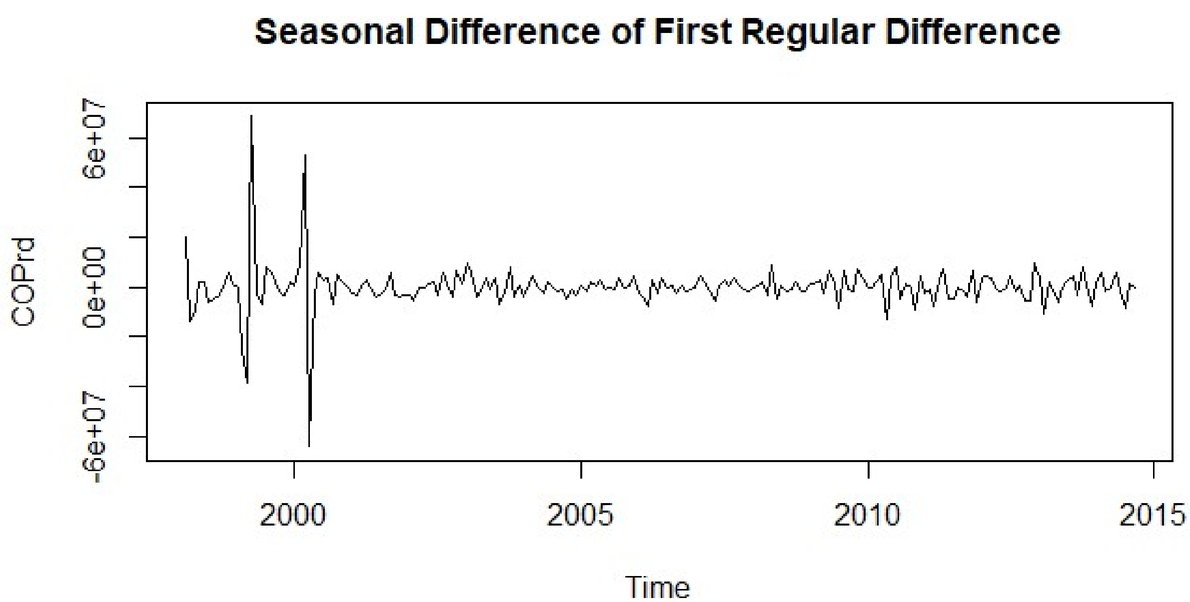

Figure 8. Seasonal Differencing of First Order Regular Difference

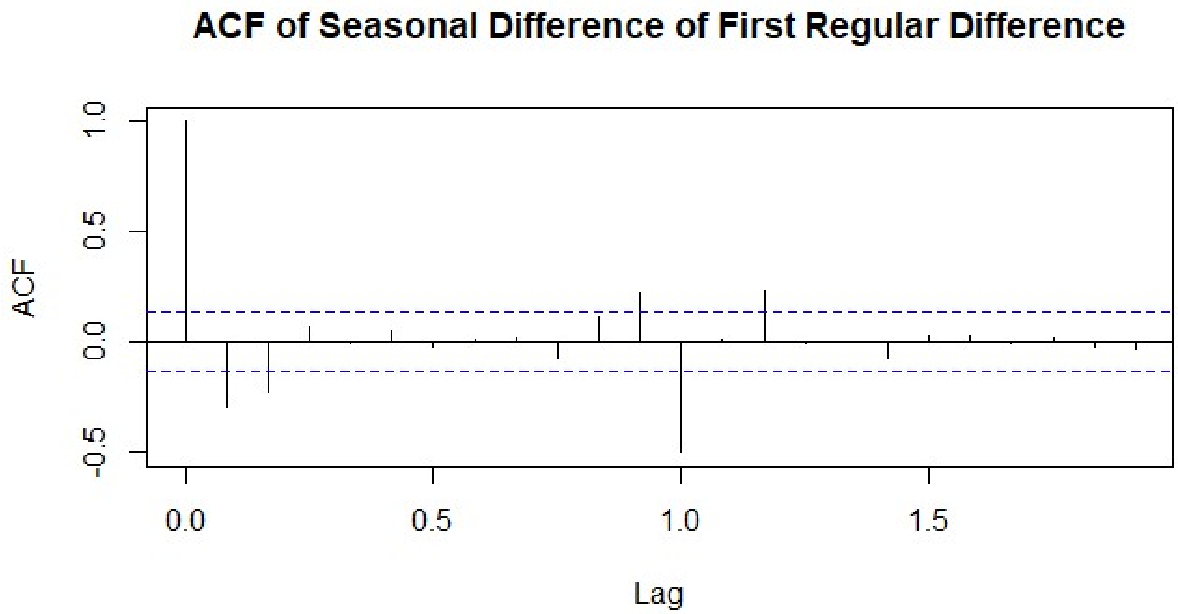

Figure 9. ACF of Seasonal Differencing of First Order Regular Difference

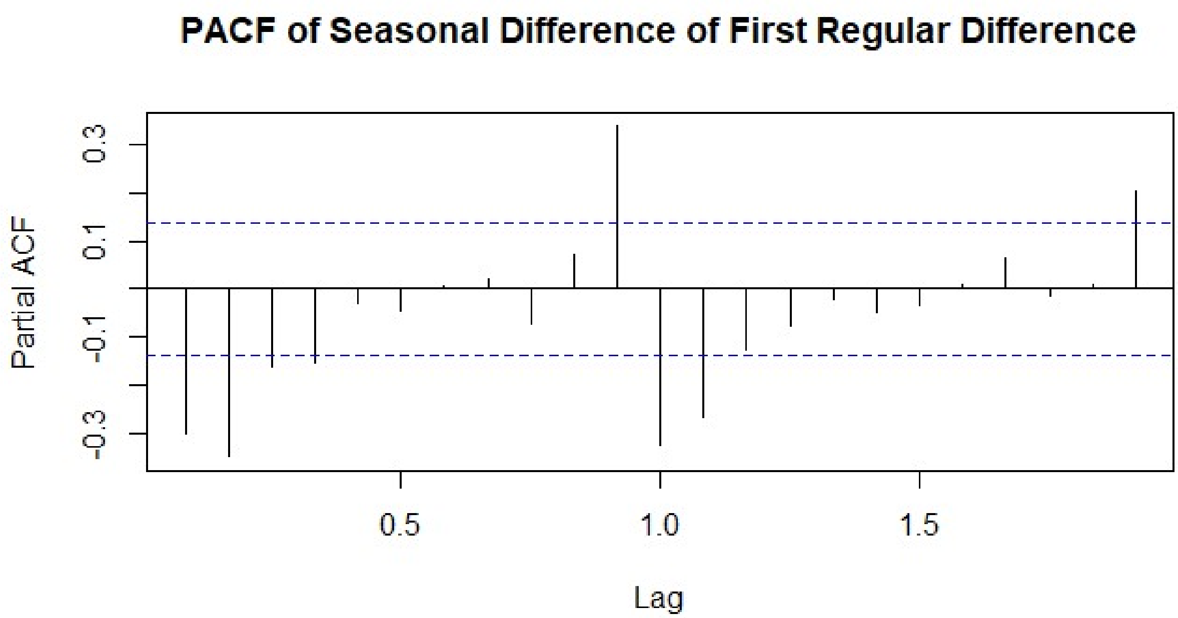

Figure 10. PACF of Seasonal Differencing of First Order Regular Difference

Because ACF and PACF indicated signs of seasonality since they repeat themselves at lags that are multiples of the number of periods per season (here the period = 12), and a Friedman seasonality test in Table 5 further confirmed that a seasonal effect is present (also, see Figure 4). A first-order seasonal differencing was done to remove the seasonal effects, and a seasonal test in

Table 6 confirmed that the seasonal effect was removed (also see Figure 11). Based on the correlogram in Figures 9 and 10, a mixed process is suggested.
Possible models identified for the COP dataset are:
$SARIMA(1,1,1)(1,1,1)_{12}$, $SARIMA(0,1,2)(0,1,2)_{12}$
$SARIMA(1,1,0)(1,1,0)_{12}$, $SARIMA(0,1,1)(0,1,1)_{12}$

Model Estimation

Determining the least squares estimates of the identified models' parameters is the next step after making a rough identification of what seems to be appropriate models for the series. Table 7 summarizes the estimated ARIMA models along with their statistics to fit the models to the data.

Table 7. Parameter Estimation for ARIMA(p,d,q)(P,D,Q)$_S$ Models

Parameters		Estimate	Std. Error	Z-value	Prob. Value
$(1,1,1)(1,1,1)_{12}$	∂_1	0.102895	0.104983	0.9801	0.3270
	$\wp_1$	-0.695992	0.072532	-9.5957	<2e-16 ***
	$\mathring{A}_1$	-0.110394	0.083553	-1.3212	0.1864
	$\mathbb{Q}_1$	-0.882415	0.077276	-11.4191	<2e-16 ***
$(0,1,2)(0,1,2)_{12}$	$\wp$	-0.579304	0.076445	-7.5780	3.508e-14 ***
	$\wp_2$	-0.083983	0.076011	-1.1049	0.2692
	$\mathbb{Q}_1$	-0.999473	0.078387	-12.7506	< 2.2e-16 ***
	$\mathbb{Q}_2$	0.122634	0.086617	1.4158	0.1568
$(1,1,0)(1,1,0)_{12}$	∂_1	-0.359280	0.066188	-5.4282	5.693e-08 ***
	$\mathring{A}_1$	-0.540229	0.058225	-9.2783	< 2.2e-16 ***
$(0,1,1)(0,1,1)_{12}$	$\wp_1$	-0.641675	0.056641	-11.3288	< 2.2e-16 ***
	$\mathbb{Q}_2$	-0.963015	0.159142	-6.0513	1.437e-09 ***

Diagnostic Check

After determining the model and estimating the parameters, diagnostic checking was carried out to see whether the fitted model was adequate. According to Table 8, the $ARIMA(0,1,1)(0,1,1)_{12}$ model has the smallest Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) values, at 6832.179 and 6822.284, respectively, out of the four estimated ARIMA models.

Table 8. Model Evaluation for ARIMA(p,d,q)(P,D,Q)$_S$ Models

Model	BIC	AIC
$(1,1,1)(1,1,1)_{12}$	6839.756	6823.264
$(0,1,2)(0,1,2)_{12}$	6839.08	6822.589
$(1,1,0)(1,1,0)_{12}$	6917.716	6907.821
$(0,1,1)(0,1,1)_{12}$	6832.179	6822.284

Table 9. Ljung-Box Test for ARIMA$(0,1,1)(0,1,1)_{12}$

Residuals from ARIMA$(0,1,1)(0,1,1)_{12}$
Q* = 17.407, df = 22, p-value = 0.7405 Model df: 2. Total lags used: 24

Figures 12, 13, and 14 demonstrate the adequacy of the fitted model by using the ACF and PACF residuals from the model to show that nearly all of the coefficients of both the ACF and PACF of the residuals are inside the significance bounds of $\pm\frac{2}{\sqrt{213}} = \pm 0.1370$. Examining the goodness of fit test once more, the model fits the data well, as shown by the Ljung-Box test statistic of 17.407 and p-value of 0.7405 in Table 9.

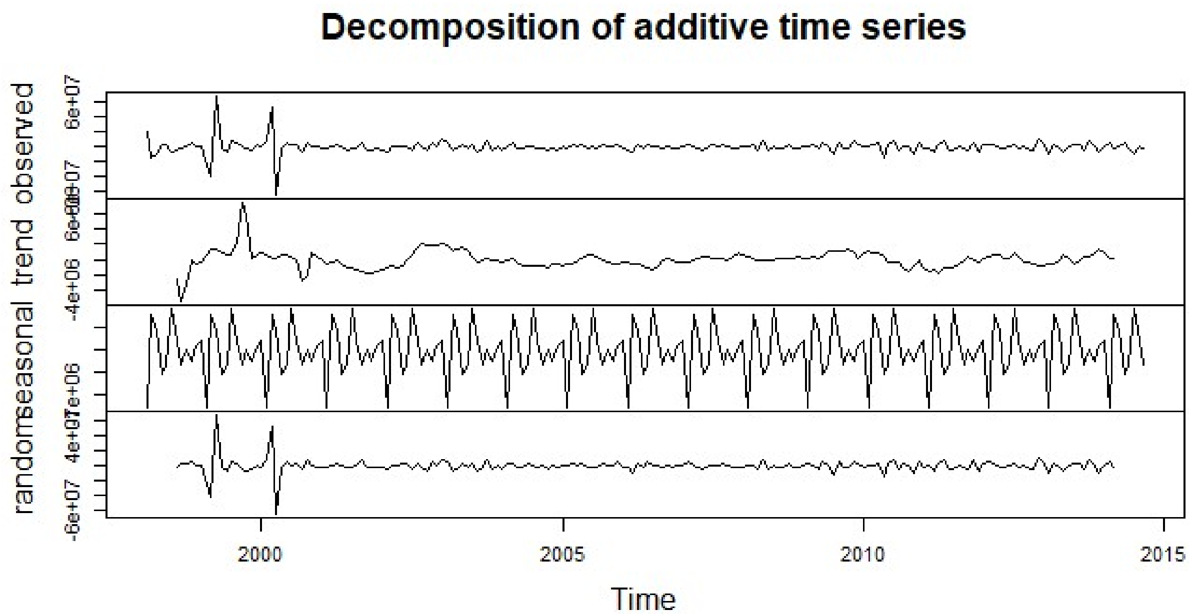

Figure 11. Decomposition after Regular and Seasonal Differencing

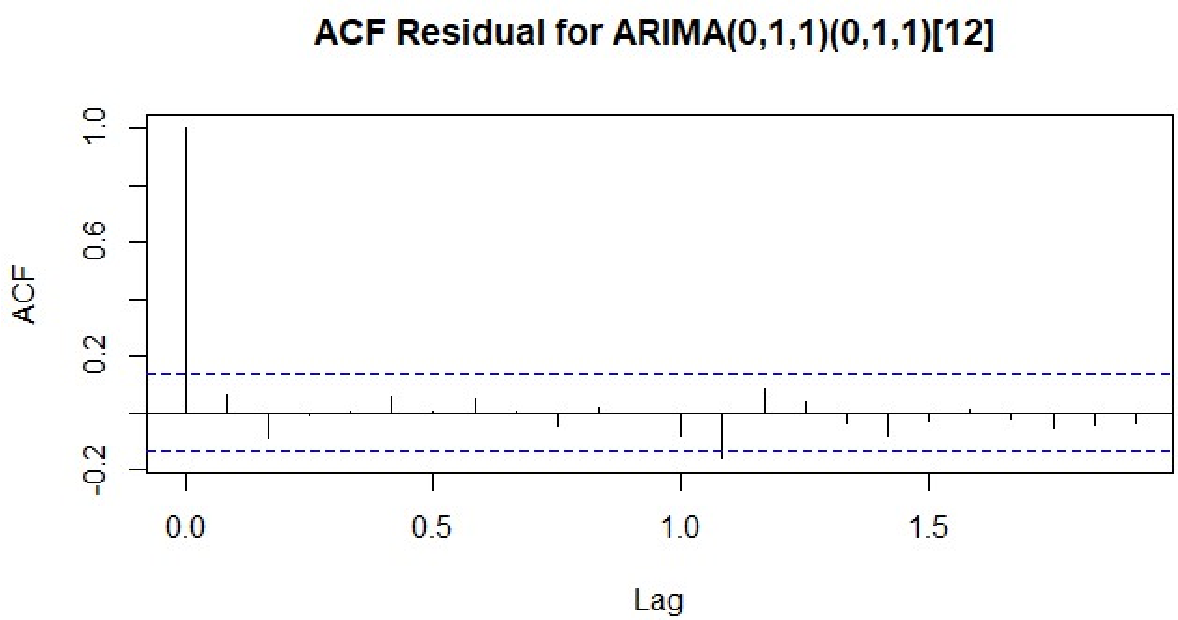

Figure 12. ACF Residuals from ARIMA$(0,1,1)(0,1,1)_{12}$ Model

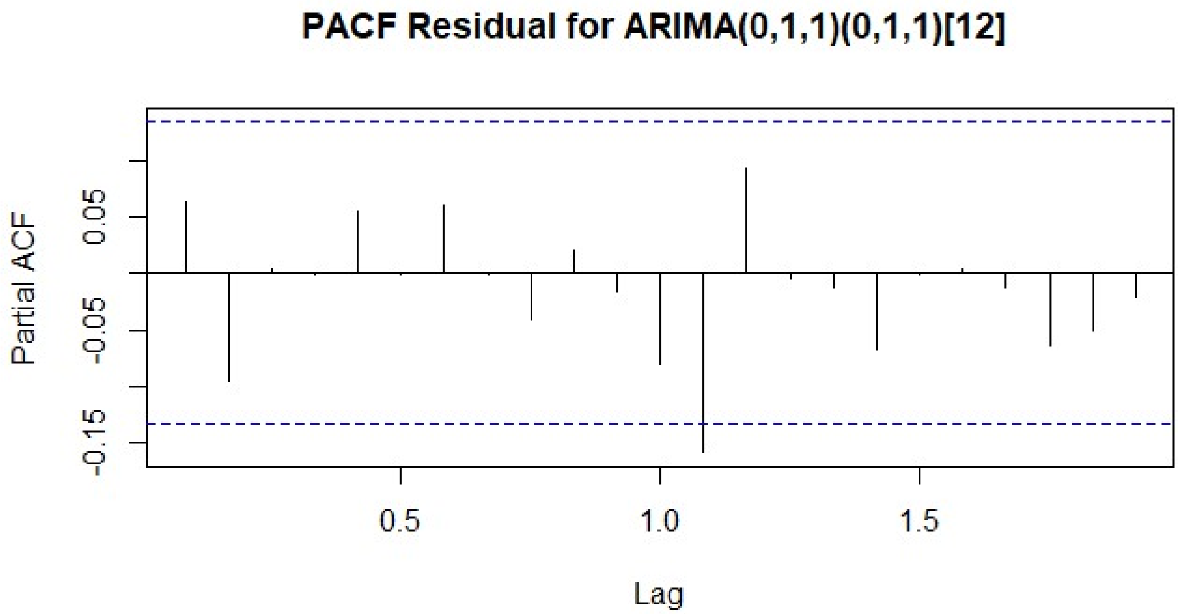

Figure 13. PACF Residuals from ARIMA(0,1,1)$(0,1,1)_{12}$ Model

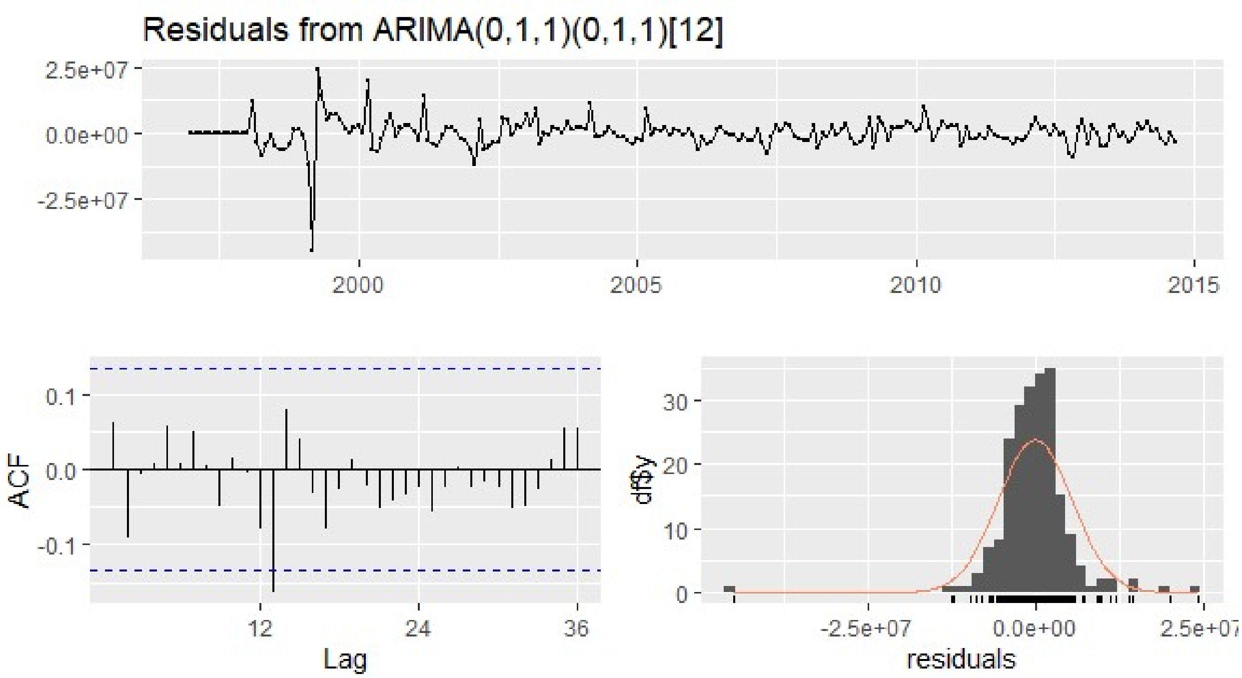

Figure 14. Residuals from Fitted ARIMA(0,1,1)$(0,1,1)_{12}$ Model

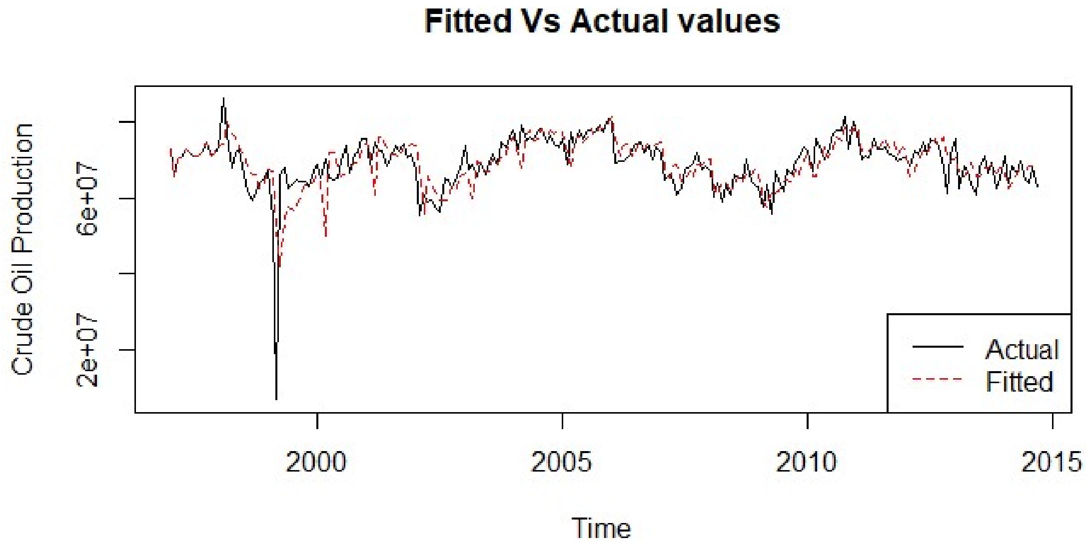

Figure 15. Fitted Vs Actual Values

Table 10. Forecasting with ARIMA(0,1,1)(0,1,1)12 Model

S/N	Period	Forecasts	95% Prediction Interval	
			Lower	Upper
214	October 2014	66993366	55666754	78319978
215	November 2014	64321971	52290163	76353779
216	December 2014	66957620	54259720	79655521
217	January 2015	67732676	54406889	81058464
218	February 2015	61435593	47507020	75364166
219	March 2015	62569469	48063137	77075802
220	April 2015	62864796	47802850	77926741
221	May 2015	64851582	49253802	80449362
222	June 2015	62964187	46848379	79079995
223	July 2015	64871982	48254287	81489677
224	August 2015	66502959	49398097	83607822
225	September 2015	64730534	47152000	82309068

5. Conclusion

Consequently, the $ARIMA(0,1,1)(0,1,1)_{12}$ model proved superior to other estimated models in the family of nested $ARIMA(p,d,q)(P,D,Q)_S$ models. Hence, it is statistically significant, appropriate, and adequate for the dataset. The plot of the fitted model agreed with the actual values, which implies that the model is a good fit and therefore suggests that the model could be used in forecasting future values of COP, Figure 15. The forecasts obtained using the fitted model revealed that there is a greater tendency for decreasing crude oil production in Nigeria (see Table 10).
Mathematically, the $ARIMA(0,1,1)(0,1,1)_{12}$ model is

$$\nabla\nabla_{12}F_t = (1 + \wp_1\Omega)(1 + \mathbb{Q}_{12,1}\Omega^{12})\varpi_t \quad (19)$$

So that,

$$F_t = F_{t-1} + F_{t-12} - F_{t-13} + \varpi_t - 0.6417\varpi_{t-1} - 0.9631\varpi_{t-2} + 0.6179\varpi_{t-13} \quad (20)$$

According to the prediction made using the fitted model in (20), Nigeria is more likely to see a decline in its production of crude oil. This can easily be explained by the lawlessness (oil theft) and corrupt practices that prevailed in the petroleum sector in Nigeria, coupled with government policy inconsistency in the sector. The decreasing rate of crude oil production has a strong effect on the Nigerian economy. Therefore, it is recommended that (i) the government regulate and supervise the operation of the petroleum sector, as is done in Saudi Arabia, Iraq, the USA, and other advanced oil-producing nations. (ii) The government should repair the local refineries and increase crude oil delivery to the local refineries to ensure a constant supply of petroleum products in the country. (iii) The government should diversify its economy to avoid distress, especially now that the crude oil price is dwindling in the international market.

Acknowledgment
We acknowledged Dr. D. W. Ebong for his insightful contributions.

References

[1].Akaike, H. (1974). A New Look at the Statistical Model Identification. I. E. E. E. *Transactions of Automatic Control, AC*, 19, 716 – 723.

[2].Box, G. and Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. *Holden – Day, San Francisco.*

[3].Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control, Revised Edition, *San Francisco: Holden-Day.*

[4].Box, G. E. P., Reinsel, G. C., and Jenkins, G. M. (1994). *Time Series Analysis: Forecasting and Control.* 3rd Ed. Prentice–Hall, England Cliffs, N. J.

[5].Clement, E. P. (2014). Using Normalized Bayesian Information Criterion (BIC) to improve Box-Jenkins model Building. *American Journal of Mathematics and Statistics*, 4(5), 214 – 221. DOI: 10.5923/k.ajms.20140405.02

[6].Clement, E. P. (2014). Time series Approach to Forex Rate Analysis between the Nigeria Naira and the U.S. Dollar. *International Journal of Statistics and Applications*, 4 (1), 76 – 84. DOI: 10.5923/j.statistics.20140401.08

[7].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 (366), 427 – 431. DOI: 10.1080/01621459.1979.10482431. JSTOR 2286348

[8].Etaga, H. O., Igwebuike, E. K. and Etaga, C. N. (2020). Time Series Models of the Crude Oil Production and Export in Nigeria (1999 - 2015). *African Journal of Mathematics and Statistics Studies*, 3(1), 1 - 24. ISSN: 2689 – 5323.

[9].Etuk, E. H., Inyang, E. J. and Udoudo, U. P. (2022). Impact of Declaration of Cooperation on the Nigerian Crude Oil Production. *International Journal of Statistics and Applied Mathematics*, 7(2), 165 – 169.

[10].Fatoki, O., Ilo, H. O. and Ugwoke, P. O. (2017). Time Series Analysis of Crude Oil Production in Nigeria. *Advances In Multidisciplinary & Scientific Research*, 3(3),1 – 10.

[11].Friedman, M. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the *Analysis of Variance. Journal of the American Statistical Association,* 32 (200), 675 – 701.

[12].Friedman, M. (1939). A Correction: The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association (American Statistical Association)* 34 (205), 109. DOI: 10.2307/2279169. JSTOR 2279169.

[13].Friedman, M. (1940). A Comparison of Alternative Tests of Significance for the Problem of M Rankings. *The Annals of Mathematical Statistics*, 11(1),86–92. DOI: 10.1214/aoms/1177731944. JSTOR 2235971

[14].Hyndman, R. J. and Khandakar, Y. (2008). Automatic Time Series Forecasting: The Forecast Package for R. *Journal of Statistical Software*, 27 (3), 1 – 22.

[15].Ijeoma, G. (2022). Time Series Analysis on Nigerian Monthly Crude Oil Prices. *Repository.mouau.edu.ng:* Retrieved Dec 28, 2023, from https:// Repository.mouau.edu.ng/work/view/ time-series-analysis-on-nigerian-monthly-crude-oil-prices-7-2

[16].Inyang, E. J., Etuk, E. H., Nafo, N. M., and Da-Wariboko, Y. A. (2023). Time Series Intervention Modelling Based on ESM and ARIMA Models: Daily Pakistan Rupee/Nigerian Naira Exchange Rate. *Asian Journal of Probability and Statistics*, 25(3), 1-17. Article no.AJPAS.106693. ISSN: 2582-0230. DOI: 10.9734/AJPAS/2023/v25i3560

[17].Inyang, E. J., Nafo, N. M., Wegbom, A. I., Da-Wariboko, Y. A. (2024). ETS - ARIMA Intervention Modelling of Bangladesh Taka/Nigerian Naira Exchange Rates. *Science Journal of Applied Mathematics and Statistics*,12(1),1–12.

[18].Inyang, E. J., Nsien, E. F., Clement, E. P., and Danjeh, A. G. (2022). Statistical Investigation of the impact of global oil politics: An Interrupted Time Series Approach. JP Journal of Mathematical Sciences, 32(1& 2), 1 – 13.

[19] Inyang, E. J., Clement, E. P., Raheem, M. A., and Ntukidem, S. O. (2024). A Pulse Intervention Modeling of Paediatric Anaemia Prevalence. Asian Journal of Statistics and Applications. 1(1), 1-15. https://doi.org/10.47509/AJSA.2024.v01i01.01

[20] Inyang, E. J. (2024). Intervention Model Based on Exponential Smoothing Methods and ARIMA Modelling of the Nigerian Naira Exchange Rates, International Journal of Scientific Research in Mathematical and Statistical Sciences, 11(1), 24-33.

[21].Kruskal, W. H. and Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of American Statistical Association*, 47 (260), 583 – 621.

[22].Ljung, G. M. and Box, G. E. P. (1978). On a Measure of a Lack of Fit in Time Series Models. *Biometrika*, 65 (2), 297 – 303. doi: 10.1093/biomet/65.2.297.

[23].Moffat, I. U. and Inyang, E. J. (2022). Impact Assessment of Gap on Nigerian Crude Oil Production: A Box-Tiao Intervention Approach. *Asian Journal of Probability and Statistics*, 17(2), 52 – 60. https://doi.org/10.9734/ajpas/2022/v17i230419

[24].NNPC (2014). Nigeria National Petroleum Corporation. *Nnpc business| Upstream venture| Oil production*. www.nnpcgroup.com

[25].NNPC (2014). Nigeria National Petroleum Corporation. Oil and Gas Statistical *Bulletin|MonthlyPerformance*. www.nnpcgroup.com

[26].Ollech, D. and Webel, K. (2020). A Random Forest-Based Approach to Identifying the most Informative Seasonality Test. *Deutsche Bundesbank's Discussion Paper Series 55/2020*.

[27].Omekara, C. O., Okereke, O. E., Ire, K. I. and Okamgba, C. O. (2015). ARIMA Modeling of Nigeria Crude Oil Production. *Journal of Energy Technologies and Policy*, 5(10). ISSN: 2224 – 3232.

[28].Sadeeq, S. A. and Ahmadu, A. O. (2018). A Study of the Suitable Time Series Model for Monthly Crude Oil Production in Nigeria. *International Journal of Mathematics and Physical Sciences Research*, 6(1), 106 – 112.

[29].Schwarz, G. E. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6 (2), 461 – 464. Doi : 10.1214/aos/117644136, MR 0468014.

[30].Pandher, S. S., Sardarli, A., and Volodin, A. (2022). Forecasting of Immigrants in Canada using Forecasting models. *Journal of Probability and Statistical Sciences*, 20 (1), 98 – 107.
[31].Shittu, O. I. and Inyang, E. J. (2019). Statistical Assessment of Government's Interventions on Nigerian Crude Oil Prices. *A Publication of Professional Statisticians Society of Nigeria, Proceedings of 3rd International Conference*, 3, 519-524.
[32].R Core Team (2022). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria*. URL https://www.R-project.org/.
[33].Taiwo, A. I., Adeleke, K. A. and Adedotun, A. F. (2021). Time Series Analysis of Nigerian Monthly Crude Oil Price. *FUPRE Journal of Scientific and Industrial Research*, 5(1), ISSN: 2579 – 1184.

Appendix: R Code

```
#Required package
library(forecast)
library(tseries)
library(seastests)
library(lmtest)
library(ggplot2)

#Data
getwd()
setwd("C:/Users/Elisha Inyang/Desktop/UG_Paper")
Eji<-read.csv("COProd.csv",header = T)
attach(Eji)

#Time Series Analyis
E<- ts(Eji, start=c(1997, 1), end=c(2014, 9), frequency=12)

#Time Plot
plot(E,main="Crude Oil Production")
acf(E,main="ACF of Crude Oil Production")
pacf(E,main="PACF of Crude Oil Production")

#Time Series Decomposition
decompose=decompose(E)
plot(decompose)

#Unit Root Test at Level
adf.test(E)

#Seasonality Test
kw(E, freq = 12, diff = T, residuals = F, autoarima = T)
fried(E, freq = 12, diff = T, residuals = F, autoarima = T)

#First Differencing
d1<-diff(E)
```

```
plot(d1,main="First Regular Difference")
acf(d1,main="ACF of First Regular Difference")
pacf(d1,main="PACF of First Regular Difference")
```

#Unit Root Test at First Difference

```
adf.test(d1)
```

#Seasonal Differencing

```
Sd2<-diff(d1,12)
plot(Sd2,main="Seasonal Difference of First Regular Difference")
acf(Sd2,main="ACF of Seasonal Difference of First Regular Difference")
pacf(Sd2,main="PACF of Seasonal Difference of First Regular Difference")
```

#Decomposition After Seasonal Differencing

```
decompose=decompose(Sd2)
plot(decompose)
```

#Seasonality Test After Seasonal Differencing

```
kw(Sd2, freq = 12, diff = T, residuals = F, autoarima = T)
fried(Sd2, freq = 12, diff = T, residuals = F, autoarima = T)
```

#Model Fitting

```
fit1<-arima(E,order = c(1,1,1),seasonal = list(order=c(1,1,1),period=12),include.mean=FALSE)
coeftest(fit1)
BIC(fit1)
AIC(fit1)

fit2<-arima(E,order = c(0,1,2),seasonal = list(order=c(0,1,2),period=12),include.mean=FALSE)
coeftest(fit2)
BIC(fit2)
AIC(fit2)

fit3<-arima(E,order = c(1,1,0),seasonal = list(order=c(1,1,0),period=12),include.mean=FALSE)
coeftest(fit3)
BIC(fit3)
AIC(fit3)

fit4<-arima(E,order = c(0,1,1),seasonal = list(order=c(0,1,1),period=12),include.mean=FALSE)
coeftest(fit4)
BIC(fit4)
AIC(fit4)
```

 # Forecasts with C.I.

```
Forecast.model1 <- Arima(window(E,end=c(2014, 9)),order = c(0,1,1),
```

```
                    seasonal = list(order=c(0,1,1),period=12))
forecast(Forecast.model1,h=12)
```

#Ljung-Box Test

```
checkresiduals(fit4)
```

 # Plot the fitted values with actual values

```
plot(E,ylab ="Crude Oil Production",xlab = "Time",col = "black",main="Fitted Vs Actual values")
lines(fitted(fit4), lty=2,col = "red")
legend("bottomright", legend = c("Actual", "Fitted"),col = c("black", "red"), lty =c(1,2))
```