Int. J. Nonlinear Anal. Appl. 16 (2025) 4, 345–359
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2024.32531.4839

Comparison of the accuracy of Box-Jenkins and Holt-Winters methods for forecasting OPEC crude oil price

Kobara Hosseini, Mohammad Ali Keramati*, Mohammad Ali Afshar Kazemi, Zadollah Fathi

Department of Management, Center Tehran Branch, Islamic Azad University, Tehran, Iran

(Communicated by Seyed Hossein Siadati)

Abstract

Various methods have recently been proposed to predict essential economic and non-economic variables. Each forecasting method has its own advantages and disadvantages based on the nature of the input data. Box-Jenkins and Halt-Winters methods are among the new approaches for increasing the accuracy of forecasting results. Therefore, this research aimed to predict OPEC average oil price data for June 2022 to May 2024 based on the data from 2003 to 2022 using Box-Jenkins and Holt-Winters methodology with a single variable. The errors of Box-Jenkins methodology, among the time series, processes ARIMA 5.1.5, ARIMA 4.1.5, ARIMA 3.1.5, and ARIMA 5.1.3 have the best accuracy with MSE of 61.86, 63.21, 63.29, and 63.62, respectively. The accuracy of the Holt-Winters method was not appropriately compared to the time series method due to the nature of the data.

Keywords: forecasting, time series, Box-Jenkins, Holt-Winters
2020 MSC: 91B84

1 Introduction

Researchers have introduced sophisticated forecasting techniques in various mathematical, statistical, and artificial intelligence fields. Each forecasting method has its own advantages and disadvantages based on the nature of the input data. In recent years, the use of a novel combined approach in forecasting has gained attention from researchers, aiming to improve the accuracy of results. The combination of multiple forecasting methods using scientific and theoretical foundations can lead to improved accuracy. Autoregressive Integrated Moving Average (ARIMA) time series method based on the Box-Jenkins methodology is one of the methods for predicting various economic and non-economic variables. Evaluation indicators demonstrate higher accuracy for this method among statistically-based financial methods. The Holt-Winters method has also shown decent accuracy in forecasting within the realm of mathematical finance. Therefore, considering the presented approaches and techniques, this study aimed to predict crude oil prices as one of the influential macroeconomic indicators and compare the accuracy of the predicted approaches.

The reason for choosing the price of oil is that crude oil is one of the primary commodities in almost all sectors of an oil-based economy, especially in the industrial production sector [16]. According to the fundamental principles of economics, the price of crude oil is determined by the balance of supply and demand, and its forecasting is very complex. However, predicting this indicator is necessary, but it comes with many challenges with different results;

*Corresponding author
Email addresses: marjan.2014h@gmail.com (Kobara Hosseini), mohammadalikeramati@yahoo.com (Mohammad Ali Keramati), m_afsharkazemi@iauec.ac.ir (Mohammad Ali Afshar Kazemi), z_fathi46@yahoo.com (Zadollah Fathi)

Received: November 2023 *Accepted:* January 2024

therefore, various reasons have been proposed for this necessity as a research gap. First, changes in oil prices often directly affect the prices of related products, such as fossil fuels and other industrial raw materials, and consequently impact the prices of final products. Therefore, crude oil prices have always been essential and influential in assessing macroeconomic risks and forecasting. Second, oil price fluctuations can affect the value chain of related products and some products used in production. For example, when the price of oil increases, transportation costs will rapidly increase, and operational profits in this sector will decrease. On the other hand, when oil prices decline, the incentive for production by oil refining companies and operational profits will also decrease. Therefore, when companies can accurately predict oil price fluctuations, they can significantly reduce cost risks and maintain sustainable business profitability. Thus, predicting the price of crude oil has always been a central issue in economics [18].

On one hand, this forecasting significantly affects the domestic economy and global governance. On the other hand, oil price fluctuations and shocks significantly affect the real and virtual economy [3, 9, 13], domestic trade, and imports and exports [15]. Moreover, accurate forecasting of crude oil prices is essential for three main groups of stakeholders in economic activities, including policymakers to allocate budgets optimally [11, 16], for companies to ensure rational production decisions and cost management, and for investors to effectively allocate their assets [16, 17]. Therefore, more accurate forecasting of crude oil prices can significantly affect economic decision-making for different stakeholders and policymakers, and it has always been important from this perspective. Considering the above points, this study focuses on predicting crude oil prices as one of the most complex, volatile, and opaque financial markets. In this regard, the Box-Jenkins and Holt-Winters methodology introduced the most suitable and precise results based on several comparison criteria. The results can apply to economists and policymakers who are always seeking accurate forecasting of crude oil prices because oil price forecasting plays an influential role in government policies, affects the performance strategies of international financial markets, and optimizes long-term production levels. Therefore, this study aimed to find the optimal model for predicting oil prices, combining the Box-Jenkins and Holt-Winters approaches.

2 Theoretical foundations of research

2.1 Time series forecasting based on Box-Jenkins methodology

The objective of analyzing a time series is to describe, explain, and forecast the future values of a process. Describing a process involves plotting data, determining its stationary and non-stationary, and examining the autocorrelation of the series. Prediction involves estimating future values of the series based on observed data [7]. Time series can be divided into two types: stationary and non-stationary series. A series is considered stationary if it has no regular changes in its mean and variance and any strong cyclical variations have been removed. Non-stationary series can be transformed into stationary series by differencing or stabilizing their variances. The changes observed in time series can result from natural or other factors, so it is necessary to identify and measure their components. The essential goal of time series analysis is to find the model of changes and predict the future. Time series forecasting using the Box-Jenkins method is one of the techniques. In time series forecasting, the values of the data sets which are unknown at the time of the analysis are estimated. The Box-Jenkins approach is one of the techniques for predicting the behavior of a time series, which is based on a wide range of forecasting models for a time series. A combination of autoregressive and moving average models is the general group of models for a time series in the Box-Jenkins methodology, known as ARIMA models in statistics whose model is written as ARIMA(p, d, q), where p determines the order of autoregression (AR) and represents the dependence on the previous values. In this model, the regression of each element is determined based on its prior values. The autoregressive process is applicable in cases where the current value of the time series depends on its previous values plus a random coefficient. In the following, q determines the order of the moving average, which defines the dependence of the series on current and past random elements. The combined autoregressive pattern with moving average order (p, d, q) is represented. Forecasting time series models is based on four stages: pattern identification, parameter estimation, pattern adequacy testing, and forecasting.

2.2 Box-Jenkins methodology based on econometric time series

2.2.1 Econometric modeling

Econometrics is one of the modern and essential branches of economics that can significantly contribute to a better understanding the relationships between economic and social phenomena. Econometric models explain an economic system using simultaneous equations representing the interrelationships between measurable variables. The equations are based on theory or empirical observations of variables' behavior or technical relationships. When the analyzed model includes one or more lagged variables as explanatory variables for the dependent variable, it is called a dynamic model.

2.2.2 Characteristics of a Good Forecasting Model

A.C. Harvey [6] provides the following criteria for evaluating the quality of a model:

Logical lack of parsimony variables

A model can never perfectly describe reality as it is, and a complex model lacking any scientific value needs to be presented to accurately describe reality. Simplification and abstraction are essential in any modeling. Therefore, the principle of logical lack of parsimony variables causes a model to be as simple as possible, considering only the key variables in the analysis and thereby eliminating all random and partial effects by the disturbance or error term.

2.2.3 Identifiability

This principle states that the estimated parameters should yield unique values for a given dataset. In other words, there would be only one for each specific parameter.

2.2.4 Goodness of fit

The main goal of regression analysis is to explain variations in the dependent variable through significant parsimony variables in the model. Therefore, a temporal model is considered good and measured by R^2, even if it is as high as possible. However, the R^2 criterion should not be regarded as alone and should be used with other criteria.

2.2.5 Theoretical consistency

It is possible for a model, despite having a high R^2, to not be a good model due to the incorrect signs of one or more of its coefficients. In such cases, the results should be interpreted with caution.

2.2.6 Generalizability and predictability

Quoting Friedman [5], "...the only proper test for validating a hypothesis (model) is to compare its predictions with experiments." A high value implies the model's generalizability, and this value is specific to a particular sample and for that particular sample. Here, the power of prediction and generalization for a period outside the sample period is interesting.

2.2.7 Subsidiary forms of regression models

These forms include constant elasticity (double-log), semi-log, and inverse models, as explained in the following.

1. Constant elasticity models (log-linear) or Log-Log (Double-Log) models These models are commonly used in demand-related studies to estimate price and income elasticities of demand. These models are linear regarding the logarithm of variables (X, Y) and parameters (β, α), so they can be estimated using the OLS method. Log-linear or Log-Log (Double-Log) models were named in this way because of their linearity. These models are the same in angle and tension coefficients. In other words, the coefficients of each parsimony variable in the model measure the elasticity of Y concerning X or the percentage change in Y for a 1% change in X. Equation (2.1) expresses the general form of the model:

$$\ln Y_i = \alpha + \beta_2 \ln X_i + U_i. \tag{2.1}$$

   In which, ln presents the natural logarithm and Ln $\beta = \alpha_1$.

2. Semi-logarithmic models (Log-Lin and Lin-Log). Only one of the two variables, Y and X, is in logarithmic form, including two models in this category. If the dependent variable (Y) is logarithmic, it is called Log-Lin. In this class of models, the angle coefficient α_2 measures the constant relative or proportional change in Y for the absolute change in X.

$$\ln Y_i = \alpha_1 + \beta_2 X_i + U_i. \tag{2.2}$$

   If parsimony (X) is logarithmic, it is called Lin-Log. In this class of models, the angle coefficient β_2 measures the absolute change in Y for a relative change in X.

$$Y_i = \beta_1 + \beta_2 \ln X_i + U_i. \tag{2.3}$$

3. Inverse models
   This category of models is nonlinear in terms of variable X because it is entered in reverse in the model, but it is linear in terms of parameters β_1 and β_2. Therefore, it is a linear regression model, and the inverse model is shown as equation (2.4).

$$Y_i = \beta_1 + \beta_2(1/X_i) + U_i. \tag{2.4}$$

   In this model, as X increases indefinitely, the $(X_i/1)\beta_2$ component tends to zero β_2 is constant, and Y tends asymptotically or marginally to the value of β_1.

2.3 Holt-Winters exponential smoothing methods

In this research, the Holt-Winters exponential smoothing method was used. The difference between this method and the single or double exponential adjustment method is that the simple or double method has a constant parameter in the equation, but in the Halt method, two parameters, α, and β, are used in the equation. This method is used for series with linear time trend and non-seasonal series. The general form of the Holt-Winters model is Equation (2.5).

$$\hat{Y}_{t+k} = a + bk \tag{2.5}$$

in which, a and b are permanent components and defined by Equations (2.6) and (2.7).

$$A(t) = \alpha y_t + (1-\alpha)(\alpha(t-1) + b(t-1)) \tag{2.6}$$

$$B(t) = \beta(a(t)–a(t-1)) + 1–\beta b(t-1) \tag{2.7}$$

in which, $\alpha > 1$, $\beta, \gamma > 0$ are moderating factors. Forecasts are calculated by Equation (2.8).

$$Y_{t+k} = a(t) + b(t)k \tag{2.8}$$

These forecasts are used in the linear trend with intercept $(T)a$ and slope $(T)b$.

2.3.1 Multiplicative Halt-Winters method (with three parameters)

This method is used for series with linear time trends ty test that was performed in th and incremental (multiplicative) seasonal changes, and smoothed series are obtained as Equation (2.9).

$$\hat{Y}_{t+k} = (a + bk)c_{t+k} \tag{2.9}$$

in which, a is the constant coefficient, b is the trend, and c is the seasonal increasing factors, which are defined by Equations (2.10), (2.11), and (2.12).

$$A(t) = \alpha \frac{Yt}{Ct(t-s)} + (1-\alpha)(a(t-1) + b(t-1)) \tag{2.10}$$

$$B(t) = \beta(a(t)–a(t-1)) + (1-\beta)b(t-1) \tag{2.11}$$

$$C_t(t) = \gamma \frac{Yt}{a(t)} + (1-\gamma)Ct(t) \tag{2.12}$$

in which, $1 > \alpha$, β, and $\gamma > 0$ are the moderating factors, and s is the season frequency in place of the seasonal period. In this method, forecasts are calculated by Equation (2.13).

$$\hat{Y}_{t+k} = (a(t) + b(t)k)c_{t+k-e} \tag{2.13}$$

in which the seasonal factors were applied from the last estimated s.

2.3.2 Collective Holt-Winters method (with three parameters)

This method is used for series with linear time trend and accumulated seasonal changes. The smoothed series is obtained by Equation (2.14).

$$\hat{Y}_{t+k} = a + kb + c_{t+k} \tag{2.14}$$

in which a and b are constant components, and c is the seasonal cumulative factor. These three coefficients are defined by Equations (2.15), (2.16) and (2.17).

$$A(t) = \alpha(y_t – c_t(t–s)) + (1 - \alpha)(a(t-1) + b(t-1)) \tag{2.15}$$

$$B(t) = \beta(a(t)–a(t-1)) + 1–\beta b(t-1) \tag{2.16}$$

$$C_t(t) = \gamma(y_t–a(t+1))–\gamma c_t(t-s) \tag{2.17}$$

where $\alpha > 1$, β, and $\gamma > 0$ are the moderating factors, and s is the frequency of the season placed in place of the seasonal period.

Forecasts are calculated by Equation (2.18).

$$\hat{Y}_{t+k} = a(t) + b(t)k + c_{t+k-s} \tag{2.18}$$

3 Research questions

1. What is the accuracy level of Box-Jenkins methodology-based forecasting methods in predicting OPEC crude oil prices?
2. Which Box-Jenkins methodology-based models have the highest accuracy?
3. What is the accuracy level of the Holt-Winters forecasting method in predicting OPEC crude oil prices?
4. Among the Box-Jenkins and Holt-Winters forecasting methods, which is more accurate in predicting OPEC crude oil prices?

4 Research methodology

This descriptive, analytical, and comparative study employed past data to forecast the future. The data collection method was field collection based on global OPEC crude oil price statistics. Data from crude oil prices were used as a basis for future forecasting. The data collection tool was a monthly OPEC crude oil price checklist through the OPEC website.

The spatial scope of the research is related to the countries' economy. The OPEC organization announced statistical data of the OPEC oil price, including 250 monthly OPEC oil price data until 2022. In the current research, the monthly data is the daily average of each month. The library method collected the required statistical data, and Eviews, MATLAB, and Games software were used for data processing. The OPEC crude oil price time series was received monthly in the first stage. Then, the obtained data was used for forecasting, and the nearest neighbor method was used for missing data. Forecasting was in-sample and out-of-sample using Box-Jenkins Holt-Winters methods.

4.1 ARIMA time series prediction method based on the Box-Jenkins method

First, the stationarity status of the data was measured after collecting OPEC crude oil price data every month, and differentiation was used if the data were not random. Forecasting time series after stabilization was based on the Box-Jenkins methodology, which consisted of four stages. At this stage, 4 ARIMA time series models with appropriate accuracy in forecasting were selected as the best models.

4.2 First step (data pattern identification)

In this step, data related to oil prices and stationarity in the mean and variance of the data were evaluated by drawing autocorrelation function (ACF) and partial autocorrelation function (PACF) diagrams. ACF and PACF are the essential parameters for testing data dependence. This function measures the correlation between observations in different months and examines a single time series in the time domain. This function often provides insight into the likely pattern generating the data, which is used to identify and fit a suitable stochastic model to the data. PACF is used when it is intended to check the correlation between ($Z_t \square Z_{t+k}$), after the joint linear dependence between the variables $(Z_{t+1} + Z_{t+2}... + Z_{t+k})$ in addition to the autocorrelation between $(Z_t\Box Z_{t+k})$. The behavior of these functions in the correlogram chart is one of the key criteria for estimating the time series pattern. In instability, the desired series becomes stationary using a suitable differential series. In this step, the order of the parameter (d) for the model is selected with minimum variance by analyzing the variance of the differentiated data. On the other hand, the ranks of p and q are determined using ACF and PACF charts.

4.3 The second stage (parameter estimation)

At this stage, criteria such as mean absolute percentage error (MAPE) can be used by identifying suitable patterns in step one to compare several patterns and select the best ones.

4.4 The third step (diagnosing the pattern correctness)

In this step, the graph of the residuals is evaluated for normality and stationarity for the correctness of the model.

4.5 The fourth stage (forecasting)

The ability to forecast is one of the reasons for the popularity and spread of ARIMA modeling. The compromise model is more reliable than the traditional econometric modeling method. The forecast values in this step are determined for the future using the equations obtained in the above steps.

4.6 Using the Halt-Winters forecasting method

There are many time series that a polynomial cannot adequately model. For example, a time series with seasonal or cyclical changes cannot be easily represented by a polynomial model. Halt-Winters result is used to forecast the seasonal time series. Winters developed the exponential smoothing model in the 1960s to evaluate the data that depend on the period or seasons of the year or have a trend based on four relationships, three of which smooth the random, trend, and seasonal effects. Table 1 expresses the exponential smoothing prediction method. S is the smoothed value of the series whose seasonal effect is taken. In addition, T is the smoothed value of the trend, I is the smoothed value of the seasonal factor, X is the actual value, L is the length of the period, $F_{(t+m)}$ is the forecast of the m period after time t, α, β and γ are the constants of seasonal data smoothing. In addition, the steps and operations of forecasting the subject under study using smoothing forecasting are stated.

Table 1: OPEC crude oil price forecasting steps using the Halt-Winters method

Stages	**Description**	**Operation**
Stage one	Obtaining the smoothed value of the time series for the data	$s_t = \alpha \frac{X_t}{I_{t-1}} + (1-\alpha)(S_{t-1} + T_{t-1})$
Stage two	Obtaining the time series growth equation for the data	$T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}$
Stage three	Obtaining the surface equation for time series data	$I_t = \gamma \frac{X_t}{S_t} + (1-\gamma)I_{t-l}$
Stage four	Making predictions and obtaining values	$F_{t+m} = (S_t + T_t m)I_{t-l+m}$

5 Findings

First step: Identifying based on the Box-Jenkins method'

The appropriate model for the data is identified after stationing the data using the form of autocorrelation ACF and partial autocorrelation PACF functions. ACF and PACF functions of the oil price are shown in Figure 1.

A model for time series is prepared in the ARIMA method or Box Jenkins model, which considers the behavior of a variable or a set of variables based on their past behavior. These models are based on a specific and relatively small number of variables and do not accept the emphasis on theoretical foundations as a priority in predicting data trends. The appropriate model for each series should be selected based on the lowest MSE value based on significance to use the Box-Jenkins model to forecast oil prices for the coming months. According to the generalized Dickey-Fuller stationarity test performed in the previous part, the oil price's time series in the origin's level and width was stopped after one differentiation. Therefore, ARIMA is used to predict the series with the Box-Jenkins method.

Based on the results and the average error square MSE comparison, the ARIMA model (5.15) had the lowest error square value (61.8630) among different ARIMA models. Then, ARIMA(4.1.5), ARIMA(3.1.5), and ARIMA(5.1.3) models had the least error, and the analysis of each one was given according to Box Jenkins methodology.

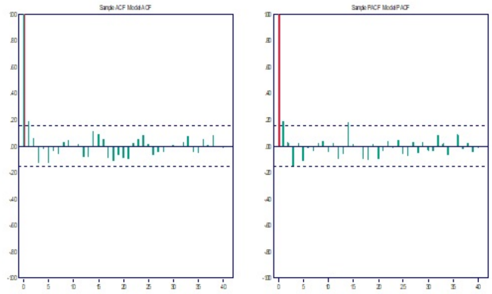

Figure 1: ACF and PACF functions of oil price

Table 2: ARIMA (p.d.q) models for oil prices

Model type	Process rank AR	Process rank MA	The order of difference	Index MSE
ARIMA (1.1.1)	1	1	1	69.7565
ARIMA (1.1.2)	1	2	1	67.6026
ARIMA (1.1.3)	1	3	1	68.1315
ARIMA (1.1.4)	1	4	1	66.9464
ARIMA (1.1.5)	1	5	1	66.4456
ARIMA (2.1.1)	2	1	1	68.3238
ARIMA (2.1.2)	2	2	1	66.4269
ARIMA (2.1.3)	2	3	1	65.9390
ARIMA (2.1.4)	2	4	1	65.9677
ARIMA (2.1.5)	2	5	1	65.9891
ARIMA (3.1.1)	3	1	1	68.1033
ARIMA (3.1.2)	3	2	1	67.1036
ARIMA (3.1.3)	3	3	1	65.8350
ARIMA (3.1.4)	3	4	1	68.0145
ARIMA (3.1.5)*	3	5	1	63.2919
ARIMA (4.1.1)	4	1	1	68.0992
ARIMA (4.1.2)	4	2	1	65.9736
ARIMA (4.1.3)	4	3	1	66.0377
ARIMA (4.1.4)	4	4	1	63.5430
ARIMA (4.1.5)*	4	5	1	63.2150
ARIMA (5.1.1)	5	1	1	66.8641
ARIMA (5.1.2)	5	2	1	66.8177
ARIMA (5.1.3)*	5	3	1	63.6299
ARIMA (5.1.4)	5	4	1	63.7311
ARIMA (5.1.5)*	5	5	1	61.8630

5.1 ARIMA model estimation (5.15)

The MLE maximum likelihood method was used to estimate model coefficients, but nonlinear methods are used when the model is nonlinear concerning parameters. ARIMA process equation (5.1) is in the form of the following equation.

$$\begin{aligned} Z(t) &= 0.4213Z(t-1) + 0.1355Z(t-2) - 0.4843Z(t-3) - 0.4666Z(t-4) + 0.3264Z(t-5) + a(t) \\ &\quad - 0.2527a(t-1) - 0.1470a(t-2) + 0.3925a(t-3) + 0.7050a(t-4) - 0.4252a(t-5) \end{aligned} \tag{5.1}$$

5.2 Diagnostic checking (ARIMA) (5.1.5)

The fit of the data selection model is checked after choosing a specific ARIMA model and estimating its parameters because another ARIMA model may provide a better fit for the data. A simple test to check this point is that the residuals from this model should be white noise, i.e., have a normal distribution, constant variance, and zero mean. The results showed that the residuals from the ARIMA process (5.1.5) are white disturbances.

Figure 2 shows the residual diagram of the ARIMA process (5.1.5).

5.3 ARIMA process prediction (5.1.5)

This step shows 24 out-of-sample forecasts for the ARIMA 5.1.5 process.

Figure 2: Residual of the ARIMA process (5.1.5)

Table 3: Forecasts of ARIMA 5.1.5

Forecast step	**Forecast month**	**Forecasted price**
1	6/2022	111.56
2	7/2022	112.34
3	8/2022	104.74
4	9/2022	107.70
5	10/2022	109.07
6	11/2022	113.11
7	12/2022	117.90
8	1/2023	116.50
9	2/2023	115.47
10	3/2023	111.62
11	4/2023	110.15
12	5/2023	112.27
13	6/2023	115.39
14	7/2023	119.70
15	8/2023	120.90
16	9/2023	119.55
17	10/2023	116.83
18	11/2023	114.46
19	12/2023	115.13
20	1/2024	117.97
21	2/2024	121.78
22	3/2024	124.20
23	4/2024	123.83
24	5/2024	121.59

Figure 3: Forecasting the 5.1.5 process

5.4 ARIMA model estimation (4.1.5)

$$Z(t) = 0.2251Z(t-1) + 0.05325Z(t-2) - 0.5725Z(t-3) - 0.3627Z(t-4) + a(t) - 0.05883a(t-1) - 0.03714a(t-2) + 0.4618a(t-3) + 0.6256a(t-4) - 0.08309a(t-5) \quad (5.2)$$

5.5 ARIMA (4.1.5) process control

Data Points =166

Sample Mean = -0.0206

Sample Variance = 63.214203

St. Error (Sample Mean) = 0.574345

Figure 4: The residual of the ARIMA process (4.1.5)

5.6 Forecasting ARIMA 4.1.5 process

Table 4: Forecasting ARIMA 4.1.5

Forecast step	**Forecast month**	**Forecasted price**
1	6/2022	116.53
2	7/2022	113.85
3	8/2022	105.14
4	9/2022	109.66
5	10/2022	109.68
6	11/2022	114.96
7	12/2022	117.56
8	1/2023	117.61
9	2/2023	115.57
10	3/2023	112.55
11	4/2023	111.62
12	5/2023	113.24
13	6/2023	116.87
14	7/2023	120.23
15	8/2023	121.43
16	9/2023	120.05
17	10/2023	117.40
18	11/2023	115.67
19	12/2023	116.33
20	1/2024	119.24
21	2/2024	122.72
22	3/2024	124.75
23	4/2024	124.33
24	5/2024	122.13

Figure 5: Forecasting ARIMA 4.1.5 process

5.7 Forecasting ARIMA 3.1.5

$$Z(t) = 0.6195Z(t-1) - 0.09256Z(t-2) - 0.6607Z(t-3) + a(t) - 0.4575a(t-1) + 0.06614a(t-2) + 0.5387a(t-3) + 0.2857a(t-4) - 0.1396a(t-5) \tag{5.3}$$

5.8 ARIMA(3.1.5) process control

Sample Mean = -0.0197

Sample Variance = 63.291775

Figure 6: Residual of ARIMA (3.1.5) process

5.9 Forecasting ARIMA (3.1.5) process

Table 5: Forecasting ARIMA (3.1.5) process

Forecast step	Forecast month	Forecasted price
1	6/2022	112.46
2	7/2022	114.06
3	8/2022	106.12
4	9/2022	109.09
5	10/2022	109.45
6	11/2022	115.21
7	12/2022	117.37
8	1/2023	118.50
9	2/2023	115.77
10	3/2023	113.13
11	4/2023	111.56
12	5/2023	113.22
13	6/2023	116.71
14	7/2023	120.33
15	8/2023	121.72
16	9/2023	120.52
17	10/2023	117.83
18	11/2023	115.93
19	12/2023	116.36
20	1/2024	119.16
21	2/2024	122.69
22	3/2024	124.90
23	4/2024	124.76
24	5/2024	122.56

Figure 7: Forecasting by ARIMA 3.1.5 process

5.10 Forecasting ARIMA 5.1.3 model

$$Z(t) = 0.7881Z(t-1) - 0.2189Z(t-2) - 0.7618Z(t-3) + 0.2800Z(t-4) - 0.1449Z(t-5) + a(t) - 0.6256a(t-1) + 0.1715a(t-2) + 0.6454a(t-3) \tag{5.4}$$

5.11 ARIMA 5.1.3 process control

Sample Mean = -0.0195

Sample Variance = 63.629162

Std.Error(Sample Mean) =0.581731

Figure 8: Residual of ARIMA 5.1.3 process

5.12 Forecasting ARIMA 5.1.3 process

Table 6: Prediction of ARIMA 5.1.3 process

Forecast step	Forecast month	Forecasted price
1	6/2022	112.16
2	7/2022	113.74
3	8/2022	106.76
4	9/2022	108.7
5	10/2022	109.18
6	11/2022	115.49
7	12/2022	117.29
8	1/2023	119.07
9	2/2023	115.65
10	3/2023	113.42
11	4/2023	111.19
12	5/2023	113.30
13	6/2023	116.46
14	7/2023	120.60
15	8/2023	121.80
16	9/2023	120.88
17	10/2023	117.85
18	11/2023	115.98
19	12/2023	116.15
20	1/2024	119.11
21	2/2024	122.64
22	3/2024	125.09
23	4/2024	124.86
24	5/2024	122.79

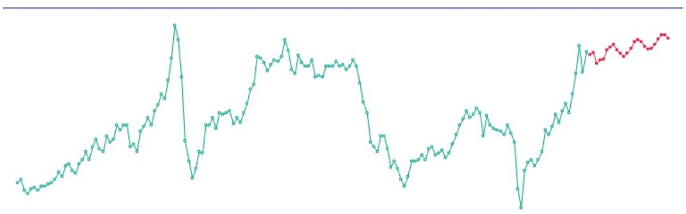

Figure 9: Forecasting ARIMA 5.1.3 process

Table 7: Input data to the Halt-Winters model

Data name	Series
Number of samples	167
Number of forecasts	24
(Seasonality) fluctuation type	Monthly
Smoothing parameters	
alpha	0.96
Beta	0.34
gamma	1

Figure 10: Forecasting residual of Holt-Winters method

5.12.1 Forecasting using the Halt-Winters method

Table 7 shows the data and smoothing parameters based on the Halt-Winters forecasting method.

Sample Mean = 67.7784

Sample Variance =0.718376E+03

The predicted values based on the Holt-Winters forecasting method are shown in Table 8 and Figure 11.

Table 8: Holt-Winters forecasting values

Forecast step	Forecast month	Forecasted price
1	6/2022	113.34
2	7/2022	115.55
3	8/2022	117.42
4	9/2022	114.83
5	10/2022	113.70
6	11/2022	115.12
7	12/2022	115.70
8	1/2023	116.59
9	2/2023	117.77
10	3/2023	117.70
11	4/2023	121.00
12	5/2023	120.34
13	6/2023	122.55
14	7/2023	124.43
15	8/2023	121.83
16	9/2023	118.49
17	10/2023	120.70
18	11/2023	122.12
19	12/2023	122.70
20	1/2024	123.59
21	2/2024	124.77
22	3/2024	124.07
23	4/2024	128.00
24	5/2024	129.12

Table 9 presents the comparison of the accuracy of forecasting methods.

Based on the results, the combined mode played an essential role in improving the accuracy of the forecasting methods and significantly enhanced the MSE, MAPE, AIC, and BIC indices.

Figure 11: Holt-Winters process forecasting

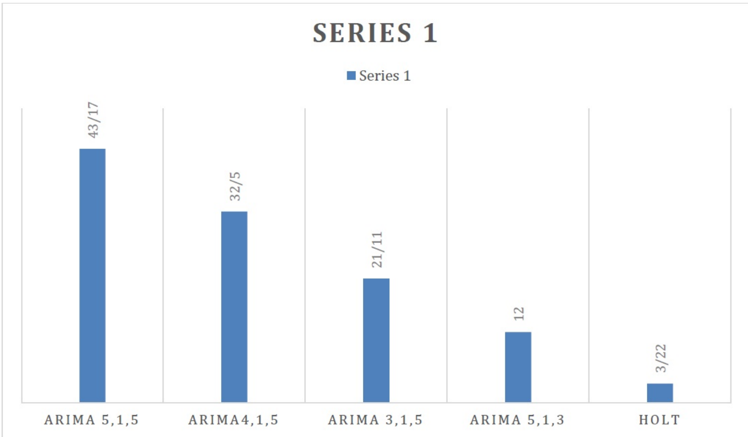

Figure 12: Importance percentage of forecasting methods

Table 9: Accuracy of forecasting methods

	MSE	MAPE	AIC	BIC
ARIMA 5.1.5	61.8630	10.1%	1184	1175
ARIMA 4.1.5	63.2150	11.7%	1194	1184
ARIMA 3.1.5	63.2919	12.02%	1202	1196
ARIMA 5.1.3	63.6299	12.46%	1214	1198
HOLT	69.2561	17.25%	1424	1412

6 Discussion and conclusion

1. What is the accuracy level of Box-Jenkins methodology-based forecasting methods in predicting OPEC crude oil prices?
   Based on the results, ARIMA 5.1.5, ARIMA 4.1.5, ARIMA 3.1.5, and ARIMA 5.1.3 exhibited superior accuracy among ARIMA models compared to other methods. Error analysis revealed that the MSE for ARIMA 5.1.5 was 61.86, with AIC and BIC indices at 1184 and 1175, respectively. The MSE for ARIMA 4.1.5 was 63.215, with AIC and BIC at 1194 and 1184, respectively. The accuracy of ARIMA 3.1.5 was found to be 63.2919, and the AIC and BIC scores were equal to 1202 and 1196, respectively. The MSE for ARIMA 5.1.3 was 63.6299, with AIC and BIC as much as 1214 and 1198, respectively. Following the evaluation of different ARIMA processes, four models were selected for the combination stage. The results were consistent with those of Karakurt et al. [8], and Śmiech et al. [14].
2. Which Box-Jenkins methodology-based models have the highest accuracy?
   Four indicators were used to assess the accuracy of the forecasting processes under the Box-Jenkins framework: Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and the AIC and BIC indices. Among the 25 models examined in this study, the ARIMA 5.1.5 process exhibited superior accuracy, indicated by lower MSE, MAPE, AIC, and BIC values. These results were in line with those of Karakurt et al. [8], Śmiech et al. [14], Lyu et al. [10], and Abd Elaziz et al. [1].
3. What is the accuracy level of the Holt-Winters forecasting method in predicting OPEC crude oil prices?
   The alpha coefficient was 0.96, the beta was 0.34, and the gamma was 1 in forecasting using the Holt-Winters method. The accuracy of the Holt-Winters forecasting method indicated that the error, based on the MAPE

index, was approximately 17.25%, which is relatively high. Abdollahi et al. [2] and Fan et al. [4] reported similar high MAPE values.

4. Among the Box-Jenkins and Holt-Winters forecasting methods, which is more accurate in predicting OPEC crude oil prices?
   The comparison between the Holt-Winters and Box-Jenkins methodologies revealed that the error rate of Holt-Winters is higher than that of the Box-Jenkins approach. Research by Omidi and Omidi [12] reported higher accuracy for time-series methods compared to Holt-Winters, which is consistent with the results of this research.

According to the results, the following recommendations can be made:

1. The appropriate accuracy of the forecasting methods indicated their ability to predict and estimate the trend governing economic variables, including oil price accuracy. Therefore, trustees of this field are recommended to use the combined approaches presented in this study to make effective decisions and policies.
2. The results emphasized the priority of using random forecasting methods among the approaches used in the current research. Therefore, it is suggested to use this method to predict the price of oil, considering the accuracy of the ARIMA method compared to other methods.
3. According to the results of improving the accuracy of combined forecasting methods, it is suggested to use the approaches of this research in order of accuracy priority to forecast other essential and influential variables in the economy, such as capital market indices and financial markets.
4. The design of forecasting software based on the methodology of this research in accordance with the mission of economic specialists and experts can improve the speed of the forecasting process and increase their accuracy.

References

[1] M. Abd Elaziz, A.A. Ewee,s and Z. Alameer, *Improving adaptive neuro-fuzzy inference system based on a modified salp swarm algorithm using genetic algorithm to forecast crude oil price*, Natural Resources Res. **29** (2020), 2671–2686.

[2] H. Abdollahi, *A novel hybrid model for forecasting crude oil price based on time series decomposition*, Appl. Energy **267** (2020), 115035.

[3] L. Chen, Z. Zhang, F. Chen, and N. Zhou, *A study on the relationship between economic growth and energy consumption under the new normal*, Nat. Account. Rev. **1** (2019), no. 1, 28–41.

[4] D. Fan, H. Sun, J. Yao, K. Zhang, X. Yan, and Z. Sun, *Well production forecasting based on ARIMA-LSTM model considering manual operations*, Energy **220** (2021), 119708.

[5] M. Friedman, *The methodology of positive economics*, D.M. Hausman (Eds.), The philosophy of economics: an anthology, Cambridge University Press, 1994.

[6] A.C. Harvey, *The Econometric Analysis of Time Series*, MIT Press, 1990.

[7] M. Istan, N. Husainah, M. Murniyanto, A. Suganda, I. Siswanti, and M. Fahlevi, *The effects of production and operational costs, capital structure and company growth on the profitability: Evidence from manufacturing industry*, Account. **7** (2021), no. 7, 1725-1730.

[8] I. Karakurt, *Modelling and forecasting the oil consumptions of the BRICS-T countries*, Energy **220** (2021), 119720.

[9] T. Li and G. Liao, *The heterogeneous impact of financial development on green total factor productivity*, Front. Energy Res. **8** (2020), 29.

[10] Y. Lyu, S. Tuo, Y. Wei, and M. Yang, *Time-varying effects of global economic policy uncertainty shocks on crude oil price volatility: New evidence*, Resources Policy **70** (2021), 101943.

[11] A. Murat and E. Tokat, *Forecasting oil price movements with crack spread futures*, Energy Econ. **31** (2009), no. 1, 85–90.

[12] N. Omidi and M.R. Omidi, *Estimating accident-related traumatic injury rate by future studies models in Semnan province, Iran*, Health Emergen. Disasters Quart. **3** (2021), no. 4, 191–198.

[13] P. Saculsan and T. Kanamura, *Examining risk and return profiles of renewable energy investment in developing countries: The case of the Philippines*, Munich Personal RePEc Arch. (2019), no. 97473. https://mpra.ub.uni-muenchen.de/id/eprint/97473

[14] S. Śmiech, M. Papież, M. Rubaszek and M. Snarska, *The role of oil price uncertainty shocks on oil-exporting countries*, Energy Econ. **93** (2021), p. 105028.

[15] L. Yu, R. Zha, D. Stafylas, K. He and J. Liu, *Dependences and volatility spillovers between the oil and stock markets: New evidence from the copula and VAR-BEKK-GARCH models*, Int. Rev. Financ. Anal. **68** (2020), 101280.

[16] Y. Zhang, F. Ma and Y. Wang, *Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors?*, J. Empir. Financ. **54** (2019), 97–117.

[17] Y. Zhao, J. Li and L. Yu, *A deep learning ensemble approach for crude oil price forecasting*, Energy Econ. **66** (2017), 9–16.

[18] Y. Zheng and Z. Du, *A systematic review in crude oil markets: Embarking on the oil price*, Green Finance **1** (2019), no. 3, 328–345.