

# A study of univariate forecasting methods for crude oil price

Mei-Ling Cheng and Ching-Wu Chu

*Department of Shipping and Transportation Management,  
National Taiwan Ocean University, Keelung, Taiwan, and*

Hsiu-Li Hsu

*Department of Air and Sea Logistics and Marketing,  
Taipei University of Marine Technology, Taipei, Taiwan*

Received 27 September 2021  
Revised 7 November 2021  
4 December 2021  
Accepted 6 December 2021

## Abstract

**Purpose** – This paper aims to compare different univariate forecasting methods to provide a more accurate short-term forecasting model on the crude oil price for rendering a reference to manages.

**Design/methodology/approach** – Six different univariate methods, namely the classical decomposition model, the trigonometric regression model, the regression model with seasonal dummy variables, the grey forecast, the hybrid grey model and the seasonal autoregressive integrated moving average (SARIMA), have been used.

**Findings** – The authors found that the grey forecast is a reliable forecasting method for crude oil prices.

**Originality/value** – The contribution of this research study is using a small size of data and comparing the forecasting results of the six univariate methods. Three commonly used evaluation criteria, mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percent error (MAPE), were adopted to evaluate the model performance. The outcome of this work can help predict the crude oil price.

**Keywords** Forecasting accuracy comparison, Univariate forecasting models, Crude oil price

**Paper type** Research paper

## 1. Introduction

Nearly one-third of global energy consumption uses crude oil, so crude oil is the most crucial energy resource on Earth. Petroleum products are also made of refined crude oil. Crude oil is the world's leading fuel, and its prices have a significant influence on the global environment and the economy.

The Organization of Petroleum Exporting Countries (OPEC) controls over 40% of the global oil supplies. The OPEC influences oil prices by setting production levels to meet the global demand for crude oil, increasing or decreasing production. When demand exceeds supply, the prices will go up, and when supply exceeds demand, the prices go down. In addition, politics, interest rates and natural disasters, such as COVID-19, can also impact crude oil prices.

Crude oil prices affect many levels, including global economic and trade development, national security, economy, central bank policy and corporate profitability. If we can make accurate predictions, the uncertainty about the future can be reduced. From a government perspective, making accurate forecasts on crude oil prices and implementing appropriate measures can reduce domestic economic shocks and increase the well-being of people.

The fuel cost accounts for somewhere between 22 and 38% of total expenses in the airline industry and is the first or second highest cost item. Therefore, it is essential to operate fuel-efficient aircraft and execute other strategies that can help mitigate jet fuel price volatility. Before the pandemic, a fuel hedging strategy was used by airlines to reduce their exposure to



---

fuel prices. In an environment of demand uncertainty and travel restrictions, including quarantine requirements imposed by governments, a rise in crude oil prices would decrease airline profitability (Midas Aviation).

The above situation in the airline industry is also applicable to the container shipping industry. The bunker cost is the first or second highest cost item. A decline in crude oil prices ultimately flows through to the price of marine fuel. The IMO 2020 rule requires ships without exhaust-gas scrubbers to switch from 3.5% heavy fuel oil to cleaner 0.5% sulfur fuel called very low sulfur fuel oil (VLSFO). The average VLSFO price at the world's top four bunker hubs was \$494 per ton, which was much higher than one year ago. An increase in crude oil price would decrease the container shipping industry's profitability.

There is no doubt that crude oil price forecasts are beneficial to governments as well as industries. Thus, forecasting crude oil prices has been an important subject of research by both academia and industry. The purpose of this paper is to make comparisons on different univariate forecasting methods to provide a more accurate short-term forecasting model on the crude oil price for rendering a reference to authorities or managers.

Our paper is organized as the following: [Section 2](#) provides the literature review. [Section 3](#) presents the methodology. The comparison of the results from all methods is discussed in [Section 4](#). Finally, [Section 5](#) provides some concluding remarks.

## 2. Literature review

[Bashiri Behmiri and Pires Manso \(2013\)](#) provided a comprehensive review of crude oil price forecasting techniques. They categorized the existing forecasting techniques into the two main groups of quantitative and qualitative methods. The quantitative method includes econometrics methods and nonstandard methods. Among them, econometrics models are grouped into three models: time series models, financial models and structural models. On the other hand, the main nonstandard methods that are most frequently applied in oil price forecasting are artificial neural networks and support vector machines. On the other side, qualitative methods estimate the impact of infrequent events, such as wars and natural events on oil prices; these approaches, such as Delphi method, belief networks, fuzzy logic and expert systems, and web text mining method, recently obtained more popularity among oil price forecasting literature.

Econometrics models are the most frequently used methods in existing oil price forecasting literature. Most of the literature studies on econometric methods to predict oil prices use time series models to make predictions. Time series models predict future oil prices based on historical data. In these models, the future price behaviors are deduced from their historical data. Time series models include moving average, exponential smoothing, decomposition method, autoregressive integrated moving average (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH) and vector autoregressive (VAR).

ARIMA models are widely used in forecasting crude oil prices. [Ahmad \(2011\)](#) applied the ARIMA approach for the time series analysis of monthly average prices of Oman crude oil. The author recommended a seasonal ARIMA for short-term forecasting. [Xiang and Zhuang \(2013\)](#) collected the monthly price of Brent crude oil from November 2012 to April 2013 and constructed an ARIMA model that provided a good prediction result. [Zhao and Wang \(2014\)](#) proposed an ARIMA model to predict world crude oil based on data from 1970 to 2006. They found that the model was able to describe and predict the average annual price of world crude oil for the short-term forecast. [Mensah \(2015\)](#) examined the monthly Brent oil price for the last two decades using ARIMA techniques and compared different ARIMA models. The result showed that the ARIMA (1,1,1) model was the best forecasting model amidst the volatilities in the oil price. [Faruk \(2018\)](#) used the ARIMA to construct the best model for the average monthly price of crude oil in Nigeria and recommended that the government should change

the oil price benchmark and replace the MA model with the ARIMA model to ensure more excellent financial stability and efficient fiscal management in Nigeria. [Selvi et al. \(2018\)](#) constructed the ARIMA model to forecast the average annual crude oil price of the period 2017–2021 based on the data of 1946–2016.

The ARIMA model is a linear model, which captures time series linear characteristics. However, some authors believed it might not be easy to grasp the risky and volatile data and explore other models. [Sadorsky \(2006\)](#) compared different forecasting models to forecast petroleum prices. West Texas Intermediate (WTI) daily future prices of crude oil, heating oil and unleaded gasoline were used. The results showed that the GARCH type models outperformed the other techniques. [Yaziz et al. \(2011\)](#) used Box–Jenkins and GARCH models to construct a forecast model of the daily price of West Texas crude oil. The study collected a large amount of historical data and found that GARCH (1, 1) captured the volatility of the data. [Nwafor et al. \(2018\)](#) compared the prediction ability of eight models based on the OPEC Reference Basket, and the results showed that the GARCH(1,1) model provided more accurate daily price forecasts. [Gupta et al. \(2020\)](#) collected a total of 1,500 records of the closing price of crude oil (5 years and 11 months). They proposed a forecasting model based on the back-propagation learning algorithm. The main advantage of this kind of neural network is to build a model for the nonlinear and complex relationship between input and output.

The multivariate time series model can be used to explain the causal relationship between variables. The advantage of using this type of model is that it can explain the causes and consequences in logical analysis and help the decision makers maintain a keen eye on oil price trends. But whether it is more accurate and stable is indecisive. [Ye et al. \(2005\)](#) proposed a forecast model for the monthly spot price of West Texas crude oil. They incorporated variables such as oil inventories, oil production, imports and demand on spot oil prices into the model. Since the total oil inventory level measures the balance between oil production and demand, the model could reflect the changing market conditions of crude oil prices.

[Beckers et al. \(2015\)](#) incorporated relevant variables into the vector auto-regression model. These variables include consumer price index (CPI) inflation, dollar exchange rates, US three-month and ten-year Treasury yields, spreads between long-term and short-term interest rates, oil supply and demand from major producers or regions, global industrial production, the Organisation for Economic Co-operation and Development (OECD) inventory demand and other variables. Although the vector auto-regression model has good forecast performance, it is still difficult to predict in the long-term sample with structural change.

[Lyu et al. \(2017\)](#) collected oil price data for 2003–2015, used a combination of factor analysis and data correlation to select the factors affecting West Texas crude oil prices for world crude oil demand, world crude oil supply, US crude oil inventories and the US dollar index. The authors incorporated these variables into the established inverted neural network and an ARIMA model to predict West Texas crude oil prices. The results indicated that the inverted neural network model had a better performance.

Crude oil prices are highly uncertain. Sometimes the prices remain stable; sometimes, the prices are unstable. To build a model, it must collect a tremendous amount of data. This process is very time-consuming, so a simple model using a small data size is getting popular. Scholars have tried to use the gray forecast in crude oil price forecasting in recent years, which brings a new application to this field. [Lin \(2009\)](#) used the grey forecast model to forecast the average monthly price of WTI crude oil. [Xu \(2010\)](#) applied the grey model to forecast China's crude oil consumption. [Huang et al. \(2016\)](#) utilized the grey prediction model to predict global crude oil consumption. [Duan et al. \(2018\)](#) made forecasting of crude oil consumption in China using a grey prediction model. [Norouzi et al. \(2020\)](#) adopted the gray forecasting model to predict the OPEC crude oil prices. The results of the gray forecast model are better than that of long-term data.

From the literature, we found that the time series model is the widely adopted method for crude oil price forecasting. It is also evident that most studies have focused on long-term forecasting and used a large data size. There is limited research using a small data size and comparing the performance of different methods on short-term forecasting. This paper attempts to bridge this gap in the application, so we aim to find a practical and yet highly accurate model. In this research, six different univariate methods, namely the Grey Forecast, the hybrid grey model, the multiplicative decomposition model, the trigonometric regression model, the regression model with seasonal dummy variables and the seasonal autoregressive integrated moving average (SARIMA), have been used.

### 3. Methodology

#### 3.1 Grey forecasting model

Grey theory, developed originally by Deng (1989), is a generic theory that deals with systems characterized by poor information or insufficient information. The primary merit of a grey model is that it needs fewer data to generate forecasts.

GM (1, 1) is a notation for a first-order and single-variable grey model. The constructing process of GM (1, 1) can be described as follows.

Denote the original data sequence as

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)), \quad (1)$$

where  $x^{(0)}(k)$  stands for the original data sequence in period  $k$ . The following sequence  $x^{(1)}$  is defined as

$$x^{(1)} = \left( \sum_{k=1}^1 x^{(0)}(k), \sum_{k=1}^2 x^{(0)}(k), \dots, \sum_{k=1}^n x^{(0)}(k) \right) = (x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(n)). \quad (2)$$

Equation (2) was generated based on the accumulated generating operation (AGO) of Equation (1).

To test whether the sequence was acceptable for constructing the model, we calculated the class ratio  $\sigma(k)$  as follows:

$$\sigma^{(1)}(k) = \frac{x^{(1)}(k-1)}{x^{(1)}(k)}, \quad k \geq 2, \quad (3)$$

If the original sequence  $\sigma^{(1)}(k) \in (0,1)$ , then  $x^{(1)}(k)$  was suitable for constructing the model.

After completing the class ratio test, we constructed the GM(1, 1) model by establishing a first-order differential equation for  $x^{(1)}(k)$  as follows:

$$\frac{dx^{(1)}}{dk} + ax^{(1)} = b \quad (4)$$

where  $a$  and  $b$  denoted the coefficients to be determined.

Next, we applied the ordinary least squares method to Equation (4) to estimate the coefficients of  $a$  and  $b$ . After obtaining the estimated coefficients,  $\hat{a}$  and  $\hat{b}$ , we substituted  $\hat{a}$  and  $\hat{b}$  in the following equation:

$$\hat{x}^{(1)}(k+1) = \left( x^{(0)}(1) - \frac{\hat{b}}{\hat{a}} \right) e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}} \quad \text{and} \quad \hat{x}^{(1)}(1) = x^{(0)}(1) \quad (5)$$

The forecast value of the time series could easily be calculated using the inverse accumulated generating operation (IAGO) to convert to  $x^{(0)}(k)$  as follows:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad (6)$$

### 3.2 Hybrid grey forecasting model

Peng and Chu (2009) found that the classical decomposition model was the best model for forecasting container throughput with seasonal variations. We modified the same forecasting model proposed by Peng and Chu (2009) by combining the grey model with the multiplicative decomposition model. Since the size of the initial sequence in the grey forecast could affect the forecast performance, we searched for the initial sequence's size with the lowest prediction errors. The constructing steps are as follows.

- (1) Step 1: The first step is to remove the seasonal component from the original data; this step is equal to the calculation of the multiplicative decomposition model discussed in Section 3.3. The predicted results were used as the input in Steps 2 and 3.
- (2) Step 2: We constructed a grey forecast model, as discussed in Section 3.1, based on the time-series data obtained in Step 1.
  - By setting  $K$ , the initial sequence's size to four, we obtained the predicted values and prediction errors.
  - We changed the size of  $K$  to 5 and continued the same calculation processes until  $K = 8$ .
  - We identified a value of  $K$  that provided the lowest prediction error.
- (3) Step 3: We conducted a grey forecast month by month based on the data for the same month of different years.
- (4) Step 4: Comparing the results from both Step 2 and 3, we chose  $K$ 's size with the lowest prediction error.

### 3.3 Multiplicative decomposition model

In the decomposition model, time series can be decomposed into four separate factors: trend, seasonal, cyclical and irregular factors. The decomposition model is based on intuition rather than theory. Two types of decomposition models: multiplicative and additive models, are commonly used in practice. In this study, we adopted the multiplicative model and expressed the time series as follows:

$$Y_t = TR_t \times SN_t \times CL_t \times IR_t, \quad (7)$$

where  $Y_t$  is the observed value of the time series in time period  $t$ .

$TR_t$  is the trend factor in time period  $t$ ,

$SN_t$  is the seasonal factor  $t$  in time period  $t$ ,

$CL_t$  is the cyclical factor in time period  $t$  and

$IR_t$  is the irregular factor in time period  $t$ .

First, we calculated a 12-period moving average and denoted it as  $MA_t$  for period  $t$ . Next, the centered moving average at time  $t$  was calculated as  $CMA_t = \frac{1}{2}(MA_t + MA_{t+1})$ . Since the trend and cyclical components were included in the centralized moving average series,

i.e.  $CMA_t = TR_t \times CL_t$ , in Equation (7), we calculated the product of the seasonal and irregular components of the time series as follows:  $SI_t = SN_t \times IR_t = \frac{Y_t}{CMA_t}$ .

To remove the irregular factor from  $SI_t$ , we estimated the average of the observations in month  $t$  for four successive years to obtain the seasonal factor for month  $t$ , expressed as  $SN_t = \frac{1}{4}(SI_t + SI_{t+12} + SI_{t+24} + SI_{t+36})$ .

Dividing  $Y_t$  by the seasonal index  $SN_t$ , we generated a deseasonalized series ( $\frac{Y_t}{SN_t}$ ), for estimating the trend factor  $TR_t$ . Next, the deseasonalized time series data were used to estimate  $TR_t$  by a linear trend model as follows:

$$TR'_t = \alpha + \beta t + \varepsilon_t \quad (8)$$

To obtain the point estimates for  $\alpha$  and  $\beta$ , we employed the least squares method to Equation (8) and obtained the following estimate for the trend factor:

$$\widehat{TR'_t} = a + bt \quad (9)$$

### 3.4 Trigonometric regression model

The trigonometric model can forecast time series that exhibit seasonal variations. We adopted a general specification that allows the modeling of a more complicated increasing seasonal pattern, suggested by Bowerman and O'Connell (1993), as follows:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 \sin \left[ \frac{2\pi t}{L} \right] + \beta_3 t \cos \left[ \frac{2\pi t}{L} \right] + \beta_4 \cos \left[ \frac{2\pi t}{L} \right] + \beta_5 t \cos \left[ \frac{2\pi t}{L} \right] + \beta_6 \sin \left[ \frac{4\pi t}{L} \right] + \beta_7 t \sin \left[ \frac{4\pi t}{L} \right] + \beta_8 \cos \left[ \frac{4\pi t}{L} \right] + \beta_9 t \cos \left[ \frac{4\pi t}{L} \right] + \varepsilon_t, \quad (10)$$

where  $L$  denotes the number of periods within a year in the data. In our case, we considered the monthly data so  $L$  is 12.

### 3.5 Regression model with seasonal dummy variables

The regression model can provide an advantageous technique for analysis variance by using seasonal dummy variables. When analyzing a time series with seasonal variation, we often use a model of the following form:

$$Y_t = TR_t + SN_t + \varepsilon_t,$$

where  $TR_t$  is the trend in time period  $t$ ,  $SN_t$  is the seasonal factor in time period  $t$  and  $\varepsilon_t$  denotes the error term in time period  $t$ . Assuming that seasonal variations could be represented by dummy variables with one for each month, we have the season factor as follows:

$$SN_t = \sum_{i=1}^{11} \beta_{si} X_{si,t}, \quad (11)$$

where  $X_{si,t} = \begin{cases} 1 & \text{if period } t \text{ is month } i, i = 1, 2, 3, \dots, 11, \\ 0 & \text{otherwise.} \end{cases}$

Substituting Equation (11) into the above model of the time series, we obtained the regression model as follows:

$$Y_t = \beta_0 + \beta_1 t + \beta_{s1} X_{s1,t} + \beta_{s2} X_{s2,t} + \dots + \beta_{s11} X_{s11,t} + \varepsilon_t. \quad (12)$$

Applying the least squares method to Equation (12), we expressed the forecasted value of the time series as

$$\hat{Y}_t = b_0 + b_1t + b_{s1}X_{s1,t} + b_{s2}X_{s2,t} + \dots + b_{s11}X_{s11,t}. \quad (13)$$

### 3.6 ARIMA

This research's seasonal model is the most general form of a univariate class of models originally suggested by Box and Jenkins (1976). It has been extensively studied and used in different fields. An important concept for the model building process is that the time series data is required to be stationary. It implies that the probabilistic structure of the series does not change over time. The SARIMA is an extension to ARIMA that supports time series with a seasonal component by incorporating seasonal factors into ARIMA model. The SARIMA model is usually denoted as SARIMA( $p, d, q$ ) ( $P, D, Q$ ) $s$ , where  $p$  is the autoregressive order,  $d$  is the number of differencing operations,  $q$  the moving average order and  $P, D$  and  $Q$  are the corresponding seasonal orders. The SARIMA has the following form:

$$\varphi_p(B)\Phi_p(B^s)\nabla^d\nabla_s^D Z_t = \theta_q(B)\Theta_q(B^s)\varepsilon_t, \quad (14)$$

where

$$\psi_p(B) = (1 - \psi_1 B - \psi_2 B^2 - \dots - \psi_p B^p),$$

$\Phi_p(B^s)$  is the seasonal operator of order  $P$ ,

$B$  is the backshift operator with  $B^m(Z_t) = Z_{t-m}$ ,

$s$  is the season length,

$\nabla^d = (1 - B)^d$  is the non-seasonal operator,

$\nabla_s^D = (1 - B^s)^D$  is the seasonal differencing operator,

$Z_t$  is the stationary data at time  $t$ ,

$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \text{ and}$$

$\Theta_q(B^s)$  is the seasonal operators of order  $Q$ , and  $\varepsilon_t$  is the white noise with zero mean and variance.

After transforming the original data into a stationary time series, a model building process included identification, parameter estimation and diagnostic checking. A tentative autoregressive moving average process was developed at the identification stage based on the theoretical model's estimated shape. This process of comparison allowed the definition of  $p$  and  $q$ .

Having specified an initial model, we estimated the unknown parameters in Equation (14) by inspecting the behavior of the auto-covariance function and partial autocorrelation function.

At the final stage, diagnostic checking, we examined the residuals to determine the model's adequacy. If the model is adequate, it was used in the forecasting.

## 4. Data and results

In this section, we first describe the data used for the study. Results from all six forecasting models are reported next. Then, we evaluate the results by the three criteria and compare the forecasting accuracy of the models.

#### 4.1 Time series data on crude oil price

The monthly data on the average crude oil price is collected from January 2015 to December 2019. We split the data into two sets: an in-sample data set for estimation and an out-of-sample data set for prediction. The in-sample data cover the period from January 2015 to December 2018, while the out-of-sample data are from January to December 2019.

#### 4.2 Results for the models

**4.2.1 Grey forecasting.** With the grey model, the forecast values are susceptible to the size of the initial sequence of the time series chosen. Therefore, we conducted the grey forecast five times by varying the initial sequence and found that the lowest prediction errors were obtained when the size of the initial sequence was equal to 4. The following steps demonstrate the process of generating predicted values for the time series by the grey model:

- (1) Class ratio test:

Application of GM(1,1) model requires that the time series data pass the class ratio test first. As mentioned in [Section 3.1](#), the value of  $\sigma(k)$  must fall within the interval between 0 and 1 for the sequence  $x^{(0)}$  to fit the grey model. The results of the class ratio test suggest that the grey model is appropriate for the time series.

- (2) Accumulated generated operation (AGO):

Based on [equation \(2\)](#), we perform AGO to obtain the sequence.

- (3) Mean value generating sequence:

We calculated the mean value generating sequence based on the sequence obtained after accumulated generated operation.

- (4) Time series prediction model

Using the least square method, we obtain estimates for coefficients  $a$  and  $b$  as follows:

$$\hat{a} = 0.1868912999 \quad \hat{b} = 89.6763678173$$

These estimates are used to get

$$\hat{X}^{(0)}(k+1) = \left(1 - e^{\hat{a}}\right) \left[ X^{(0)}(1) - \frac{\hat{b}}{\hat{a}} \right] e^{-\hat{a}k} \quad (15)$$

Based on [equation \(15\)](#), the predicted values of the time series from January 2019 to December 2019 are calculated and presented in [Table 3](#).

**4.2.2 Hybrid grey forecast.** Following the same calculation procedures described in [Section 3.2](#) for the hybrid grey model, we can obtain the predicted crude oil prices from January 2019 to December 2019. The detailed results are also given in [Table 3](#).

**4.2.3 Multiplicative decomposition model.** Based on the multiplicative decomposition model discussed in [Section 3.3](#), we summarize the calculation results for crude oil prices in [Table 1](#).

**4.2.4 Trigonometric regression model.** We estimate [equation \(10\)](#) using statistical analysis system (SAS) statistical software and obtain the following results:



$t$	$Y_t$	12 MA	CMA = $TR_t$ $*CL_t$	$SN_t*IR_t$	$SN_t$	$TR'_t = Y_t/$ $SN_t$	$\hat{TR}_t = a+bt$	$\hat{Y}_t =$ $\hat{TR}_t \times SN_t$
1	47.4				0.982	48.27	41.4572	40.7068
2	50.83				0.959	52.99	41.9014	40.1909
3	47.85				0.985	48.57	42.3455	41.7182
4	54.44				1.029	52.92	42.7896	44.0147
5	59.26				1.066	55.58	43.2337	46.0992
6	59.8				1.032	57.96	43.6779	45.0624
7	51.16	48.759	48.094	1.064	1.010	50.65	44.1220	44.5683
8	42.86	47.429	46.576	0.920	0.951	45.08	44.5661	42.3712
9	45.48	45.723	45.303	1.004	0.977	46.54	45.0102	43.9851
10	46.2	44.883	44.321	1.042	1.022	45.22	45.4544	46.4419
11	42.64	43.759	43.242	0.986	0.992	42.98	45.8985	45.5326
12	37.19	42.725	42.265	0.880	0.995	37.37	46.3426	46.1242
13	31.44	41.804	41.543	0.757	0.982	32.02	46.7867	45.9399
14	30.35	41.282	41.360	0.734	0.959	31.64	47.2309	45.3028
15	37.77	41.439	41.426	0.912	0.985	38.34	47.6750	46.9688
16	40.96	41.413	41.567	0.985	1.029	39.82	48.1191	49.4968
17	46.85	41.721	41.845	1.120	1.066	43.94	48.5633	51.7819
18	48.75	41.969	42.583	1.145	1.032	47.25	49.0074	50.5608
19	44.89	43.198	44.078	1.018	1.010	44.44	49.4515	49.9517
20	44.75	44.958	45.918	0.975	0.951	47.07	49.8956	47.4382
21	45.17	46.878	47.370	0.954	0.977	46.22	50.3398	49.1933
22	49.89	47.863	48.288	1.033	1.022	48.83	50.7839	51.8872
23	45.62	48.713	48.782	0.935	0.992	45.99	51.2280	50.8196
24	51.93	48.851	48.702	1.066	0.995	52.18	51.6721	51.4286
25	52.56	48.553	48.627	1.081	0.982	53.53	52.1163	51.1729
26	53.40	48.701	48.838	1.093	0.959	55.67	52.5604	50.4148
27	49.58	48.974	49.163	1.008	0.985	50.33	53.0045	52.2194
28	51.17	49.353	49.422	1.035	1.029	49.75	53.4486	54.9789
29	48.50	49.492	49.954	0.971	1.066	45.49	53.8928	57.4646
30	45.17	50.416	50.655	0.892	1.032	43.78	54.3369	56.0593
31	46.67	50.894	51.358	0.909	1.010	46.20	54.7810	55.3351
32	48.03	51.823	52.188	0.920	0.951	50.52	55.2251	52.5052
33	49.71	52.553	53.107	0.936	0.977	50.87	55.6693	54.4014
34	51.56	53.660	54.291	0.950	1.022	50.46	56.1134	57.3325
35	56.71	54.923	55.813	1.016	0.992	57.17	56.5575	56.1066
36	57.67	56.703	57.626	1.001	0.995	57.94	57.0017	56.7330
37	63.70	58.549	59.552	1.070	0.982	64.87	57.4458	56.4060
38	62.17	60.555	61.381	1.013	0.959	64.82	57.8899	55.5267
39	62.86	62.207	63.058	0.997	0.985	63.81	58.3340	57.4700
40	66.32	63.908	64.705	1.025	1.029	64.47	58.7782	60.4610
41	69.86	65.503	65.493	1.067	1.066	65.52	59.2223	63.1474
42	67.33	65.483	65.126	1.034	1.032	65.26	59.6664	61.5577
43	70.74	64.768			1.010	70.03	60.1105	60.7186
44	67.85				0.951	71.36	60.5547	57.5723
45	70.13				0.977	71.76	60.9988	59.61
46	70.69				1.022	69.19	61.4429	62.78
47	56.48				0.992	56.93	61.8870	61.39
48	49.09				0.995	49.32	62.3312	62.04
49	51.43				0.982		62.7753	61.64
50	54.91				0.959		63.2194	60.64
51	58.10				0.985		63.6635	62.72
52	63.81				1.029		64.1077	65.94

**Table 1.**  
Predicted crude oil  
price using  
multiplicative  
decomposition model

(continued)

Table 1.

$t$	$Y_t$	12 MA	CMA = $TR_t$ * $CL_t$	$SN_t^*IR_t$	$SN_t$	$TR_t' = Y_t'$ $SN_t$	$\hat{TR}_t = a+bt$	$\hat{Y}_t =$ $\hat{TR}_t \times SN_t$
53	60.78				1.066		64.5518	68.83
54	54.61				1.032		64.9959	67.06
55	57.33				1.010		65.4401	66.10
56	54.76				0.951		65.8842	62.64
57	56.90				0.977		66.3283	64.82
58	53.91				1.022		66.7724	68.22
59	56.92				0.992		67.2166	66.68
60	59.77				0.995		67.6607	67.34

$$\begin{aligned}
 \hat{y} = & 40.66382 + 0.45501 t + 3.57426 \sin\left(\frac{2\pi t}{12}\right) - 0.10327 t \sin\left(\frac{2\pi t}{12}\right) \\
 & - 4.11685 \cos\left(\frac{2\pi t}{12}\right) + 0.05154 t \cos\left(\frac{2\pi t}{12}\right) - 2.18172 \sin\left(\frac{4\pi t}{12}\right) \\
 & + 0.05425 t \sin\left(\frac{4\pi t}{12}\right) + 3.97381 \cos\left(\frac{4\pi t}{12}\right) - 0.15647 t \cos\left(\frac{4\pi t}{12}\right)
 \end{aligned}
 \tag{16}$$

where the number in parentheses indicates the  $t$ -value for the estimated coefficient. In addition,  $R^2 = 0.4775$  and adjusted  $R^2 = 0.3538$ . For the analysis of variance, the  $F$ -value for the overall significance of the model is 3.86 with a  $p$ -value of 0.0015, suggesting that the model is acceptable empirically. Equation (16) is then used to forecast the crude oil price from January 2019 to December 2019. Detailed forecast results are summarized in Table 3.

4.2.5 Seasonal dummy regression. Like the trigonometric regression model, we estimate equation (11) using the SAS statistical package. However, before running the regression, we conducted a logarithmic transformation of the dependent variable. The results of the estimation are reported as follows:

$$\begin{aligned}
 \hat{y}_t = & b_0 + b_1t + b_{s1}X_{s1,t} + b_{s2}X_{s2,t} + \dots + b_{s11}X_{s11,t} \\
 = & 35.03896 + 0.46437t + 4.91305X_{s1,t} + 4.86118X_{s2,t} + 4.72431X_{s3,t} \\
 & + 7.696744X_{s4,t} + 10.39808X_{s5,t} + 9.07871X_{s6,t} + 6.71684X_{s7,t} \\
 & + 3.75997X_{s8,t} + 5.0456X_{s9,t} + 6.54374X_{s10,t} + 1.85687X_{s11,t}
 \end{aligned}
 \tag{17}$$

where, again, the number in parentheses is the  $t$ -value for the coefficient estimate. The  $F$ -value for the overall significance of the model is 2.52 with a  $p$ -value of 0.0166; while  $R^2 = 0.4635$  and adjusted  $R^2 = 0.2795$ . Based on the results reported in equation (17), we calculate the predicted crude oil prices from January 2019 to December 2019. The detailed results are listed in Table 3.

4.2.6 ARIMA model. By examining the historical data, we found that a trend existed in the data. Hence, the original series needs a first differences or other forms of differencing to produce a stationary series. After conducting several differences, we found that the second differences can produce a stationary data for the original data. From Figure 1, we can see that the original data becomes stationary after a second differences.

Once we have transformed the original data into stationary time series, we use the autocorrelation function (ACF) and partial autocorrelation function (PACF) to identify tentative Box-Jenkins models. Three possible models are summarized in Table 2. A good way

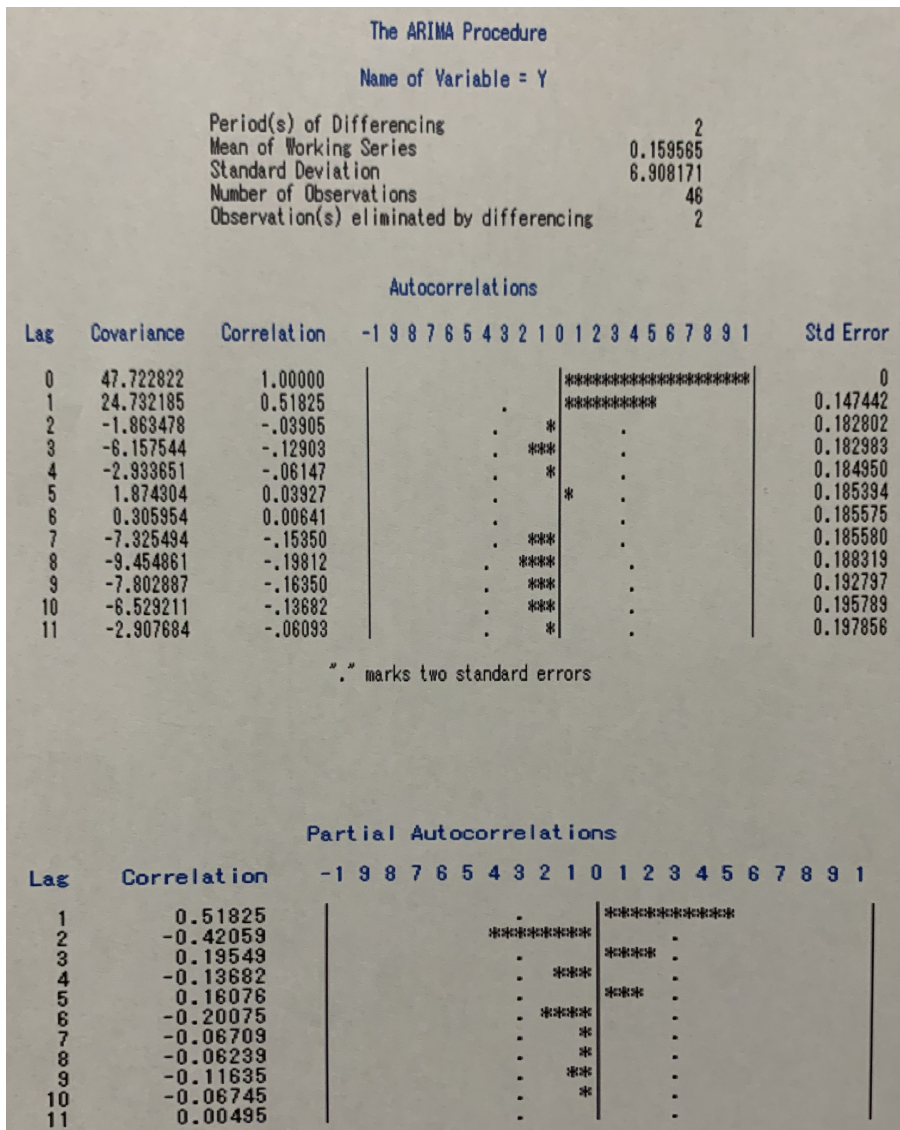


Figure 1. The SAS output of the ACF and PACF for the second differences of the original data

**Table 2.** A summary of the Q\* values and prob-values for the models

	To Lag	Chi square	p-value
Model 1: ARIMA(2,2,0)	6	2.74	0.6016
$Z_t = \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \varepsilon_t$	12	7.14	0.7125
$Z_t$ is the second differences of the original time series	18	9.45	0.8936
	24	13.10	0.9303
Model 2: ARIMA(2,2,0)	6	14.86	0.0110
$Z_t = \psi_2 Z_{t-2} + \varepsilon_t$	12	21.91	0.0251
$Z_t$ is the second differences of the original time series	18	26.77	0.0615
	24	34.41	0.0595
Model 3: ARIMA(2,1,0)	6	13.19	0.0001
$Z_t = \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \varepsilon_t$	12	13.94	0.2364
$Z_t$ is the quadratic roots of the original time series	18	14.57	0.6266
	24	14.69	0.9054

**Table 3.** Actual and predicted crude oil prices Unit: USD/per barrel

	Actual price	Grey forecast	Hybrid grey forecast	Multiplicative decomposition	Trigonometric model	Seasonal dummy regression	ARIMA (2,2,0)
1	51.43	39.86	56.53	61.64	59.40	62.70	52.37
2	54.91	47.32	64.30	60.64	63.65	63.11	51.77
3	58.10	57.91	66.41	62.72	66.18	63.44	51.15
4	63.81	61.80	64.30	65.94	65.01	67.15	54.39
5	60.78	68.44	69.30	68.83	62.26	70.04	55.81
6	54.61	63.55	69.51	67.06	62.09	69.19	53.46
7	57.33	51.15	60.47	66.10	66.23	67.29	56.51
8	54.76	54.11	63.90	62.64	71.80	64.80	55.90
9	56.90	55.71	66.87	64.82	73.85	66.55	58.29
10	53.91	55.89	65.99	68.22	70.29	68.51	59.74
11	56.92	54.35	70.95	66.68	64.32	64.29	54.30
12	59.77	55.93	64.10	67.34	61.52	62.90	50.82

to check the adequacy of an overall Box–Jenkins model is to analyze the residuals obtained from the model. The Ljung–Box statistic from the SAS output is used.

Table 2 provides a summary of the Ljung–Box statistic values and prob-values for the possible models. Using the Ljung–Box statistic values and prob-values to check the model adequacy, we found that both Models 2 and 3 are not adequate since not all  $p$ -values are  $>0.05$ . Hence, it is clear that only Model 1 is adequate since the Ljung–Box statistic values (Chi-Square values) are small, and all  $p$ -values are  $>0.05$ .

The best model identified for crude oil price is ARIMA(2,2,0) and estimated as follows:

$$Z_t = 0.99775Z_{t-1} - 0.56635Z_{t-2} + \varepsilon_t \tag{18}$$

(6.73)                      (−3.84)

where the value in parentheses refers to the  $t$ -value for the coefficient estimate. Equation (18) is used to forecast crude oil price for the ARIMA specification.

#### 4.3 Comparison of forecasting methods

The predicted values of crude oil price for the out-of-sample period from January to December 2019 are computed for each of the six forecasting methods. We summarize the results in Table 3, along with actual crude oil prices for comparison.

Yokum and Armstrong (1995) conducted two studies on experts' opinions on the criteria used in selecting forecasting techniques. They found that accuracy was considered to be the most critical criterion across by most researchers. Since there is no universally accepted measure of accuracy that can be applied to every forecasting situation, several criteria are typically used to comprehensively assess forecasting models. The performance of the models often differs depending on the accuracy measure being used (Makridakis *et al.*, 1982). In this study, we consider three criteria of measuring accuracy commonly chosen to assess the six forecasting models. They are the root mean squared error (RMSE), the mean absolute error (MAE) and the mean absolute percent error (MAPE) defined as follows:

$$MAE = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n} \tag{19}$$

$$MAPE = \frac{100 \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|}{n} \tag{20}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \tag{21}$$

where  $Y_i$  and  $\hat{Y}_i$  are the actual and the predicted values of the time series in period  $i$ , respectively. Obviously, all three measures are positive in value, and the smaller is the value obtained for each of the measures calculated the better the performance of the forecasting method.

The comparative results of forecasting accuracy of the six methods are presented in Table 4. Table 4 shows that the grey forecast is clearly the best forecasting model since it has the lowest values of all three performance measures. The ARIMA appears to be the second best model for forecast accuracy regardless of the three measures used. There seems to be no major significant difference between the results from the hybrid grey forecast, multiplicative decomposition and trigonometric model. On the other hand, seasonal dummy regression is found to be the worst method for predicting crude oil prices.

In the research on the forecast, it is a fact that no single forecasting model is the best for all situations under all circumstances (Makridakis *et al.*, 1982). Nevertheless, from all performance measurements reported above, we found that, in general, the grey forecast outperformed other forecast methods in this study and is a reliable method for forecasting crude oil prices.

Forecasting method \ Accuracy measure	MAE	MAPE	RMSE
Grey forecast	4.53	8.12	5.77
Hybrid grey forecast	8.28	14.78	9.28
Multiplicative decomposition	8.28	14.80	8.86
Trigonometric model	8.62	15.52	10.20
Seasonal dummy regression	8.90	15.94	9.59
ARIMA(2,2,0)	4.99	8.68	5.63

**Table 4.** Performance of various methods of forecasting crude oil prices

## 5. Conclusions

In this paper, six methods, including the grey forecast, the hybrid grey forecast, the classical decomposition, the trigonometric model, the seasonal dummy variables and the ARIMA, have been applied to forecast West Texas crude oil prices based on the monthly data. We compare the predicting accuracy of the models by calculating MAE, MAPE and RMSE for each of the models.

Grey forecast is the best model for crude oil price forecasting. Comparing the results of this study with Lin's (2009) paper, it is pointed out that the grey forecast is based on a small sample, simple calculation and accurate prediction. In Lin's study, the WTI crude oil price prediction in February 2009 with an accuracy of 95.62%. The results are similar to those of this study, which had an accuracy of 98.83, 97.92, 96.45 and 95.50%, in August, September, October and November 2019, respectively.

ARIMA is the second best model for crude oil price forecasting. Although the time series of crude oil prices is seasonal, it is not very obvious. The single variable ARIMA model is based on mathematical and statistical theory and can convert the historical data of oil price into a stationary pattern, which is a reliable forecasting model in capturing the characteristics of time series. The contribution of this study is using a small size of data and comparing the performance of different methods on short-term forecasting. This paper bridged this gap in the application and found a practical and yet highly accurate model.

There is a research limit in our paper. The collapse of oil prices occurred during the time when COVID-19 had halted much of global economic activity. The phenomenon of negative oil prices in 2020 may be the reason that cannot be well explained in all forecasting methods because it is affected by the epidemic situation, other human factors and economic measures to prevent a severe economic downturn.

In the future, if the research focuses on explaining the causal relationship between variables, a multivariate approach can be used since many factors affect the crude oil price. Furthermore, it may be worthwhile to combine and explore other forecasting methods, such as neural networks, artificial intelligence or advanced data mining techniques, to predict crude oil prices.

## References

- Ahmad, M.I. (2011), "Modeling and forecasting Oman crude oil prices using Box-Jenkins techniques", *Society of Interdisciplinary Business Research (SIBR) 2011 Conference on Interdisciplinary Business Research*.
- Bashiri Behmiri, N. and Pires Manso, J.R. (2013), "Crude oil price forecasting techniques: a comprehensive review of literature", available at: <https://ssrn.com/abstract=2275428>.
- Beckers, B. and Beidas-Strom, S. (2015), *Forecasting the Nominal Brent Oil Price with VARs—One Model Fits All?*, 2015, International Monetary Fund, IMF Working papers, available at: <https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Forecasting-the-Nominal-Brent-Oil-Price-with-VARs-One-Model-Fits-All-43423>.
- Bowerman, B.L. and O'Connell, R.T. (1993), *Forecasting and Time Series: An Applied Approach*, 3rd ed., Duxbury Press, CA.
- Box, G.E.P. and Jenkins, G.M. (1976), *Time Series Analysis, Forecasting and Control*, 2nd ed., Holden Day, San Francisco.
- Deng, J.L. (1989), "Introduction grey system theory", *Journal of Grey System*, Vol. 1, pp. 1-24.
- Duan, H., Lei, G.R. and Shao, K. (2018), "Forecasting crude oil consumption in China using a grey prediction model with an optimal fractional-order accumulating operator", *Complexity*, Vol. 2018, pp. 1-12.

- Faruk, B.U. (2018), *An Application of Box-Jenkins Approach in Forecasting Nigeria Crude Oil Prices*, Ph. D. Thesis, Department of Economics, Umaru Musa Yaradua University, Katsina.
- Gupta, N. and Nigam, S. (2020), "Crude oil price prediction using Artificial Neural Network", *Procedia Computer Science*, Vol. 170, pp. 642-647.
- Huang, S.P., Zhang, L., Wang, M.H., Wang, R. and University, C.J. (2016), "Analysis of global oil consumption based on grey prediction model", *Science and Technology Innovation Herald*, Vol. 34, pp. 118-121.
- Lin, A. (2009), "Prediction of international crude oil futures price based on GM(1,1)", *2009 IEEE International Conference on Grey Systems and Intelligent Services (GSIS 2009)*.
- Lyu, H. and Chang, Y. (2017), "Research on international crude oil price forecasting model", *International Journal of New Developments in Engineering and Society*, Vol. 1 No. 3, pp. 78-81.
- Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982), "The accuracy of extrapolation (time series) methods: results of a forecasting competition", *Journal of forecasting*, Vol. 1 No. 2, pp. 111-153.
- Mensah, E.K. (2015), *Box-Jenkins Modelling and Forecasting of Brent Crude Oil Price*, MPRA Paper No. 67748, posted 16 Mar 2016 00:14 UTC, Munich Personal RePEc Archive, Munich.
- Norouzi, N. and Fani, M. (2020), "Black gold falls, black plague arise -an OPEC crude oil price forecast using a gray prediction model", *Upstream Oil and Gas Technology*, Vol. 5, 100015.
- Nwafor, C.N. and Oyedele, A.A. (2018), "Forecasting OPEC oil price: a comparison of parametric stochastic models", *European Journal of Business and Management*, Vol. 10 No. 10, pp. 49-60.
- Peng, W.Y. and Chu, C.W. (2009), "A comparison of univariate methods for forecasting container throughput volumes", *Mathematical and Computer Modeling*, Vol. 50, pp. 1045-1057.
- Sadorsky, P. (2006), "Modeling and forecasting petroleum futures volatility", *Energy Economics*, Vol. 28, pp. 467-488.
- Selvi, J.J., Shree, R.K. and Krishnan, J. (2018), "Forecasting crude oil price using ARIMA models", *International Journal of Advance Research in Science and Engineering*, Vol. 7 No. 5, pp. 334-343.
- Xiang, Y. and Zhuang, X.H. (2013), "Application of ARIMA model in short-term prediction of international crude oil price", *Advanced Materials Research*, Vols 798-799, pp. 979-982.
- Xu, R.S. (2010), "Research on China's oil consumption demand forecast based on Grey System Theory", *Statistics and Decision*, Vol. 320, pp. 98-101.
- Yaziz, S.R., Ahmad, M.H., Nian, L.C. and Muhammad, N. (2011), "A comparative study on Box-Jenkins and Garch models in forecasting crude oil prices", *Journal of Applied Sciences*, Vol. 11 No. 7, pp. 1129-1135.
- Ye, M.T., Zyren, J. and Shore, J. (2005), "A monthly crude oil spot price forecasting model using relative inventories", *International Journal of Forecasting*, Vol. 21, pp. 491-501.
- Yokum, J.T. and Armstrong, J.S. (1995), "Beyond accuracy: comparison of criteria used to select forecasting methods", *International Journal of Forecasting*, Vol. 11 No. 4, pp. 591-597.
- Zhao, C. and Wang, B. (2014), "Forecasting crude oil price with an autoregressive integrated moving average (ARIMA) model", in *Fuzzy Information and Engineering and Operations Research and Management*, pp. 275-286.

### Further reading

- Aamir, M. and Shabri, A.B. (2015), "Modeling and forecasting monthly crude prices of Pakistan: a comparative study of ARIMA, GARCH and ARIMA-GARCH Models", *Science International*, Vol. 27 No. 3, pp. 2365-2371.
- He, X.J. (2018), "Crude oil prices forecasting: time series vs. SVR models", *Journal of International Technology and Information Management*, Vol. 27 No. 2, Article 2.

- Safari, A. and Mohammadi, M. (2017), "Assessing linear and nonlinear models to forecast OPEC oil prices", *Revista QUID*, (Special Issue), Vol. 1 No. 3, pp. 13-20.
- Tularam, G.A. and Saeed, T. (2016), "Oil-price forecasting based on various univariate time-series models", *American Journal of Operations Research*, Vol. 6, pp. 226-235.
- Yu, F., Liu, Y. and Zhang, C. (2021), "Forecasting the price of fuel oil: a STL-(ELM+ARIMA) combination approach", *Journal of Physics: Conference Series*, Vol. 1903, 012048.

**Corresponding author**

Ching-Wu Chu can be contacted at: [cwchu@ntou.edu.tw](mailto:cwchu@ntou.edu.tw)

---

For instructions on how to order reprints of this article, please visit our website:

[www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)