Journal of Econometrics and Statistics
Vol. 4, Issue 1, 2024, pp. 15-51
https://doi.org/10.47509/JES.2024.v04i01.02

FORECASTING CRUDE OIL MARKETS

Miss Jiaqi Li,[1] Dr. Azzam Alroomi,[2] and Professor Dr. Konstantinos Nikolopoulos[3]

[1] *https://www.linkedin.com/in/jiaqi-li-681734a2/*
[2] *aalroomi@aou.edu.kw*
[3] *kostas.nikolopoulos@durham.ac.uk*

In this research daily crude oil data from U.S, Energy Information Administration from 2000-2019 is explored to test the forecasting accuracy by drawing the comparison between multiple models. Forecasting models discussed in the research cover regression, artificial neural network (ANN), exponential smoothing (ES), and autoregressive integrated moving average (ARIMA). We primarily aim to determine which mode provides the optimal forecasting results for WTI and Brent market, two major international light oil markets. The data is split into training, validation, and testing parts, with different purposes of modelling. Based on the adopted evaluation metrics, ARIMA model exhibits the optimal performance in validation data for both markets; while seasonal exponential smoothing model achieves the best 10-day and 20-day ahead forecasting.

Keywords: Crude Oil Forecasting; Regression; Neural Networks; Exponential Smoothing; ARIMA;

1. Introduction

Crude oil is vital to global economics. One of the observations is that uncertainty in oil price has directly led to recession in 1980 and 1982 (Elder and Serltetis, 2010). Also, nearly two thirds of the world energy demands are met from crude oil (Slim, 2015). In recent years, it has been evidenced that the large fluctuations in oil market has depressed current investment (Wang, Wu and Yang, 2016). Due to the significance of crude oil, forecasting crude oil arouses huge attention from scholars and researchers; nevertheless, forecasting oil price is a challenging research subject. One of the challenges refer to the uncertainty in oil price, since several factors can affect crude oil price. The internal factors consist of inventory, supply, and demand; in the means time, the external factors involve weather, policy, and wars etc. The variations of these factors make the oil prices unpredictable. Previously, conventional statistical and econometric techniques are popular in oil forecasting (Slim, 2015). Since 1973, scholars considered that the linear relation between price and economy has diminished (Elder and Serletis, 2010). The nonlinearity of crude oil price is attributed to the uncertainty (e.g., transaction cost and the absence of riskless arbitrage) (Fattouh, 2009). Over the years, several representative oil forecasting models include linear regression, Generalized Heteroscedasticity (GARCH), and vector auto-regression (VAR). GRACH model has been employed extensively to characterize the dynamics of oil return (Wang, Wu and Yang, 2016). Also, ARMA is a classical linear model; GARCH and Exponential Smoothing (ES) model can be expressed as ARMA-type model. (Bollerslev and Mikkelsen, 1996). In recent studies, certain advanced and artificial intelligence models outperform conventional models in forecasting than, since these models can characterize the non-linear relation in the oil market. Artificial Neural Network (ANN), a representative AI model with excellent performance in forecasting short term crude oil and economic index (Kulkarni and Haidar, 2009). However, each model has its own weaknesses. Most GARCH-class model only captures the short memory; ANN's forecasting result is difficult to test; Markov switching multifractal (MSM) parameter k's value is setting given user's experience. The challenges of forecasting crude oil is not only comes from model selection, but also the uncertainty of price pattern. There is a continuous debate in crude oil dynamics, one perspective is that world oil market is one great pool; the other perspective says that oil market is globalized, of which price shocks is transferring from the market to the next (Fattouh, 2009). This research aims to find appropriate forecasting models for WTI and Brent markets. In this research, three types of models, traditional, more advanced, and artificial intelligence, are attempting for justify the optimal model. There are rare researches that are covering all three types of models. Also, both univariate and multivariate models are taken to fit daily light crude oil spot prices from 2000 to 2019. Regression and neural network models are taken as multivariate forecasting models; exponential smoothing and ARIMA models are taken as univariate forecasting modes for WTI and Brent oil prices. To build the

Article History
Received : 10 January 2024; Revised : 11 February 2024; Accepted : 22 February 2024; Published : 15 April 2024

To cite this paper

Miss Jiaqi Li, Dr. Azzam Alroomi, & Professor Dr. Konstantinos Nikolopoulos (2023). Forecasting Crude Oil Markets. *Journal of Econometrics and Statistics*. 4(1), 15-51.

model with high forecasting accuracy, the crude oil data is split into training, validation, and test dataset. Training data helps the user identify the parameters for each model, and validation dataset helps the user to assess the fitted model's performance. The evaluation metrics consist of MSE, RMSE, MAE, and AIC. The optimal forecasting model for Brent and WTI oil markets should exhibit the lowest error rate in evaluation metrics. Finally, testing data assesses the model's performance in 10 and 20 day-ahead forecasting. Based on the past papers, there are few researches that focusing on predicting 10 and 20 day-ahead oil price based on daily crude oil price; scholars are very interesting in forecasting 1 day-ahead price. In the research, the various types of forecasting models are discussed in literature review. Then, the methodology and model selection process is following literature review. Modelling section presents the model's parameter selection and adjustment process. Lastly, the model performance and conclusion are drawn based on the taken forecasting model.

2. Literature Review

Oil has acted as the crucial source of energy for human since the 19th century. Compare to modern society, oil is even more important to people's life, since machinery, transportation, and heating are three knowable usages of oil. In the past few decades, economists found the unstable crude oil price adversely affects the country's economy, especially in gross domestic product (GDP), employment rate. In Hamilton (1983) research, seven of the eight post-war recessions in the U.S. have been preceded by a surge in the price of crude oil. From the economics perspectives, the giant oil fluctuations' adverse effects are vital to several measures of investment, durable consumption, as well as aggregate output (Elder and Serletis, 2010). The fluctuations in the oil market affect the demand and supply of other energetic comedies, so does their prices. Also, the fluctuant oil price negatively impacts each country in the world, since the fluctuations impact both the oil export countries' economy stability and the import countries'. Thus, forecasting crude oil's demand and volatility is of rising implication for economy reasons, especially the wide variations of price in recent years.

2.1. Crude Oil

1973 is a breakeven point in crude oil history, since the relation between oil prices and economics has changed. Hooker (1996) defined the linear relation is diminished after 1973, and three critical reasons to explain the oil prices is no longer Granger cause. They are sample stability, oil prices are endogenous, and linear and symmetric misrepresent the form of the oil price interaction. Before 1973, the U.S. had the dominated power in controlling crude oil price. From 1973 to 1979, OPEC (Organization of the Petroleum Exporting Countries) increased crude oil price, and the price was rising from $3.4 to 12$ per barrel. The oil shock somehow affected the U.S. macroeconomic and caused its deep recession from 1974 and 1975. In 1985, Saudi Arabic quitted the OPEC, and the world crude oil production was up-regulated from 2 million barrels per day to 5 million barrels per day. This event directly down-regulated the world crude oil price. Since then, the model's ability to keep stable during structural break period has been increasingly significant. Salisu et al. (2012), defined recent two structural breaks occur in 1990 and 2008, corresponding to the Iraqi/Kuwait conflict and global financial crisis. Other recent notable price fluctuations cover price increase due to OPEC curtailed the production of crude oil b 4.2 million per day from 2000 and 2001 and North Korean missile launches in 2017.

2.2. Dynamics Patterns in Oil Market

Depending on crude oil's quality, oil is classified as light and heavy oil. Light oil exhibits higher quality and a higher share of light hydrocarbons, and a simple distillation can produce it; in contrast, the heavy oil exhibits lower quality since the higher sulphured contents, and it requires extra processing to produce products (Lanza, Manera and Giovannini, 2005). One factor causing price differential in the oil market is the quality since the extra refining process cost reflects the oil price. Oil markets' dynamics are closely related to oil quality. The three major oil markets (Brent, Dubai, and West Texas Intermediate (WTI) Cushing), are not only split by region but also the quality. Both Brent and WTI, the European benchmark and US benchmark individually, are light oil markets; Dubai, the Persian Gulf benchmark, is heavy oil market (Lanza, Manera, Giovannini, 2005). Among the three primary markets, WTI exhibits highest quality, followed by Brent and Dubai (Fattouh, 2010).

There are three essential perspectives regarding crude oil market dynamics. Adelman's theory compares the world oil market with the world ocean, saying that the feature is one great pool (1984). Each country should share the same level of scarcity due to a shortage of oil supply. Accordingly, the change of price is linked. However, other scholars have the opposite perspectives to those of Adelman. They consider the oil market is globalization; the scarcity in one region will transfer to other regions (Gulen 1997, 1999 & Weiner, 1991). This is because different countries hold various attitudes to the crude oil, so regional policy is crucial to impact the oil price in a specific area. One implication of globalization is the price of crude oil with similar quality should move closer together. Finally, Gulen (1999) criticized the globalization thesis and considered the world oil market is regionalized, oil prices move independently of each other in response to local market conditions and regional shocks.
Fattouh's (2009) finding, suggests the different oil markets, Brent, Dubai, and WTI, are linked. For the presence of transaction costs and absences of riskless arbitrage, the market relation is non-linearity. Due to the non-linearity and non-stationary characteristics of crude oil prices, most of the conventional methods cannot provide accurate results in forecasting oil prices (Li et al., 2018). Numerous scholars asserted that the characteristics of crude oil are nonlinear, complexity, and seasonality. Ling, Tang et al. (2013), summarized that the seasonality can be explored by one-year time scale, and it is a particular case of a cyclical pattern. The reasons for the seasonality is numerous and complicated; they are temperature, climate, as well as weather.

2.3. Forecasting Oil

Oil forecasting model can be grouped into three broad types: structural, linear, and nonlinear time series model (Moshiri and Foroutan, 2006). A structural model is not as conductive as linear and nonlinear time series model in forecasting crude oil prices. Pindyck (1999) explained the reason for this is that structural models provide some reasonable explanation of the underlying factors of demand and supply changes, but they are generally not successful in predicting oil prices. Also, the conventionally linear structure model is not as prioritized as a linear or nonlinear time series model in forecasting crude oil volatility, since dynamic pattern of oil is being complex. In contrast, linear and nonlinear models (e.g., ARCH class and ARMA models) are more accurate in forecasting oil volatility.

ARIMA
An autoregressive integrated moving average (ARIMA) model is a generalized from autoregressive moving average (ARMA) model in time series analysis. The 'I' in ARIMA refers to differentiation of the modelling. The function representing the ARIMA model is denoted ARIMA (p, d, q). The origin of the ARMA model is auto regression (AR) model and moving average (MA) model. Ahmed and Shabei (2014) has summarized the advantage of ARIMA model is showing the regression error is a linear combination of error term whose values occurred contemporaneously and at various times in the past.
Linear time series models, ARIMA (Box and Jenkins, 1976) helps characterize linear characteristic of crude oil volatility, an ideal model theoretically. It requires only the historical time series data under forecasting. In other words, the future value of the variable in ARIMA is assumed as a linear combination of past errors. Data transformation is usually vital for the classification step to ensure the accuracy of ARIMA, Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) is recommended to be applied before being fitted into ARIMA model to recognize the data trending (Ahmed and Shabei, 2014). Also, PACF and ACF are adopted to identifying the models.

The challenge in model identification has stated by Yusof, Rashid, and Mohamed (2010), the ACF and PACF are random variables, which will not give the same picture as the theoretical function. In Ahmed and Shabei (2014) study, ARIMA model has better forecasting performance under two measure of error, MAE and RMSE.

Exponential Smoothing Forecasting crude oil prices can be split into three major types, conventional time series, more advanced time series (e.g., ARIMA and artificial intelligence or machine learning models). The representative conventional time series models are simple exponential smoothing (SME). Exponential Smoothing (ES) is one of the trending models used to forecast the real-time data, exhibiting the advantages of getting operative, simple, and accurate results (Tularam and Saeed, 2016). Some ES models are specific cases of the ARIMA model. For instance, two parameters ES is equated with ARIMA (0, 2, 2); simple ES is equated with ARIMA (0,1,1). The only exception is that the three-parameter Holt-Winters model does not have the equivalent ARIMA model (Chatfield and Yar, 1988). Though ES is a simple model, its result is comparable to those complex models, and it fits well in the horizontal data pattern. Therefore, transfer the nonstationary time series data to stationery can improve to model's accuracy. Theoretically, the training data is finding the parameters by the sum of squared one-step-ahead prediction error criterion (Billah, King, Snyder, and Koehler 2006). Similar to ARMA class models, the advantage of ES models is weighted and summed past observations; the data a long time ago is allocated with a small value of weight. In ES class models, no seasonal effect, additive seasonal effect, and multiplicative seasonal effect are the three major types of computing trend. The additive seasonal ES model calculates the amount of adjustment constantly; in contrast, multiplicative seasonal effect indicates that amount of adjustments is varied (NCSS Statistical Software). The following ES models are extensively used to forecast crude oil and economic data.

1. Simple Exponential Smoothing (SES): the basic model in the ES family and a local level model. Simple ES has been commonly used for short-range forecasting, usually just one month into the future; whereas, it does not consider the seasonality in time series model. Though simple exponential smoothing is usually considered a basic model in forecasting, it can provide reasonably acceptable forecasting accuracy with minimum computation complexity (He, 2018).

2. Double exponential smoothing (DES): it is also known as Holt's linear exponential method. It has been extensively applied to data that shows a trend, the long-term increase or decrease. Suppalakpanya et al. (2019) experimentally found that DES has the smallest MAPE in forecasting crude palm oil since the palm oil price indicates rising trend. In other words, DES can forecast the crude oil price accurately either in upward or downward trend.

3. Holt-Winters (HW) is a more advanced ES family models than models mentioned previously, since it considers the seasonality factor that may impact the crude oil volatilities. In Albalawi and Alanzi (2015)'s research, HW ES is applied after they detect the seasonality in data.

Nonlinear Models

Crude oil volatility is a complicated and dynamic series, so generating a flexible nonlinear and local optimized model will be more appropriate and realistic. One evidence of nonlinear relation in crude oil prices is Chaos theory, interested and investigated by economists under the market crash of October 1987. In Chaos theory, the complex behaviour of economic series appears randomly, probably explained by a deterministic nonlinear system. Before 1973, the crude oil price was forecasted by the linear regression. However, Hooker (1996) defined three critical reasons to explain the oil prices is no longer Granger cause many US macroeconomics indicator variable in data after 1973. The sample is stable, the oil prices is endogenous, linear and symmetric, which wrongly reflects the form of the oil price interaction. Accordingly, part of scholars is prioritized ANN models to forecast oil volatility instead of ARCH class models.

Nonlinear time series models are remarkable in their prediction accuracy under the chaos theory. In the past academic papers, scholars primarily focused on three types of a forecasting model to forecast crude oil fluctuation; they are a neural network, SVR, GARCH, and MSM (Markov switching multifractal) volatility model. In the recent researches, scholars agree that different type of model has its strength and weakness, so combined forecasting models are trending for recent studies. The forecasting models introduce below are suitable for forecasting dynamics in the oil market.

Neural Networks

Artificial neural network (ANN) is considered as a method to forecast crude oil volatility, since its nonlinear methods that mimics human brain and memory. (Wang and Wang, 2016). The major benefit of applying ANN is that it can learn all types of data and forecast with a reasonable accuracy (Moshiri and Foroutan, 2006). One feature of energy market is chaos (Adrangi and Chatrath, 2001& Panas and Ninni 2000), this means the market are dynamic and unstructured, so finding a locally optimal forecast will be more realistic. In accordance with chaos theory, the nonlinear ANN-family models should exhibit better accuracy than structural and linear forecasting models. The following ANN models are representative in oil forecasting research.

1. The backpropagation neural network (BPNN) has a powerful problem-solving ability. This model is commonly used for forecasting short term crude oil. Depends on the selected data, some scholars proved that three layer perceptron has optimal forecasting accuracy in oil price; the more advanced BPNN model employed five neurons that represents five different trends in the oil market (Aloui, 2015).

2. Multilayer perceptron (MLP) can capture complex relation between inputs and outputs; neurons grouped into layers of different levels (Herrera et al., 2019). Three types of layers, input layer, hidden layer, as well as output layer. MLP can have more than one hidden layer. This ANN based model obtains the major idea. In Moshiri and Foroutan(2006)'s study, they used the MLP model to forecast oil volatility, and then compared forecasting error (MSE, RMSE, MAE) with ARIMA and GARCH models. MLP has the lowest error rate among three models.

3. Feedforward network (FFNN): Recurrent neural networks, ERNN, a special type of recurrent neural networks, acquires the recent event to forecast future output (Wand and Wang, 2016). In Kulkarni and Haidar's research (2009), FFNN was applied for short-term crude oil forecast, with the rise in forecast horizon, the predict accuracy is decreasing. A three-layer FFNN is first set up as the benchmark. Besides, ANN model has the lowest MSE, MAE, RMSE compared with GARCH and ARMA.

Among several type of neural network, Wang and Wang (2016) combined the Multilayer perception and Elman recurrent neural network (ERNN) model, termed as ST-ERNN. The motivation of ST-ERNN is by capturing both of historical and recent crude oil data to prevent losing of meaning information from past to recent. Since the structural break may exist in the historical data rather than recent data, a long memory model is vital to enhance prediction accuracy in out-of-sample data. Compared with BPNN, ERNN's result is closer to real data from 1990s to today, a noticeable fluctuation period (Wand and Wang, 2016).

However, the concern raised by using ANN is that it may be extreme useful in particular scenarios, but not generally apply to others scenarios. Also, the data's stationary should be examined in ANN model. If data is non-stationary, the transformation process can up-regulate the ANN model's in-sample accuracy and out-of-sample accuracy. Besides the neural network, other nonlinear time series models should be employed as well. As Moshiri and Foroutan (2006) mentioned in their research (e.g., Threshold AR, Exponential AR, or Generalized AR).

SVR

The support vector regression (SVR) model is another popular machine learning model for oil forecasting. Support vector machines (SVM) was first developed for classification case; SVR is later developed for regression case (Yasin et al. 2016). Compared with ANN model, SVR model has several similar aspects to ANN's. First, both are easy to overfit. Second, they are not statistical based model, so they cannot fit test data based on parameters estimated from the training data (He, 2018). One thing that SVR is better than ANN is that SVR can overcome the overfitting, the serious challenge in machine learning models. Evidence can be found in Yasin et al. (2016) research, they found the

SVR's forecasting result is following the same pattern of the actual oil price.

Family of ARCH Models

Family of ARCH is popular and extensively used model to forecast oil return volatility. Engle (1982) introduced The ARCH (Autoregressive Conditional Heteroskedastic) model. It was defined as mean zero, serially uncorrelated process with non-constant variance conditional on the past, but constant unconditional variance. In contrast to ARIMA model, ARCH model can capture the heteroscedastic outcomes of a time series procedure, which ARIMA mode is not able to capture. In practical application, ARCH has been widely adopted in macroeconomic field, like estimates of the variance of United Kingdom inflation. The models that list below are members of ARCH family model, and they are commonly used for forecasting oil price.

1. GARCH: Under the effect of ARCH model, the generalized ARCH model was first presented by Bollerslev (1986). Compared with the ARCH model, GARCH exhibits more flexible lag structure (Bollerslev 1986). GARCH model is a trending method to forecast oil volatility. In Moshiri and Foroutan (2006), they selected GARCH as the prior non-linear model.

2. IGARCH: In Kang et al. (2009), CGARCH and FIGARCH exhibit better out-of-sample forecasting accuracy than GARCH and IGARCH; FIGARCH exhibits high forecasting accuracy in Brent and Dubai oil price, CGARCH is committed to WTI.

3. FIGARCH: Baillie et al. (1996) presented a fractionally integrated GARCH (FIGARCH) model that allowing for a factional integrated process in conditional variance.

4. CGARCH: Engle and Lee (1999) presented the component-GARCH (CGARCH) models distinguishing between short-run and long-run persistence of volatility.

Weakness of ARCH class

However, the family of GARCH model has 2 limitations in forecasting crude oil volatility. First, most of GARCH-class models only captures the short memory and they are unable to capture structural break. As mentioned in above, structural break is an infrequent shift or shock without information. FIGACH, with a long memory and the remarkable prediction accuracy, seems to be a fiction due to unaccounted structure break. GARCH models cannot identify the timing of structural shift and appearances of unstructured shift (Lamoureux and Lastrapes, 1990). To address such problems, the model should be combined with current economic context, e.g., conducting risk-premium analysis (Lamoureux and Lastrapes, 1990).

In the recent studies, AFIGARCH (Adaptive FIGARCH) has been used to capture crude oil's long memory and structural break characteristics; however, if the sample period does not cover structural break, the model's accuracy will not be as conductive as other ARCH class's models. The unfavourable sample performance can be explained by AFIGARCH easiness of overfitting, since it requires considerable parameters to be estimated. Due to estimates, the result of the model is not easy to interpret. AIIGARCH is similar to neural network. Overfitting is because model's in-sample performance is better than out-of-sample performance, i.e., the lower forecast power in the real-world case. Second, GARCH-class models not able to multiscale data. GARCH model are always associated with dynamics of squared returns, so family of GARCH model fails to capture the multivalent feature of crude oil volatility.

Markov Switching Multifractal

After discovering limitation on GARCH-class model, scholars proposed a novel model, Markov switching multifractal (MSM). Theoretically, benefits of applying MSM approach are the capturing of the stylized facts of crude oil fluctuations, which are multi-fractality or multi-scaling, long memory, as well as structural breaks (Wang et al., 2016). Multifractal process is a recent formalism to model the time series of returns in finance, and its characteristics include hyperbolically decaying auto-covariance (long memory) and fat tails, implying different powers of the measure(Lux, 2012). Univariate MSM can be extended to multivariate ones, which can model the covariance between 2 different asset returns.

Empirically, Wang et al. (2016) indicated that MSM model is more accurate both in-sample result and out-sample result than GARCH or historical volatility models. However, the ongoing discussion about MSM is how to define the value of k. theoretically, higher value of k gives the favourable model confidence. In Lux (2012)'s finding, the general method of moments (GMM) approach was proposed to estimate multifractional parameters, which can be used in case which Maximum likelihood (ML) is not applicable or computationally infeasible.

2.4. Characteristics of Crude Oil Data

Knowing that crude oil price refers to a time series data. In Kalekar(2004)'s study, he defined the seasonality in time series is the tendency of time-series data to exhibit behaviour that repeats itself L periods. In time series modelling, whether the target value is stationary or not should be checked by plotting the line or scatter chart. Most types of the time-series data are non-stationary data, so does crude oil spot price. Ahmed and Shabri (2014) summarized that the characteristics of stationary data are stable mean and autocorrelation. If data is non-stationary, the transformation will be necessary in data pre-processing step for number of models (e.g., family of ARMA and ES models).
Spot oil price data is extensively used to forecast crude oil volatility, and its volatility reveals the volatility of current and future values of oil production consumption and inventory demand (Pindyck, 2004). Some scholars highlighted the relationship between crude oil price and stock market prices (Wand and Wang, 2016); others stressed the relation between spot prices and future price. Kulkarni and Haidar (2009), used commodity future price to forecast spot price. Compare with crude oil spot price, the superiority of future price is that reacts faster to new information, since the characteristics of commodity are more favourable by investors, including low transaction costs, high liquidity and low cash in up-front. On the other hand, the pressing concern of applying future price to forecast spot oil price is whether the two prices are correlated or not. Numerous of journals has stated the low correlation between future price and spot oil price; some scholar stated the relationship depends on periods (Minimol, 2018). Those scholars who realize the weak correlation may exist between future and spot oil price still insist on using the future price to forecast oil price. Another motivation for seasonal testing is the long duration of spot oil prices; As a result, seasonal visual examination may not be accurate sufficiently. This is primarily because they believe future price can influence oil price at least in some degree, since high commodity price impact the crude oil spot price by weaken buyer's purchasing power and economic growth.

3. Methodology

The literature review suggest that in the economic filed, more and more advanced prediction models are developed used to predict the volatility of crude oil. Scholars believe that complicated models exhibit better forecasting accuracy for recent crude oil data, but they also agree that different advanced model's prediction power might heavily on the data selection, pre-processing, and evaluation metrics. From this perspective, it is arduous to judge whether complex or classical type of model is fitting better in data that has been selected in this paper. In other words, this research is aiming to compare both classical and complex models forecasting performance, and their performance in 10 and 20-day ahead forecasting. The following section is discussing the motivation for selecting the models and their parameters. The reason that testing data is relatively small range of data due to the high fluctuation in the crude oil price, a short forecasting period can reflect what is going on currently.

3.1. Model Selection

3.1.1. Multivariate Model versus Univariate Model

The two types of time series forecasting model are multivariate and univariate models. Univariate model forecasts the oil volatilities by analysing the historical price. The single input of forecasting model is spot oil price. The advantage of the model is it provides more accurate results than multivariate model and tends to reflect changes in the macroeconomic environment more rapidly. In the economic modelling, exchange rate forecasting, univariate time series fits better than multivariate time series for most currencies (Azubuike and Kosemoni, 2017). Crude oil price is related to economic activity, so the economic forecasting models inspire model's selection process in crude oil forecasting. Besides, the weakness of univariate model is that it does not consider the relevant factors that impact price (Tularam and Saeed, 2016).

Different than univariate forecasting model, multivariate models always contain numerous of variables correlated with crude oil price. Gabralla and Abrahma (2013) summarized five driving factors affecting the crude oil price, including gold, demand of the international oil, supply of international oil, political factor, and natural disaster/oil for heating. These variables can act as inputs in multivariate oil forecasting models. Some popular multivariate models are ANN, VAR, regression models, since these models are prioritized to have sufficient inputs and then capture the relation between input and targets.

Though multivariate model covers number of important factors affecting crude oil price (e.g., GDP, supply and demand), it increases the difficulty to explain the result and variables. Moreover, factors (GDP and Consumer Price Index (CPI)) does not match the crude oil data daily frequency. In other words, if user selects daily crude oil data as input for multivariate model, the oil driving factor (e.g., GDP) cannot be considered. From this perspective, multivariate models are generating less convincible results than univariate models.

In this research, both univariate and multivariate models are adopted. The multivariate models' inputs are calculated from daily spot oil prices. The motivation of selected daily crude oil data as input is stating in the next section. The methodology of converting univariate to multivariate follows the Azubuike and Kosemoni's research (2017), which they convert the daily exchange rate to average monthly exchange rate when fitting a multivariate time series model.

3.2. Crude Oil Data

3.2.1. Data, range frequency

User can collect the data purposefully after defining the targeting model types. Data's selection process is associated with research aim, which is to find the optimal model in a short forecasting period. Though there are various types of crude oil in the market, this research will only focus on two major prices that exhibit high level similarity, WTI and Brent. These two markets present high level of similarity in crude oil quality. Crude oil in WTI and Brent are recognized as higher quality (light and sweet) than Dubai. Since the oil prices are depending on its quality, the assumption made in data selection process is WTI and Brent's prices and its fluctuations are similar. When data's feature is similar, the identical forecasting models can fit into data, and the forecasting results can be comparable. Later in the descriptive analysis, this assumption is being validated.
WTI and Brent's spot oil prices are collected from U.S. Energy Information Administration (EIA). Compared with commodity future prices, spot oil prices reflect the variations in the macroeconomic environment more rapidly. Data's frequency can vary with research purpose and model type. Some scholars employ the weekly data, others use daily or intraday data. For a short-term forecast, the daily or intraday data is prioritized; weekly and monthly data are prioritized for other forecasting horizons, since it is less noisy (Kulkarni & Haidar, 2009). Thus, the daily spot oil price data of WTI and Brent will be used since the objective of research paper is short time forecasting.
The line chart in Figure 1 downloaded from EIA has demonstrated the crude oil fluctuation from 1986 to 2018 in unit of dollar. From 1986 to 2000, price was less fluctuation; price became unstable after 2000, and showed dramatic up and down these years. Based on visual inspection of line chart, non-stationarity and uncertainty pose serious threat in crude oil forecasting, especially for data after

year of 2000.
Range and frequency are two crucial factors affecting the model's performance. Data's frequency has been taken as daily based, and the next step is to define range. Based on the pattern of fluctuations, the range of selection is from 2000 to 2018, the wavier period is in the Figure 1. Data visualisation helps understand the data. Thus, another line chart is created by SAS Enterprise Miner; Figure 2 captures WTI and Brent's daily price from 2000 to 2019.

FIG. 1. Annually Spot Price of Crude Oil in US dollar (1986-2018)

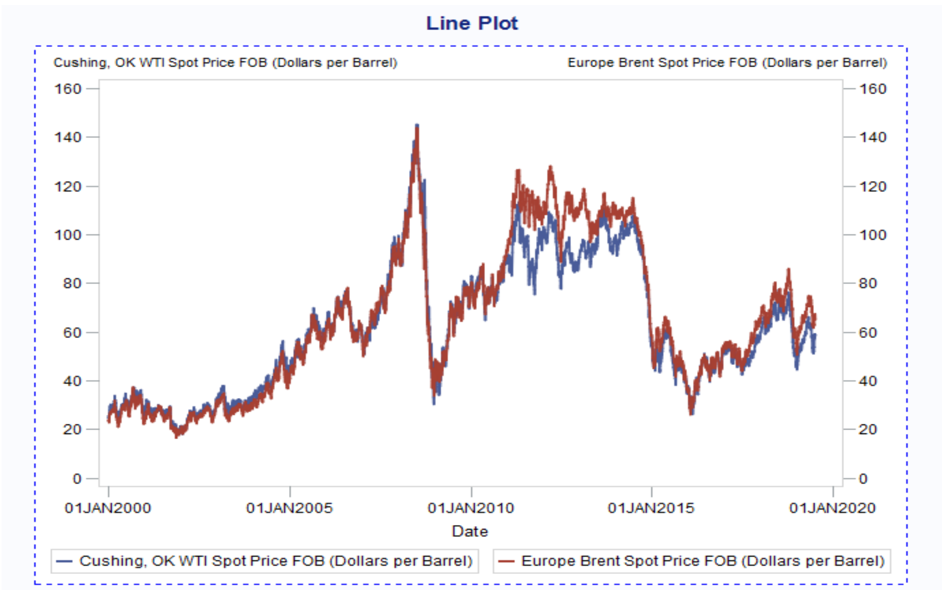

FIG. 2. Daily Crude Oil Spot Price from 2000 to 2019 (in US dollars)

3.2.2. The Descriptive Analysis of Data

The descriptive analysis of data help define the data's characteristics, and pre-processing method before fitting into model. From the Table 1, there are 4978 observations from January 1st 2000 to July 1st 2019. In contrast, Brent's average price is slightly higher than WTI, nearly $2.36 higher per barrel; besides, Brent's standard deviation is larger than WTI's, i.e., Brent prices market is more unstable than WTI. Another noticeable statistic is there are 84 and 27 missing values for WTI and Brent respectively. By pre-processing the missing values may take the potential risk of changing the patter, it can increase the model's generalizability. I general, the two ways to process missing values are either deleting or imputing. Since the missing amount is not large than the whole database, Python is applying to delete the missing values in the dataset. In Table 2, the descriptive analysis result is not obviously changed after the missing values are deleted. There are 4867 observations for Brent and WTI individually, and no change for the mean, minimum and maximum. From the descriptive analysis, the crude oil return descriptive statics shows the high-level similarities, so it illustrates the high level of integration between the world's large oil markets. Based on the finding, most of WTI's model's parameters should show high level of similarity as Brent's. In the modelling step, WTI's forecasting model will first fit data, and it provides some clues for s Brent market's models.

3.2.3. Criteria of Selection Process

The model selection and creating process follows three major criteria. First, model should capture long memory. The advantage of long memory model can capture key historical event (e.g., structural breaks in the past). By including black swan event in the dataset, whether the model has strong forecast power and stability in fat tail event can be tested. Second, model should capture the multi-scaling feature. For instance, different methods of normalization, computing the oil return differently impacts the result. To justify the optimal scaling method, one way is the user' applying personal experience. Another approach is for the model to automatically select the best features. Finally, an appropriate model should have high accuracy and stability in out-of-sample data. In this research, multiple linear regression model is created as baseline for complex models. More advanced linear based models ARIMA and Exponential smoothing is created later.

The nonlinear and machine learning models are necessary to apply under chaos theory in crude oil. The nonlinear model neural network fits data to compare the liner models' result. None linear models (e.g. ANN) may be overfitting and into local optimal easily. Though the performance is ideal in in-sample data, its forecasting accuracy in the real-world data is not conductive. Part of researches focused on forecast 1 trading day ahead to valid model performance, whereas longer and multiple forecasting horizons is meaningful to test model performance by including more rare events. Accordingly, 10, 20 trading day ahead in out-of-sample data are significant to validate model long term forecast accuracy.

TABLE I. Descriptive Analysis of Spot Oil Prices

Variable	Mean	Std. Dev.	Non Missing	Missing	Min	Median	Max	Skewness	Kurtosis
Cushing Price	62.02724	26.53666	4894	84	17.5	59.23	145.31	0.367874	-0.76648
Brent Price	64.39039	30.31785	4951	27	16.51	60.56	143.95	0.393839	-0.9517

TABLE II. Descriptive Analysis of Spot Oil Prices after Deleting Missing Value

Variable	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
Cushing Price	62.03044	26.51565	4867	0	17.5	59.23	145.31	0.368505	-0.76306
Brent Price	64.73514	30.36025	4867	0	16.51	60.98	143.95	0.379098	-0.96699

3.3. Preliminary Analysis of Data

Two analytical methods are applied for checking the existence of seasonality in WTI and Brent oil prices. First, visual inspection of line chart is for roughly check, and then autocorrelation test (ACF) is for more accurate check. In the Figure 2, one unit of the horizontal axis represents five years, but it is not accessible to inspect seasonality visually. Alternatively, reconstructing the line chart with two years per unit is more visual inspection friendly. The reconstructing oil return line chart of WTI and Brent are displaying on Figure 3 and Figure 4 respectively. WTI spot oil price at the beginning of the 2000 is high, and then it starts descending till the end of June. Likewise, Brent's line chart illustrates the price in summer period is less than winter period. This observation validates the seasonality existing in crude oil market, and the world's oil market consistency. Most of periods in the chart shows the same cyclical fluctuation; however, the exceptions are found frequently from the start of 2008 to 2012. This phenomenon revealed that seasonal is not the only influence factor, other factors like economics are also powerful to impact crude oil price. In later models identify stage, seasonality-based model will be one of them.
Another motivation for seasonal testing is the long duration of spot oil prices; As a result, seasonal visual examination may not be accurate sufficiently. Then second method is applied, i.e., autocorrelation and partial autocorrelation tests. If the data without seasonality, there will be less correlated between lags, and each lag's coefficient will stay from upper confidence limit to lower limit confidence limit. In other words, the correlated lags illustrate seasonality. From the Figure 5, in 16 lags, most of coefficients located within the limit, besides lag 1, 3 ,5, and 7. In WTI, seasonality probabilities are important for early data, but not so important for recent data. In Figure 6, the Brent's autocorrelation chart, lag 6 and 14 is exceeding the confidence limit, but others stay within the limit. Seasonality is not detected in the Brent market. In summary, for the prediction of oil fluctuations, seasonal factors seem to be less important during structural damage, but seasonal model and non-seasonal model were fitted and their prediction accuracy was compared in the dataset.

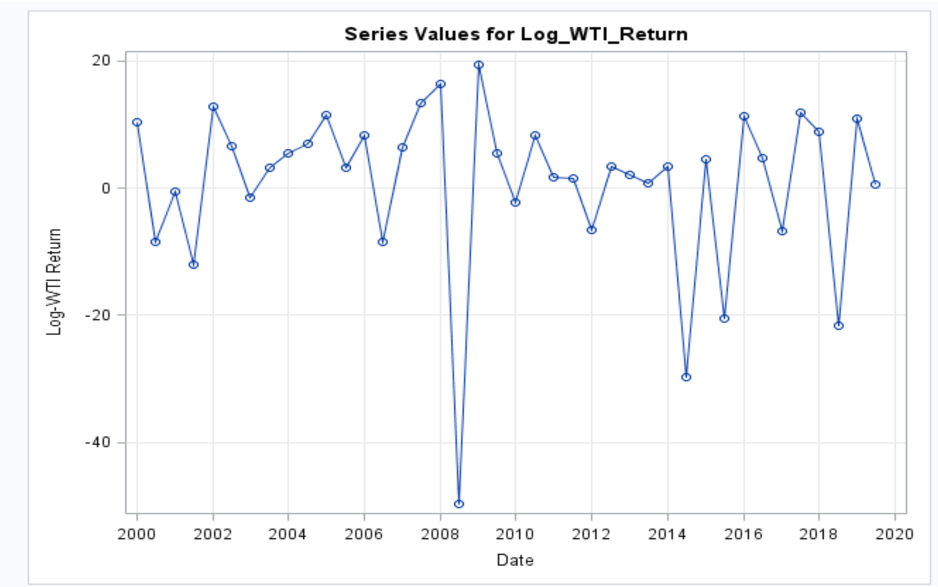

FIG. 3. WTI Oil Return Line Chart for Detecting Seasonality

3.4. Process of Modelling

Initially, preparing the preliminary analysis of crude oil for both WTI and Brent gives a better understanding to the data. In preliminary analysis, descriptive statistics is adopting for the basic statistics

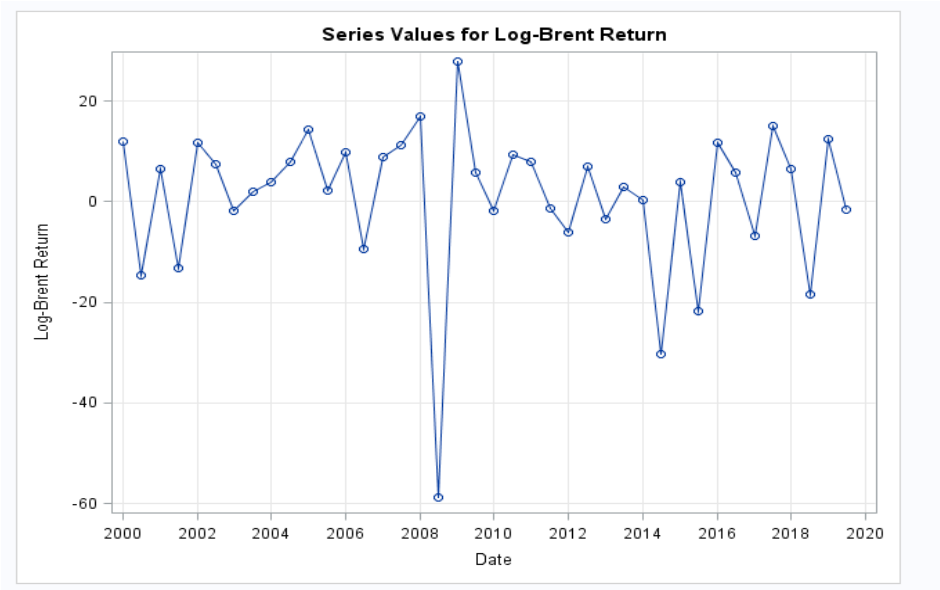

FIG. 4. Brent Oil Return Line Chart for Detecting Seasonality

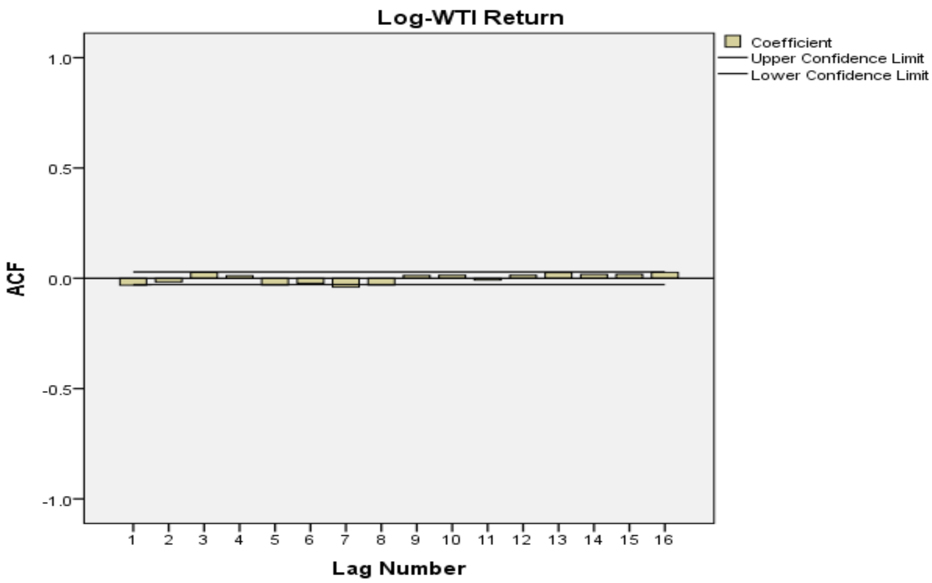

FIG. 5. WTI Oil Return ACF Chart

metrics cover the mean, standard deviation, maximum, and minimum, etc. Furthermore, autocorrelation test was performed for WTI and Brent for identifying the seasonality in crude oil price. Overall, preliminary analysis assists user to select model and identify the parameters in the modelling stage. The data is split into training, validation dataset, 90% and 10% respectively; 80% and 20% for ANN to avoid overfitting. Test dataset is for 10 day-ahead forecasting is from June 18th 2019 to July 1st 2019; 20 day-ahead forecasting is from June 4th 2019 to July 1st 2019. Finally, a stationary time series model is built enabling model's prediction to be more stable under black swan event. The two data transformed methods are logarithmic return and classical return. Logarithmic return method is to calculate crude oil price return by the following function:

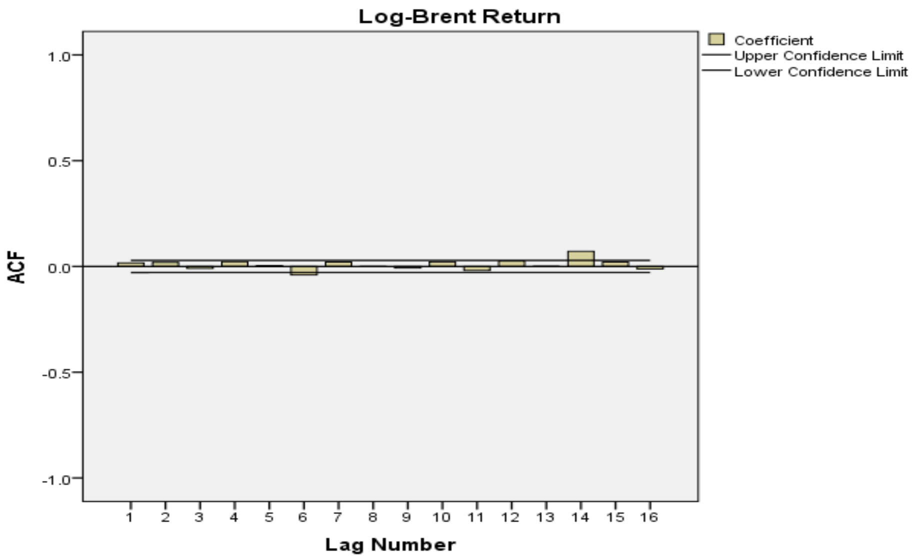

FIG. 6. Brent Oil Return ACF Chart

$r_t = 100 * (log(pt) - log(Pt - 1))$
The return is calculated by current logarithmic price minus yesterday's logarithmic price. Excel is employed to calculate logarithmic return, so the base 10 logarithm of number is applied as default. Another method is classical return, and it is calculating by following function:
$R = \frac{r_f - v_i}{v_i}$
In this method, return is equated with current crude oil price minus yesterday's price, and then divided by current price. In the modelling step, two data transform methods are comparing under certain evaluation metrics, so the better one will be applying to remaining of study.

3.5. Forecasting Evaluation Metrics

After defining the forecasting models, the next step is to compare forecasting price to actual price in out-of-sample dataset. This step helps user select the optimal performance forecasting model. It is known that the complex model has favourable accuracy in training data, but there may be overfitting for the validation data. The optimal forecasting model should have relatively small prediction errors in validation dataset. To assess forecasting model's performance, two types of evaluation metrics is applied in this research, scale-dependent measures and information criterion. In Chen, Twycross, and Garibaldi (2017) research, they asserted that performance of forecasting is varies with the accuracy measure methods have been used. In other words, crude oil forecasting models' performance is based on the selected evaluation metrics.

Crude oil volatilities' forecasting accuracy is depending not only on the forecasting models, but also on evaluation metrics. Appropriate crude oil forecasting models may achieve unfavourable results, since inappropriate evaluation metrics misleads the results. Scale-dependent measures is based on absolute or squared errors, and it is useful in comparing forecasting methods on same set of data (Chen, Twycross, and Garibaldi, 2017). Three popular scale-dependent measures are mean absolute error (MAE), mean squared error (MSE), and error root mean squared error (RMSE). MAE calculates arithmetic mean of absolute errors, so it is easy to compute and understand; whereas, bias may occurr when data has large outliers. MAE's formula is expressed as
$MAE = \frac{1}{n} \sum_{t=1}^{n} |e_t|$

Unlike MAE's calculation, MSE calculates arithmetic mean of squared errors. Its formula is:
$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2$
RMSE is the square root of MSE, so it is sensitive to data's outliers as well, similar to MSE and MAE's shortcoming. Its formula is:
$RMSE = \sqrt{MSE}$
Sum of squared errors (SSE) is sum of squared the difference between observation and overall average.
$SSE = \sum_{t=1}^{n}(X_i - \bar{X})^2$
In brief, the scale-dependent evaluation metrics depend on scale of data, so it is crucial to user to smooth the outliers before imputing data into forecasting models. Either the data transform or data normalization is strongly recommended in the data pre-processing step.

Information criterions (IC) is another type of evaluation metric used in the article. Representative information criterions cover Akaike Information Criterion (AIC) Empirical Information Criterion (EIC), Schwarz's Bayesian Information Criterion (BIC). The motivation of applying IC is that it is a simple method to choose from a range of competing models. On the other hand, IC is highly theoretical, so it is not clear how this will work in practice, or whether users know which information criteria are best suited for predictive tasks (Billah, Hyndman, and Koehler 2003). The performance of information criterion depends on various factors (e.g., forecasting horizons, types of models). AIC is a common criterion to assess the autoregressive model's performance, since it balances the quality and model complexity. AIC can be expressed by following equation:
$AIC = e^{\frac{2k}{n}} \frac{RSS}{n}$
K denotes the number of repressor's (including intercept), n is the number of observations and RSS is the residual sum of squares of the model. The model with a smaller value of AIC is more conductive to user.

To assess forecasting model performance, model producing the lowest forecasting error is selected, since it means more accurate between forecast and actual value. Depending on models' type and data pre-processing methods, the evaluation metrics may be slightly different.

4. Modelling: Multivariate Forecasting Models

This research starts with the classical forecasting method and follows by the AI model. The multiple linear regression (MLP) is the benchmark. As the chaos theory and uncertainty exist in the crude oil market, model that can capture the nonlinear relation may exhibit better forecasting accuracy. Quadratic regression and ANN models are applied after the conventional statistical model is built. The data pre-processing method is min-max normalization. Unlike the time series models proposing under stationary data assumption, the normalized data is prioritized both in regression and ANN models. However, one limitation of ANN is that parameters adopted in training data are not available to test in 10 day and 20 day ahead forecasting, since it is black box model. Due to the limitation of ANN, linear regression and quadratic regression models are adopted to compare the forecasting performance in validation dataset instead of doing 10 day and 20 day ahead forecasting.

4.1. Regression Model

Regression function is a basic analysis to find the relationship between target and inputs, and multiple linear regressions (MLR) is created as benchmark for other models. The general linear regression equation is
$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{p-1} x_{i,p-1} + \in$
MLR requests one dependent variables and more than one independent variables; however, the WTI and Brent's crude oil spot price are univariate. It is necessary to convert the univariate data into multi-variable data and pre-process the data.

4.1.1. Data Normalization

Unlike the pre-processing step in time series model, ARIMA and ES, regression models do not require to transform the data from the non-stationary to stationary data, whereas normalization is highly recommended. Regression is sensitive to outliers, and existence of outliers will down-regulate the model's accuracy. According the descriptive statistics and the history of crude oil spot prices, outliers are detected in high frequency around 2008. Min-max and Z score, min-max normalization is taken between two popular normalization methods. This normalization method is scaling the price data from zero to one, so the standard deviation will end up in a smaller number. The Min-max normalization function is presented as:
$X_n = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}$
Z-score is calculated based on the data's mean and standard deviation; however, non-stationary data's mean and standard deviation is not sufficiently stable. The application of Z-score method may cause the normalized data not to be in the same range, so this method is rejected.
$X_n = \frac{Value - \mu}{\sigma}$
As regression model is the based on the multiple inputs, univariate data must transfer to multivariate data. In the later computation, the original spot prices are replaced by the normalized price, since the large-scale data reduces the impact of small-scale data. The major idea of regression model is to analyse the historical data to forecast future. The normalized price as the result, 2,3,4,5 day's average spot oil price, logarithmic return, volatilities are included in the input variables. In brief, the data cover one dependent variable is normalized crude oil price, and 11 independent variables, and a time ID. In the Table 3, all the variables for WTI and Brent are shown.

TABLE III. WTI Regression Inputs and Target List

Number	**Name**	**Role**
1	Min-max Normalization Price	Target
2	Yesterday Price	Input
3	2 Days Average Price	Input
4	3 Days Average Price	Input
5	4 Days Average Price	Input
6	5 Days Average Price	Input
7	WTI Return (Log)	Input
8	WTI Volatilities (Log Squared)	Input
9	WTI Volatilities (Log Absolute)	Input
10	WTI Return (Classic)	Input
11	WTI Volatilities (Classic Squared)	Input
12	WTI Volatilities (Classic Absolute)	Input

4.1.2. Multiple Linear Regression Parameters Selection:

Different than univariate time series model, the 11 predictive inputs of regression models are created by user. The user created inputs raising concern about whether those inputs are significant to target or not. To eliminate the concern, significant test is applying to MLR model. The significant value is 0.05, and the null hypothesis is predictive input is not significant to the target. If predictive input's p value is greater than 0.05, the null hypothesis is accepted; otherwise, the null hypothesis is rejected.
As the result, only WTI 4 and 5-day average are not significant enough to the target, since their p value are greater than the 0.05, which is 0.57 and 0.42 respectively. In this statistical test, it proves the recent WTI prices are more significant to the forecasting target, and that forecasting the short-term oil prices is more accurate than the long-term. On the other hand, the top three important variables are WTI return, WTI volatilities, and WTI yesterday price. In contrast, the white-box models, MLR, has the benefits of easy to explain, examine; moreover, predictive variables' coefficients clearly show the significance and effect level to the target. In WTI's MLR model, three justifications are attempting. In the first attempt, MLR includes all the input variables. Secondly, the insignificant variables are dropped. By dropping the WTI 4-day and 5-day average price, the value of SSE, MSE, and RMSE are

all increasing. This attempt is rejected since its high error rate. In the final attempt, only one input that has major effect to the target is kept, but this attempt is failed with high error rate. In contrast, MLR's coefficients allocated weights for each variable, insignificant variables are not necessary to drop; accordingly, WTI's MLR includes all the 11 predictive variables.
Brent's MLR justification is following the WTI's process. The process is starting with significant test, most of inputs are not significant, since their P values are greater than 0.05. One possible explanation can be multicollinearity in time series data. To solve this issue, the stepwise algorithm is generating in the next step. In the next step, the top five insignificant variables are dropped in the system to compare the results with the original result. After several inputs justifications, the model has the lowest value of error rate, MSE, RMSE, and SSE, includes all 11 inputs in the function.

4.1.3. Quadratic Regression Parameters Selection

Nonetheless, the straight line MLR may not fit into data as well as other models. From the WTI and Brent spot oil price line chart, the high fluctuation characteristics is weakening the MLR prediction power, a straight line not seems fit the high volatilities very well; moreover, the crude oil price may not always increase or decrease in linear trend. One assumed trend is damped trend, oil price should be in the long-term view. Another assumed trend is multiplicative, and the convinced example is the structure break in 2008. Prices has a huge drop and down during 2008 to 2012, so multiplicative trend seems to be more accurate for forecasting than linear trend under uncertainty. Based on the varied trends of oil price, a complex form of the regression is in need to develop. In regression family models, polynomial regression enables a single predictor variable in several powers, so they a complex form of linear regression models. In SAS Enterprise Miner software, the degree of polynomial is defined by user. By setting the polynomial value as two, quadratic regression is applied for WTI and Brent markets. In contrast, the quadratic regression for WTI and Brent has the better prediction accuracy than the MLR, and the detailed results will be discussed in the later. To reduce the amount of computation and to optimize quadratics function, three types of variable selections are applying in WTI and Brent's data. The methodology behind stepwise method is added or deleted one variable at a time. Forward selection method means variables are added to model one at a time. In backward elimination method, it covers all 11 variables initially, with variables subsequently being eliminated one at a time. The motivations of applying all three algorithms to both market's quadratic regression models is variable deleted or added in the early stage may not important in a later stage. Also, the volatilities pattern in the different markets can be varied. For instance, WTI spot prices are more correlated with the recent data; however, Brent seems not have a specific pattern of the data fluctuation.

4.2. Neural Network (ANN)

4.2.1. Data Pre-processing

Min-max method of normalization is applied to pre-processing process, and all ANN family models use the same dataset as regression models to make the results comparable. Normalization is crucial and necessary step in data pre-processing, since ANN is a sophisticated form of the linear regression model, and the outliers may affect the ANN's forecasting accuracy. Li et al. (2018) introduced two benefits of normalization; one is accelerate the search for the optimal solution for gradient descent, and the other is to allow the features of different dimensions to be comparable numerically and possibly improve the accuracy. In order to establish a fair comparison, this research adopts a common normalized method (the Min-Max normalization) for AI-based predictors. ANN's target variable is same as regressions, which is spot oil price after min-max normalization. WTI and Brent spot prices are dropped since they are normalized value; it can mislead the ANN model. Since the non-stationary data enable ANN to capture the general characteristics, transforming nonstationary to stationary data is not necessary for ANN models (Kulkarni and Haidar, 2009).
The data partition step in the ANN model should be extra careful, since ANN model's training and validation weights may be different than other models. Such difference referring to ANN is a complex

model mimicking human brain. Complex model can capture the brilliant in-sample accuracy, but poor performance in the out-of-sample validation. ANN model may be too complex to overfitting. To avoid overfitting, a larger amount of the validation dataset can help the early stopping of ANN. In this research, the 20% validation data and 80% of training data will be applied after trying several different portions of division. Principle Component Analysis (PCA) is a feature selection method that extensively used in ANN models for reducing dimension. In concept, PCA filters the irrelevant features in order to improve the prediction accuracy. In Grigoryan (2015), he states the major idea in PCA is to find the component vectors that interpret the maximum possible amount of variance by linearly transformed components. From several methods to compute PCA, a correlation method is adopted to extract principal competent in this research. However, the inputs variables are not considerable in this case. PCA is useful in large amount of input, and probable not necessary in this model. The result is to be compared for the models with or without PCA. Among various types of ANN, multilayer Perceptron, Generalized Linear are developed in SAS Enterprise Miner.

TABLE IV. WTI's Principle Components List

Principle Component 1	**Principle Component 2**	**Principle Component 3**	**Principle Component 4**
WTI 3 days Average	Log absolute WTI volatilities	Basic WTI Return	Basic squared WTI volatilities
WTI 4 days Average	Log squared WTI volatilities	Log WTI Return	Basic absolute WTI volatilities
WTI 2 days Average	Basic absolute WTI volatilities		Log absolute WTI volatilities
WTI 5 days Average	Basic squared WTI volatilities	Log squared WTI volatilities	
WTI Yesterday price			

4.2.2. ANN Model Parameters' Setting

The Generalized Linear ANN is the benchmark, since it does not have hidden layers. Subsequently, multilayer ANN fits data. One advantage of MLP model is user can set more than one hidden unit. As for the hidden nodes' selection, the number is determined using the practical guidelines used by previous researchers ($2n + 1, 2n, n,$ and $n/2$). The number of n is 11 in this model, so the hidden units 23, 22, 11, and 5 are tried in the model. Also, 1 hidden unit is trying, since 1 hidden unit is sufficient in most of cases; the inputs are not large enough to carry number of hidden layers. In SAS Enterprise Miner, the training and validation data is split into 100 units, and each unit covers nearly 204 observations of training data, and 51 observations of validation data. In those 11 variables, part of them may be highly correlated to another, PCA can help user eliminate the worries. Other motivations of applying Principle Component Analysis is reducing the calculation cost, make the black box model easier to understand. After the principle component analysis, the ANN models inputs decreases to four principle components. However, one disadvantage of applying PCA is delete some observations. By comparing the results from the model with PCA and without the PCA, the model without the PCA exhibits the better performance for both WTI and Brent markets.

TABLE V. WTI Hidden Units and Error Evaluation

Number of Hidden Units	**Validation SSE**	**Validation MSE**	**Validation RMSE**
ANN(Multilayer)n=1	0.034591	0.00003401	0.005832
ANN(Multilayer)n=2	0.022548	0.00002217	0.004709
ANN(Multilayer)n=4 (Selected)	0.00297	0.000008822	0.00297
ANN(Multilayer)n=5	0.005664	0.000005569	0.00236
ANN(Multilayer)n=6	0.007166	0.00007046	0.002645
ANN(Multilayer)n=11	0.039171	0.00003852	0.006206

TABLE VI. Brent Hidden Units and Error Evaluation

Number of Hidden Units	Validation SSE	Validation MSE	Validation RMSE
ANN(Multilayer)n=1	0.053047	5.22E-05	0.007222
ANN(Multilayer)n=4	0.023117	2.27E-05	0.004768
ANN(Multilayer)n=5	0.040485	3.98E-05	6.31E-03
ANN(Multilayer)n=6 (Selected)	0.020168	1.98E-05	0.004453
ANN(Multilayer)n=7	0.038153	3.75E-05	0.006125

4.2.3. The limitation of ANN class model

ANN is a black box model, so some parameters are difficult to be explained; in the meantime, the first input that ANN model randomly choose may misleading the model result. Also, more worries are originate from its data partition method. On the whole, time series forecasting models divide the training and validation dataset by the periods; whereas, ANN is the black box method that user is unable to know the partition method in SAS Enterprise Miner. Compared with time series models, like ES and ARIMA, they can select data in earlier period as training data, and recent data as validation. For time series data, data partition that based on the date is more understandable and reasonable than randomly selected method, especially when the crude oil data has several structure breaks. In brief, ANN class model does not have strong prediction ability of time series model.

5. Modelling: Univariate Forecasting Model

Though ANN-class model can capture the complex attribute in the oil price, it has limitation in the model testing and result explanation process. In contrast, the univariate time series forecasting models have easily interpret result and have capability to evaluate the 10 day ad 20 day-ahead forecasting. Two classical univariate models are applied, Exponential Smoothing (ES) and Auto regression integrated moving average (ARIMA). Logarithmic return is adopted to transform the data from nonstationary to stationary. Under the same pre-processing method, ES and ARIMA's forecasting performances are compared

5.1. ES models

5.1.1. Data Pre-processing

In this paper, the stationary data is imputing to ES models, since imputing the same dataset enable user to compare the model's performance between ES and ARIMA. Also, a stable time series data takes concerted action to simple ES horizontal data assumption. As simple ES is the first univariate time series model, two data transformation methods are applying into ES model. The two transformation methods are logarithmic return and square root classical return, the detailed calculations have introduced in mythology part. In the next step, user has to prepare the time series data according to data's frequency; SAS Enterprise Miner can help user to prepare time series data in an easy way by setting several criteria's before the node is running. The crude oil spot price is a weekday based data, so user need to specify the weekday as time interval. To prevent any missing value that existing in the weekday data, user can replace the missing value to data constant value. Based on the finding of descriptive statistical, the structure break around year of 2018 incurred large number of outliers. By smoothing the outliers as predictive value in ES model, the pattern of imputing data might closer to horizontal, which is conform to the assumption of ES model. The last step before fitting data into ES model is the divide data into training, validation and testing. In SAS Enterprise Miner, I use WTI's classical and logarithmic return as target variable for comparing their performance. After imputing two dataset into system, the times series data preparation node is connecting to data, so the time interval can be set by weekday based, and data partition node is setting up as the as MLR model. The system automatically select the best ES model for imputing data, the following table shows the training data error metric. Classical return has higher RMSE, MSE and MAE than logarithmic return, so the classical return method is rejected. As the marketing similarity's existence, Brent is using logarithmic return as target. Overall, the log return data performs better than another, so logarithmic return

TABLE VII. Data Transformation Method Comparison

Method	**RMSE**	**MSE**	**MAE**
Logarithmic Return	1.056074	1.115293	0.815447
Classical Return	1.11307	1.278484	0.832161

price is applying for other time series models. After defining the dataset, the proportion of training and validation data need to be defined. The most favourable data partition is 90% of training data and 10% of validation data. The training data, from January 1^st 2000 to August 31, 2017, is used to find the smoothing parameters; the validation data is used for evaluating forecast accuracy and model selection after determining the parameters, which is from September 1st, 2017 to June 03rd, 2019; test data is last 10 days and 20 days.
The prior evaluation criterion is AIC, other major criterion is MAPE, MSE. Number of criterions can be used for evaluating the model performance. In Billah, King, Snyder, and Koehler (2006)'s paper, their find is AIC has a slight edge over its counterparts among AIC, BIC HQ MCp, GCV, and FPE. Though there is little to distinguish the various information criteria, the information criteria approaches outperform the encompassing approach. In order to prevent different evaluation metrics demonstrate different best model, AIC is set up the prior criteria to evaluate the ES family models.

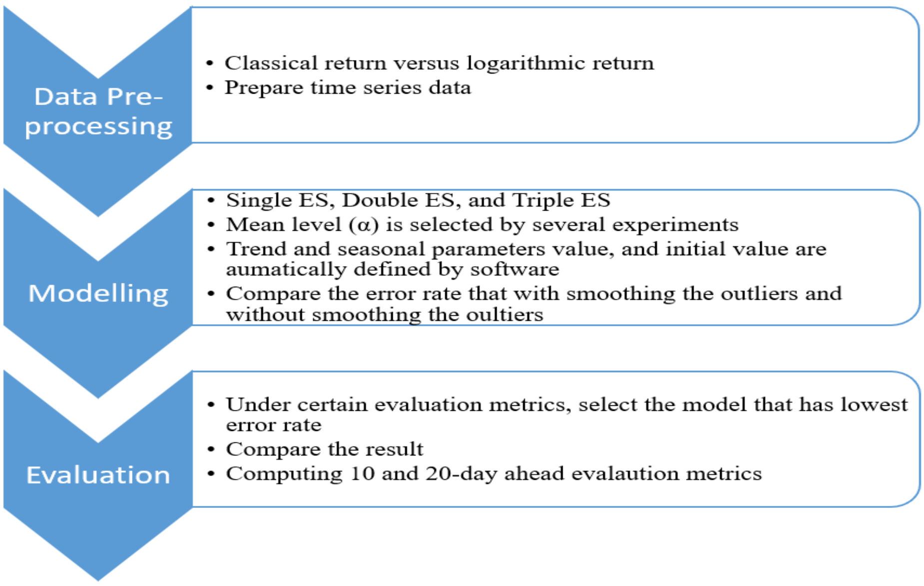

FIG. 7. . Exponential Smoothing Modelling Process Flow

5.1.2. Model Selection

The motivations of employing ES model to crude oil data are declining weight on older data, easy to compute, and minimum data is required (Holt, 2004). First, descriptive statistics help user to select appropriate ES family models by visually inspecting two market's crude oil spot price line chart. Three data trends of ES models are linear, exponential and damped trend model. Linear trend means the price increase or decrease fixed amount each period, which is an ideal model theoretically, exponential trend indicates the price increase or decrease by specific value of factor. These two trends are fitting into data later, the only exception is damped trend ES model. The damped trend means the crude oil price will remain stable in some point's future, it may happen in the long term perspective; however, this research focused on the short period forecasting, the damped trend is excluded in models selection. Based on the descriptive analysis, seasonality of data should be considered as one crucial component that influences the model accuracy. In ES family models, Triple ES, an extended model based on Double ES, is able to a'alysed data's seasonality; thus, besides the double ES's mean level and trend parameters, seasonality is the additional parameter enable Triple ES be different than other ES models. In this case, winter additive ES and additive seasonal ES model are fitting to the two crude oil markets' data. In triple ES model, there are two types of seasonality, multiplicative and additive. In this research, the multiplicative seasonality is not considering, since it makes model become complicated. On the other hand, additive methods are commonly used for constant data. In the data pre-processing step, crude oil data has been transformed to stationary data, which means it is constant and prioritized by additive seasonality. In the later modelling process, I am going to compare the model with smoothing outliers and without smoothing outliers, and then select the model with lower error rate.

5.1.3. Simple ES

Simple ES is also called as single ES model, it assumes the data has the stable mean, and no trend and seasonality; moreover, it is ARIMA (0,1,1) practically, it performs well in short period forecast. The assumption of ES model the data pattern is nearly horizontal, and no trend or seasonal variation exists in the historical data. In family of ES models, simple ES is setting as benchmark for others. The simple ES formula shows below:

$$S_t = \alpha y_{t-1} + (1 - \alpha) S_{t-1}$$

The function that forecasting the next point is

$$S_{t+1} = \alpha y_t + (1 - \alpha) S_t$$

where $0 \leq \alpha \geq 1$
The level is slowly changing over time. The starting value of simple ES is generating by the software automatically. The user only needs to set up the level parameter, α. The value smoothing parameters α, is subjective and relying on the user's experience. An increasing value of α represents recent data carries more weights than aged. In SAS Enterprise Miner, the default setting of α is 0.05, and it brings out the best in sample performance compared with α value of 0.1, 0.2, and 0.9 separately. By comparing the result, the default setting 0.05 performs best among three values. One explanation of small value of α can be the dataset has transferred from non-stationary to stationary, so smaller value of α fits data better when data is stationary. The single ES is a basic model in univariate time series forecasting, since the model does not consider seasonal and trend. In preliminary analysis, seasonality impacts the crude oil price in some measure, and it will be considered in advanced ES model later in the paper.

5.1.4. Linear ES

The linear ES is an extended form of simple ES, since it includes level and trend factors in its function. The motivation of fitting double ES model into crude oil data assumes that the trend of crude oil may slightly increase year by year. Though the ACF chart and visual inspection did not show the obviously trend in WTI and Brent prices, the trend may cause by the healthy 3% US dollar inflation every year; moreover, the data range is about 18 years, which is large enough to consider the inflation impact to crude oil price. In SAS Enterprise Miner, Additive and multiplicatively trend can be selected by user;

however, the multiplicatively trend will increase the model complexity, it will not consider in Double ES model. Double ES with an additive trend is called Holt's Linear ES, and its formula shows below:

$$S_t = \alpha y_t + (1-\alpha)S_{t-1}$$
$$b_t = \Gamma(S_t - S_{t-1}) + (1-\Gamma)b_{t-1}$$

Besides the smoothing parameters α, γ is representing the trend. In the software, the initial value and parameter γ is generated by the system, parameter α's value is same as simple ES which is 0.05. After setting down the parameters, SAS Enterprise Miner generating the following figures, Figure 8 that shows Brent's linear ES forecasting. From the chart it shows the trend of oil return is going downwards.

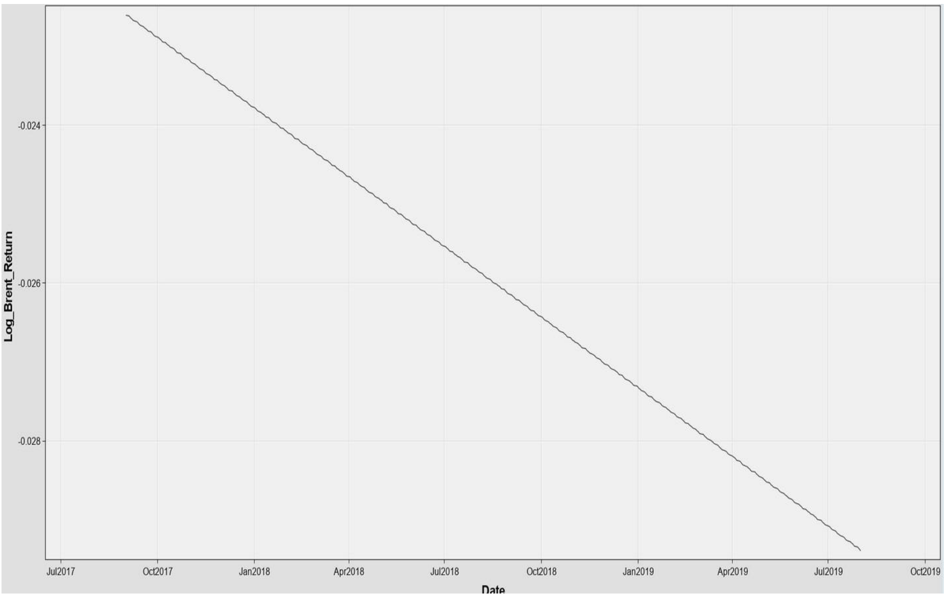

FIG. 8. The Screenshot of Brent's Linear ES Forecasting Result from SAS Enterprise Miner

5.1.5. Triple ES

Based on the descriptive analysis, seasonality of data should be considered as one crucial component that influences the model accuracy. In ES family models, triple ES is an extended model based on Double ES. Triple ES can analyze data's seasonality, so three parameters are included in the formula. Holt-Winters ES is created by Winters and Holt, as an extended model of the Double ES. Holt-Winters method includes three smoothing parameters, they are α in level mean function; γ in trend function, and δ in seasonality function. In the following functions, S_t, b_t, and I_t represent the mean level, trend, and seasonal individually.

Overall Smoothing:

$$S_t = \alpha y_t + (1-\alpha)S_{t-1}$$

Trend:

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}$$

Seasonal:

$$I_t = \beta\frac{y_t}{S_t} + (1-\beta)I_{t-L}$$

Forecast:

$$F_{t+m} = (S_t + mb_t)I_{t-L+m}$$

In general, users need to define the three parameters based on experience. SAS Miner Enterprise simplifies the parameter selection process, and only parameter α needs to be defined by the user; other parameters are generated in the system automatically.
Two types of Triple ES are applied to WTI and Brent. The first one is the Additive Winter model, which includes seasonality and trend. The second model is additive seasonal, which only covers seasonality and level in the function. In Triple ES, WTI price is used to test outliers' effect on forecasting accuracy. Table 8 shows the performance of the model without smoothing outliers, and Table 9 shows the performance of the model with smoothed outliers. In summary, forecasting models with and without smoothing outliers do not have a huge forecasting difference. The model that smooths outliers is kept, since the ES model is based on the linear assumption.

TABLE VIII. WTI ES model Error rate Without Smoothing Outliers

Model Name	**MAE**	**MSE**	**RMSE**	**AIC**
Additive Winters Exponential Smoothing	0.73884	1.10642	1.051867	472.0199
Additive Seasonal Exponential Smoothing	0.738076	1.105163	1.051613	467.7991

TABLE IX. WTI ES model Error rate After Smoothing Outliers

Model Name	**MAE**	**MSE**	**RMSE**	**AIC**
Additive Winters Exponential Smoothing	0.738844	1.106424	1.051867	472.0119
Additive Seasonal Exponential Smoothing	0.738076	1.10589	1.051613	467.7991

1. Additive Winter Model.
   After defining the parameter α in SAS Miner Enterprise and smoothing the outliers, the Holt-Winters ES shows the oil return in a slowly downward trend for both WTI and Brent markets. Figure 9 is additive winter's forecasting result in WTI market. One possible explanation of the trend can be the structure break in 2008 leads the price has huge increase and following a drop down, so the software captures the characteristics, the trend oil price is going down.

2. Additive Seasonal Model
   Different than conventional Holt-Winters model, this triple ES eliminates the trending factor in the function. The system default model selection criterion is MSE, and the parameter α is same as additive winter which is 0.05. Also, smoothing the outliers' option is selected in the systems before running the model. After the setting up the model, SAS Enterprise Miner automatically selected additive seasonal model as the best model in ES-family.

5.2. ARMA Family Models

5.2.1. Data Pre-processing

The assumption of ARIMA is that crude oil price is a linear function. The linear function is written as:

$$Y_t = b_t + \epsilon_t$$

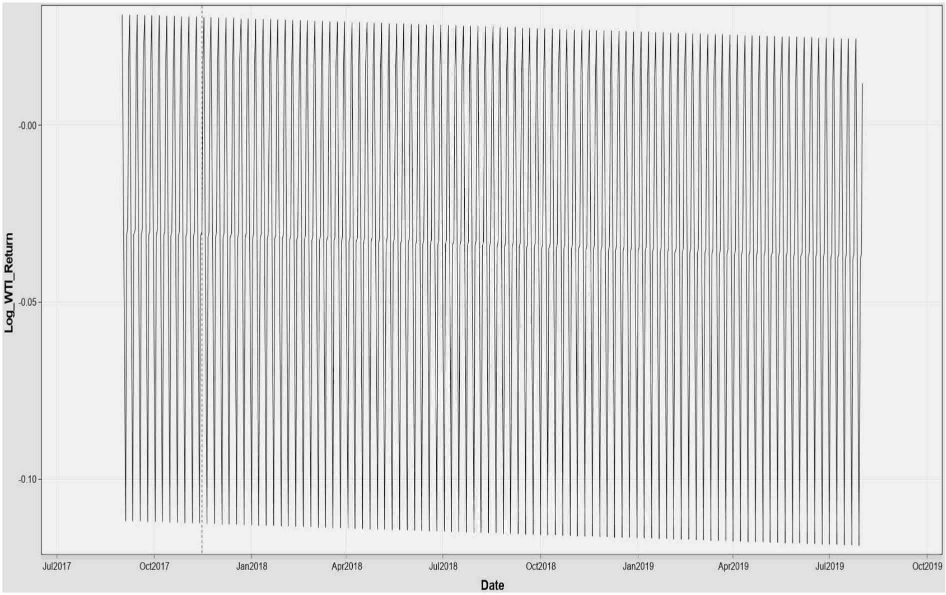

FIG. 9. A Screenshot of WTI Additive Winter Model Forecasting Result from SAS Enterprise Miner

The error term is normally distributed and independent, with a mean of zero.

The motivation of the ARIMA model is that it is a univariate model, and a good ARIMA model requires at least 50 observations, with a large sample size required for seasonal time series (Okrie, 2014). By performing sufficient observations, the 19 years daily data, ARIMA model applies to forecast crude oil price.

Before applying the model, whether the data is stationary should be checked. ARIMA model is based on stationary data, so it has no trend or heteroscedasticity. In time series data, stationary means the mean, variance, and autocorrelation remain stable over time. Methods of checking stationarity include visually inspecting the price plot, mean, ACF, PACF, and IACF plot. The Autocorrelation (ACF) is the correlation between a variable lagged one or more periods and itself (Selvi, Shree, and Krishnan, 2018). The lag k ACF function is shown below:

$$r_k = \frac{\sum_{i=1}^{N-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}$$

The PACF function at time lag k is the correlation between Y_t and Y_{t-k}, and it plays a crucial role in data analysis aimed at identifying the extent of the lag in an AR model. From the descriptive analysis, the WTI and Brent's price plots show high fluctuations, indicating they are not moving around a constant mean based on visual inspection. Additionally, the WTI and Brent's standard deviations are large, further demonstrating the non-stationary feature in crude oil.

To remove the non-stationarity of time series data, two metrics in the data pre-processing step can be used to calculate crude oil return: logarithmic return and classical return. Their detailed calculating functions have been discussed in the methodology chapter. Based on ES model's findings, logarithmic return has better prediction performance than the classical return method; accordingly, ARIMA is going to employ the same database as ES. After the logarithmic return, the average prices of WTI and Brent are near to zero, so the non-stationary data has been transformed into stationary data.

After the data transformation, ACF and PACF tests are applied to validate the crude oil's non-stationary feature in a more accurate way. By inputting the crude oil spot price data into SAS Enterprise Miner, the system will generate the autocorrelation test and divide the data into 24 lags automatically. Okrie (2014) experimentally found that crude oil price is non-stationary, since the ACF plot decays very slowly to zero. From Figure 5, the WTI's ACF decays slowly, so the autocorrelation plots demonstrate that the daily crude oil price data is non-stationary. In Figure 6, Brent's ACF also shows the data's non-stationary feature. The crude oil return data for both markets is non-stationary, so the ARIMA model must fit into oil return data instead of the ARMA model. Later, the difference (d) with a value of one is applied in the ARIMA model.

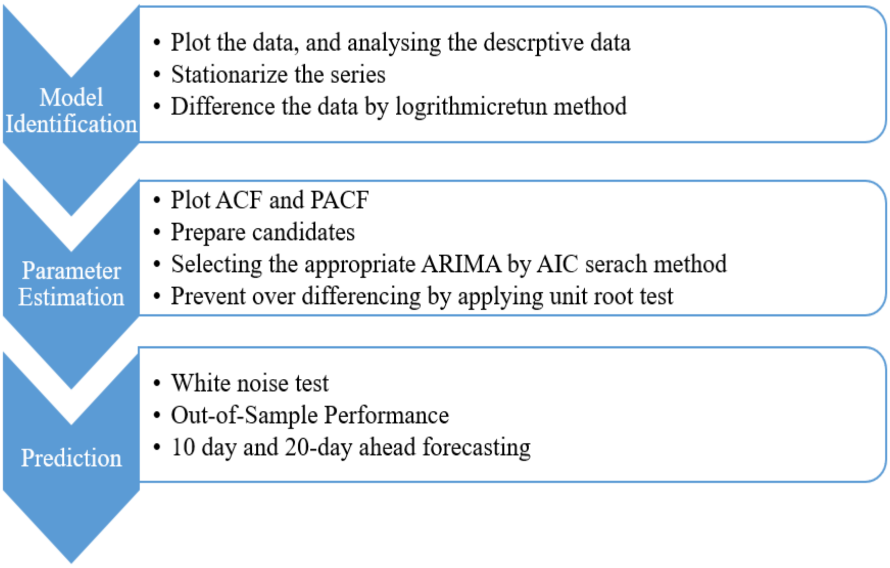

FIG. 10. ARIMA Model's Process Flow Chart

5.2.2. ARIMA's Parameters Setting

ARIMA model combines the AR and MA. AR displays the variable of interest is regressed on its own lagged values; MA shows the regression error is virtually a linear combination of error terms (Selvi, Shree, and Krishnan, 2018). If $q = 0$, then the model refers to AR model. If $p = 0$, the model is MA type. If $d = 0$, then it is ARMA model.

1. AR Model:
   To understand ARIMA and its parameters, it is crucial to break down the ARIMA into AR and MA models. The Auto-Regressive model (AR) is p in ARIMA model:

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p} + \epsilon_t$$

   where:
   Y_t is the dependent variable
   Φ is estimated coefficients
   ϵ_t is the error term which represents the variable that is not considered in the model
   $Y_t$is dependent variable Y_t and $Y_(t-k)$ by plotting the PACF, people can find the appropriate lag p in an AR(p) model or ARIMA(p, d, q) model.

2. Moving Average model (MA)
   Moving average is emphasized as the function below:

$$Y_t = \Theta_0 - \Theta_1 \epsilon_{t-1} - \Theta_2 \epsilon_{t-2} - \ldots - \Theta_q \epsilon_{t-q} + \epsilon_t$$

   where:
   Θ represents the estimated moving average parameter ϵ_t is the random error term

3. ARIMA Model:
   ARIMA model can be stated as follows:

$$Y_t = \Theta_0 - \Theta_1\epsilon_{t-1} - \Theta_2\epsilon_{t-2} - \ldots - \Theta_q\epsilon_{t-q} + \epsilon_t$$

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \ldots + \Phi_p Y_{t-p} - \Theta_0 - \Theta_1\epsilon_{t-1} - \Theta_2\epsilon_{t-2} - \ldots - \Theta_q\epsilon_{t-q} + \epsilon_t$$

Due to continuing change in the world oil market, a short-term forecast is discussed in this research. The Box-Jenkins model, ARIMA, is the proper model to forecast short-term time series data; it is developed in four essential steps, they are model identification, parameter estimation, diagnostic checking, and model utilization. In time series analysis, ARIMA model is a generalization from ARMA model. Table 10 shows the major idea about model selection. ARMA model is more suitable than AR and MA, since both ACF and PACF die down.
As the ACF charts also demonstrate the non-stationary feature in the crude oil data, ARIMA model should urgently fit the oil return data. The order of model is commonly written as (p, d, q); p is the order if autoregressive, d is the differencing and q is the moving average. The default value of d is one in SAS Enterprise Miner, whereas user can vary the value to a larger integer to remove data's non-stationarity. In this research, the default value of differencing (p) is not changed, since pre-processing the data with logarithmic return method makes data stationary. In Okorie (2014), he stated that usually the first difference is sufficient to coerce a non-stationary timer series to stationary and the second difference is seldom required. In other words, this research focuses on justifying the value of p and q to build a good ARIMA forecasting model for WTI and Brent.

TABLE X. ACF and PACF chart help to identify ARIMA's Parameter

Model	**ACF**	**PACF**
AR (p)	Dies down	Cut off after lag q
MA (q)	Cut off after lag p	Dies down
ARMA (p,q)	Dies down	Dies down

In concept, the value range of p and q should be between zero and two inclusive, or the total number of parameters should be below three. Box and Jenkins (1976) presented partial autocorrelation (PACF) statistics and autocorrelation test (ACF) methods to build ARIMA. WTI's ACF and PACF chart are displayed in Figure 5 and Figure 11 separately. In WTI's ACF, there is one spike indicating MA (1); in PACF, there is one spike, which means AR (1). With all parameters defined, the WTI market's ARIMA model is $ARIMA(p, d, q) = (1, 1, 1)$. Brent's ACF and PACF chart are shown in Figure 6 and Figure 12. There are two spikes in PACF and ACF, so Brent's ARIMA $(p, d, q) = (2, 1, 2)$.
Another method to define ARIMA parameters is p, indicating the cut-off lag in the PACF, and q, indicating the decay number in ACF. In WTI's ACF and PACF chart, $p = 3$, $q = 1$. In Brent's ACF and PACF, it's $p = 6$, and $q = 1$. Though the ACF and PACF provide the values of parameters directly, visual checks are insufficiently accurate. By comparing several different ARIMA models, the one with the lowest AIC is considered the optimal model. Table 11 shows the different attempts with their own AIC. Initially, ARIMA $(1, 1, 1)$ is fitted to the data for the benchmark in both markets. After trying different p and q values, ARIMA $(3, 1, 5)$ has the lowest AIC value, 13540.24, in the WTI market; ARIMA $(1, 1, 1)$ performs best in the Brent market. After defining the ARIMA's parameters, the model requires the absolute value of t to be greater than 1.96, and the p-value should be less than 0.05. If the statistical data do not satisfy these two conditions, the model is inadequate.

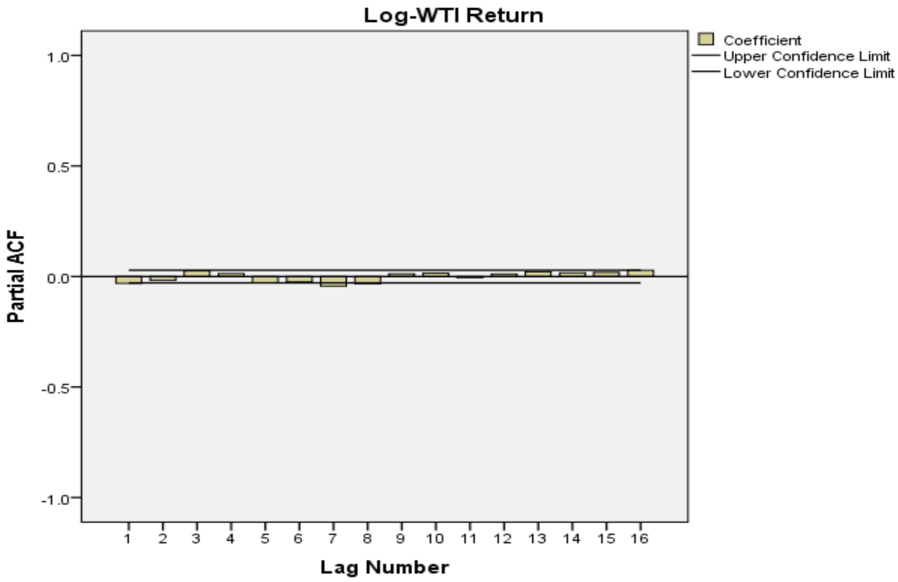

FIG. 11. WTI Oil Return PACF

FIG. 12. Brent Oil Return PACF

TABLE XI. ACF and PACF chart help to identify ARIMA's Parameter

Model	**ACF**	**PACF**
AR (p)	Dies down	Cut off after lag q
MA (q)	Cut off after lag p	Dies down
ARMA (p,q)	Dies down	Dies down

TABLE XII. Brent ARIMA Model's Parameter Selection

ARIMA	**p**	**d**	**q**	**AIC**
Attempt 1	1	1	1	12805.11
Attempt 2	2	1	2	12807.2
Attempt 3	1	1	2	12806.73
Attempt 4	2	1	1	12806.54
Attempt 5	1	1	3	14611.02
Attempt 6	7	1	7	15914.23

5.2.3. Unit Root Test

Unit root test is the stationary test in the ARIMA model. In usual, the test is applied before identifying the ARIMA model, so user can identify the order of difference. However, unit root test prevents over differencing in this case. For WTI and Brent markets, the $d = 1$ is employed to test, first ordinary difference, is sufficient to transform the data. In the unit root test, the hypothesis is presented below:

$$H_0 : \text{The time series data is not stationary}$$

$$H_1 : \text{The time series data is stationary}$$

The default significant value in the SAS Enterprise Miner is 0.05. In the Table 13, WTI's ARIMA $(3, 1, 5)$ p-value is below 0.05. Thus, null hypothesis is rejected, WTI's data is stationary. Same hypothesizes are applied in Brent's unit root test, and oil return is stationary.

TABLE XIII. WTI ARIMA(3,1,5) Unit Root Test

Type	**Lags**	**Rho**	**Pr ¡ Rho**	**Tau**	**Pr ¡ Tau**	**F**	**Pr ¿ F**
Zero Mean	0	-4773.39	0.0001	-70.35	$< .0001$		
	1	-4995.12	0.0001	-49.97	$< .0001$		
	2	-4402.83	0.0001	-38.58	$< .0001$		
Single Mean	0	-4773.54	0.0001	-70.35	$< .0001$	2474.51	0.001
	1	-4995.67	0.0001	-49.97	$< .0001$	1248.47	0.001
	2	-4403.78	0.0001	-38.58	$< .0001$	744.23	0.001
Trend	0	-4774.59	0.0001	-70.36	$< .0001$	2475.07	0.001
	1	-4999.28	0.0001	-49.98	$< .0001$	1249.12	0.001
	2	-4410.05	0.0001	-38.59	$< .0001$	744.79	0.001

TABLE XIV. Brent ARIMA (1, 1, 1) Unit Root Test

Type	**Lags**	**Rho**	**Pr ¡ Rho**	**Tau**	**Pr ¡ Tau**	**F**	**Pr ¿ F**
Zero Mean	0	-4532.1	0.0001	-66.77	$< .0001$		
	1	-4426.5	0.0001	-47.03	$< .0001$		
	2	-4475	0.0001	-38.79	$< .0001$		
Single Mean	0	-4532.4	0.0001	-66.77	$< .0001$	2228.93	0.001
	1	-4427.3	0.0001	-47.03	$< .0001$	1106	0.001
	2	-4476.7	0.0001	-38.79	$< .0001$	752.43	0.001
Trend	0	-4533.5	0.0001	-66.78	$< .0001$	2229.5	0.001

6. Model Validation and Result Analysis

In this research, both the univariate and multivariate models are adopted in the modeling step whereas, the data pre-processing methods are different, univariate model's forecasting performance is not comparable to multivariate. In other words, models imputing normalized data are comparable to each other, so the linear regression models' forecasting performance is compared with ANN class models'. Models, ARIMA and ES, using stationary data, are comparable to each other. After comparing the out-of-sample validation, the 10-day ahead and 20-day ahead forecasting accuracy will be compared.

6.1. Multivariate Models' Out-of-sample Performance

6.1.1. Multiple Linear Regression and Quadratic Linear Regression

As Table 15 shows, stepwise quadratics linear regression has the lowest SSE, MSE, and RMSE among five regression-class forecasting models. Based on the validation data's performance metrics, the ranking of favorable regression models is stepwise QLR ¿ Backward QLR ¿ Forward QLR ¿ QLR ¿ MLR. MLR's error rate is larger than the QLR types model, which illustrates the more complex models are prioritized to forecast WTI oil prices since the year of 2000. Similar to WTI, Brent also showed that the quadratic regression fits better than the MLR. Both the forward and stepwise quadratics linear regressions perform well in the validation dataset. The reason that SAS Miner Enterprise software selects the forward quadratics linear regression as the optimal model is the slightly difference in performance metric cannot be seen in the table. In brief, the complex form of regression model, quadratic regression exhibits higher forecasting accuracy than MLR for WTI and Brent.

TABLE XV. WTI Regression-Class Models' Performance Evolutions

USE	Model Name	Validation SSE	Validation MSE	Validation RMSE
Y	Stepwise Quadratic Linear Regression	3.36E-07	6.62E-10	2.57E-05
	Forward Quadratic Linear Regression	3.67E-07	7.23E-10	2.69E-05
	Backward Quadratic Linear Regression	3.20E-07	1.22E-09	3.49E-05
	Quadratic Linear Regression	1.14E-06	2.25E-09	4.74E-05
	Multiple Linear Regression	4.13E-04	0.000000812	0.000901

TABLE XVI. Brent Regression-Class Models' Performance Evolutions

USE	Model Name	Validation SSE	Validation MSE	Validation RMSE
Y	Forward Quadratic Linear Regression	2.71E-07	5.34E-10	2.31144E-05
	Stepwise Quadratic Linear Regression	2.71E-07	5.34E-10	2.31144E-05
	Backward Quadratic Linear Regression	5.22E-07	1.03E-09	3.20405E-05
	Quadratic Linear Regression	1.03E-06	2.03E-09	4.50781E-05
	Multiple Linear Regression	0.006774803	1.33362E-05	0.00365188

6.1.2. Neural Network

The validation performance metrics table shows that the generalized neural network has lower error rate than multilayer neural network for both WTI and Brent prices. The simple form of ANN models without the hidden layer have higher forecasting accuracy than multilayer ANN. One explanation can be the multilayer ANN is too complicated to reach the out-of-sample forecasting success. In other words, ANN model has capability to captures the optimal result in training data, but can be overfitting in the validation or testing dataset. In the modelling process, the software selecting top 80% of data as training data, the remaining part as test data. The training data includes floated period, like economic recession in 2008; however, the prices in validation data are more stable than them in training data. In summary, the data partition method highly possible lead the multilayer ANN forecasting accuracy lower than generalized ANN.

TABLE XVII. WTI ANN-class Model Performance Evaluation

Use	Model Name	Validation SSE	Validation MSE	Validation RMSE
Y	Generalized Neural Network	0.001083	1.06E-06	0.00103173
	Multilayer Neural Network (n=6)	0.007166	0.00000882	0.00297014

TABLE XVIII. Brent ANN-class Model Performance Evaluation

Use	Model Name	Validation SSE	Validation MSE	Validation RMSE
Y	Generalized Neural Network	0.016479804	1.62043E-05	0.00402546
	Multilayer Neural Network (n=4)	0.020168162	0.000019831	0.004453205

6.2. Univariate Model's Performance

6.2.1. Exponential Smoothing

In SAS Miner Enterprise, Mean Square Error is set up as the selection criteria in Exponential Smoothing models. As a result, additive seasonal ES has the lowest error rate among all ES models for both WTI and Brent markets. Among all ES models, additive winters ES has the highest similarity to additive seasonal ES, since both are triple ES models. The only difference is additive winters ES considers three parameters (level mean, trend, seasonality); additive seasonal ES only considers level mean and seasonality. In additive winters ES, WTI and Brent's trend is going downward; however, the additive seasonal ES forecasts two market's prices as horizontal lines. In brief, the model without trend exhibits better forecasting accuracy than with trend under the same level means, 0.05. This finding illustrates that seasonality heavily impacts crude oil prices instead of trend.

TABLE XIX. WTI Additive Seasonal ES Model Performance Evaluation

Market	Model	Validation SSE	Validation MAE	Validation MSE
WTI	Additive Seasonal ES	277.2695523	0.556819168	0.606716745

TABLE XX. Brent Additive Seasonal ES Model Performance Evaluation

Market	**Model**	**Validation SSE**	**Validation MAE**	**Validation MSE**
Brent	Additive Seasonal ES	265.1664337	0.548924087	0.580232897

6.2.2. ARIMA

By several attempts, WTI ARIMA $(3,1,5)$ and Brent ARIMA $(1,1,1)$ have the lowest AIC. After defining the parameters, the user must do diagnostic checking. One primary tool to do diagnostic checking is residual analysis. The aim of the analysis is to realize whether the residuals are uncorrelated (white noise) or contain additional information that may be adopted by a more complex model (SAS Institute Inc, n.d.). In a well-fitted model, the residuals are expected to exhibit the property of white noise. The null hypothesis is that errors are random, which is white noise. The significance level will be 0.05; if the result is above 0.05, it means to accept the null hypothesis, and the white noise exists. The autocorrelation check of residuals for ARIMA $(3,1,5)$ is listed in Table 21. In this case, the values for the first 12 lags are higher than 0.05, so the null hypothesis is accepted. The ARIMA $(1,1,1)$ autocorrelation check of residual for Brent is shown in Table 22. As the values for the first 12 lags are greater than the significance level, the null hypothesis is accepted, indicating that white noise exists in the model.

TABLE XXI. ARIMA (3,1,5) Autocorrelation Check of Residuals

To Lag	**Chi-Square**	**DF**	**Pr ¿ ChiSq**	**Autocorrelation**					
6	9.16	4	0.0573	-0.036	-0.019	0.001	0.012	0	-0.014
12	17.14	10	0.0714	-0.028	-0.028	0.006	-0.011	0	0.002
18	26.85	16	0.0432	-0.003	0.028	0.028	0.006	0.003	0.023
24	35.42	22	0.035	-0.001	0.01	-0.011	-0.017	0.011	-0.035
30	42.02	28	0.0432	0.021	0.018	0.025	-0.003	0.007	0.003
36	56.43	34	0.0092	-0.018	0.018	-0.044	0.007	-0.006	-0.021
42	82.27	40	$< .0001$	-0.018	0.055	0.019	0.039	0.017	0.011
48	88.15	46	0.0002	-0.004	0.017	0.018	-0.023	0.006	0.009

TABLE XXII. Brent (1,1,1) Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr ¿ ChiSq	Autocorrelations					
6	6.27	4	0.1798	0.005	0.002	-0.012	0.013	0.001	-0.032
12	10.27	10	0.4168	0.023	-0.01	-0.006	-0.005	0	-0.014
18	26.37	16	0.049	0.012	0.043	0.025	0.021	-0.019	0.001
24	40.2	22	0.0103	-0.017	0.002	-0.026	-0.032	-0.031	0.007
30	48.54	28	0.0094	0.03	0.026	0.002	-0.008	0.011	0.006
36	57.9	34	0.0065	-0.005	-0.005	-0.024	-0.01	-0.013	-0.034
42	65.94	40	0.006	0.01	0.004	0.018	0.032	-0.013	0.007
48	69.8	46	0.0134	0.005	-0.001	0.011	-0.024	0.002	-0.009

The taken ARIMA models apply to forecast crude oil volatilities, and then the out-of-sample validation performance need to be evaluated. The validation data is about 10% at the end of observations periods, and it has not been used in modelling step. In SAS Enterprise Miner, user can set up a specific amount of observations to forecast ahead. The forecasting and actual data is compared under several evaluation metrics, SSE, MAE, MSE, RMSE. Table 23 and Table 24 shows the selected ARIMA model's performance in the validation dataset. The finding is ARIMA $(1, 1, 1)$ in Brent market are more accurate in forecasting out-of-sample data than that of WTI ARIMA $(3, 1, 5)$.

TABLE XXIII. WTI ARIMA (3, 1, 5) Out-of-sample Performance

Market	Model	Validation SSE	Validation MAE	Validation MSE	Validation RMSE
WTI	ARIMA (3, 1, 5)	274.73611	0.54995	0.601173113	0.775353541

TABLE XXIV. Brent ARIMA (1, 1, 1) Out-of-sample Performance

Market	Model	Validation SSE	Validation MAE	Validation MSE	Validation RMSE
Brent	ARIMA (1, 1, 1)	262.3164811	0.543547196	0.573996676	0.757625683

6.2.3. Performance Comparison

ES and ARIMA models' results are comparable, since both of them are using the same stationary data in modelling step. In line with the evaluation metrics tables, ARIMA model has better out-of-sample performance than the additive seasonal ES. Overall, ARIMA $(3, 1, 5)$ is the best univariate forecasting model in WTI; ARIMA $(1, 1, 1)$ is the best univariate forecasting model in Brent.

6.3. 10-day Ahead Forecasting and 20-day Ahead Forecasting

In 10-day and 20-day ahead forecasting, regression-class models and ANN models are excluded. The main reason is ANN as a black box model is challenging to interpret the model and its result. In this research, the complex forecasting models are used for short-term forecasting. Additive seasonal ES and ARIMA models' 10-day ahead forecasting performance are listed in Tables 25 to 28. From the table, additive seasonal ES is more accurate in forecasting 10-day ahead oil price for WTI and Brent, since the lower values of SSE, MAE, MSE, and RMSE. For 20-day ahead forecasting, the models' performance is listed in Tables 29 to 32. Similar to 10-day ahead forecasting, additive seasonal ES has lower error rates than the ARIMA model. In summary, the additive seasonal ES model is more accurate in forecasting short-term light oil price than the ARIMA model.

TABLE XXV. WTI Additive Seasonal ES 10 day-ahead forecasting

Market	**Model Name**	**Test SSE**	**Test MAE**	**Test MSE**	**Test RMSE**
WTI	Additive Seasonal Exponential Smoothing	8.859644038	0.727224537	0.885964404	0.941256821

TABLE XXVI. WTI ARIMA(3,1,5) 10 day-ahead forecasting

Market	**Model Name**	**Validation SSE**	**Validation MAE**	**Validation MSE**	**Validation RMSE**
WTI	ARIMA (3,1,5)	9.365580472	0.765451848	0.936558047	0.967759292

TABLE XXVII. Brent Additive Seasonal ES 10 day-ahead forecasting

Market	**Model Name**	**Validation SSE**	**Validation MAE**	**Validation MSE**	**Validation RMSE**
Brent	Additive Seasonal Exponential Smoothing	6.956191999	0.674958578	0.6956192	0.834037889

TABLE XXVIII. Brent (1,1,1) 10 day-ahead forecasting

Market	Model	Validation SSE	Validation MAE	Validation MSE	Validation RMSE
Brent	ARIMA(1,1,1)	7.311084155	1.650101241	0.731108415	0.85504878

TABLE XXIX. WTI Additive Seasonal ES 20-day Ahead Forecasting

Market	Model Name	Validation SSE	Validation MAE	Validation MSE	Validation RMSE
WTI	Additive Seasonal Exponential Smoothing	19.92866683	0.736018434	0.996433342	0.998215078

TABLE XXX. WTI ARIMA(3,1,5) 20-day Ahead Forecasting

Market	Model	Validation SSE	Validation MAE	Validation MSE	Validation RMSE
WTI	ARIMA (3,1,5)	20.20635494	0.744674277	1.010317747	1.005145635

TABLE XXXI. Brent Additive Seasonal ES 20-day Ahead Forecasting

Market	Model Name	Validation SSE	Validation MAE	Validation MSE	Validation RMSE
Brent	Additive Seasonal Exponential Smoothing	12.12291982	0.638066067	0.606145991	0.778553782

TABLE XXXII. Brent ARIMA(1,1,1) 20-day Ahead Forecasting

Market	Model	Validation SSE	Validation MAE	Validation MSE	Validation RMSE
Brent	ARIMA (1,1,1)	12.82861335	0.647933011	0.641430667	0.800893668

7. Conclusion

The purpose of this research is to find the best forecasting models for Brent and WTI, and also the model for 10 and 20 day-ahead forecasting. Varied types of forecasting models are including in the research, they are linear, advanced time series, and machine learning models, but each model has its own advantage and disadvantage in forecasting oil prices. In the multivariate models, MLR and Quadratics regression are classical statistics models, but unable to capture the nonlinear relation in the oil prices. In contrast, ANN-class models are able to capture the nonlinear relation, but they are challengeable to user for interpreting. Among MLR, quadratics regression, and ANN-class model, the quadratic regression is the best forecasting model which has lowest MSE, RMSE, and SSE in validation dataset. Model testing is not applying in multivariate models, since In the univariate time series model, ES-class models and ARIMA are fitting into the logarithmic return. ARIMA models predicting the WTI and Brent prices better in validation data, but additive seasonal ES models have lower error rate in both 10 and 20 day ahead forecasting. Overall, stepwise and forward quadratic regression are the best multivariate forecasting model for WTI and Brent individually. In univariate time series models, ARIMA (1,1,1) and ARIMA(3,1,5) are appropriate models in the forecasting; additive seasonal ES is excellent in forecasting 10 and 20 day-ahead WTI and Brent oil price.

References

[1] Arezki, R., et al. (2017). *Oil prices and the global economy.* International Monetary Fund. Retrieved from `https://www.imf.org/external/pubs/ft/wp/2016/wp16210.pdf`. Accessed 17 June, 2019.
[2] Achraf, G., & Boujelbene, Y. (2013). Contagion effect of the oil shock and US financial crisis on the GCC and BRIC countries. *International Journal of Energy Sector Management*, 7(4), 430–447.
[3] Adelman, M.A. (1984). International oil agreements. *The Energy Journal*, 5(3), 1–9.
[4] Adelman, M.A. (1994). The Economics of Petroleum Supply. *The Energy Journal*, 15(2), 238–240.
[5] Ahmed, A.R., & Shabri, B.A. (2014). Daily crude oil price forecasting model using ARIMA, generalized autoregressive conditional heteroscedastic and support vector machines. *American Journal of Applied Sciences*, 11(3), 425–432.
[6] Akpanta, A.C., & Okorie, I.E. (2014). Application of Box-Jenkins techniques in modelling and forecasting Nigerian crude oil prices. *International Journal of Statistics and Applications*, 4(6), 283–291.
[7] Al-Maamary, M.S.H., Kazem.A.H., & Chaichan, T.M. (2017). The impact of oil price fluctuations on common renewable energies in GCC countries. *Renewable and Sustainable Energy Reviews*, 75, 989–1007.
[8] Alterman, S. (2012). Natural Gas price Volatility in the UK and North America. The Oxford Institute for Energy Studies. Retrieved from `https://www.oxfordenergy.org/wpcms/wp-content/uploads/2012/02/NG_60.pdf`. Accessed 12 June, 2019.
[9] Baillie, R.T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1), 5–59.
[10] Billah, B., et al. (2006). Exponential smoothing model selection for forecasting. *International Journal of Forecasting*, 22, 239–247.

[11] Billah, M.B., Hyndman, R.J., & Koehler, A.B. (2003). Empirical information criteria for time series forecasting model selection. *Journal of Statistical Computation and Simulation*, 75(10), 831–840.

[12] Bollerslev, T., & Mikkelsen, O.H. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73, 151–184.

[13] Boshnakov, N.G. (2019). Time Series Analysis and Financial Forecasting, lecture note, School of Mathematics, University of Manchester, delivered 01 July 2019.

[14] Brocklebank, J.C., Dickey, D.A., & Choi, B.S. (2018). *SAS for forecasting time series*. 3rd ed. Cary: SAS Institute Inc. Retrieved from `https://support.sas.com/content/dam/SAS/support/en/books/sas-for-forecasting-time-series-third-edition/68469_excerpt.pdf`. Accessed 5 July, 2019.

[15] Cabedo, J.D., & Ismael, M. (2003). Estimating oil price 'Value at Risk' using the historical simulation approach. *Energy Economics*, 25, 239–253.

[16] Calvet, L.E., & Fisher, A.J. (2001). Forecasting multifractal volatility. *Journal of Economics*, 105, 27–58.

[17] Cao, B.Y., & Nasseri, H. (2014). *Fuzzy information & engineering and operations research & management*. Berlin: Springer.

[18] Chen, C., Twycross, J., & Garibaldi, J.M. (2017). A new accuracy measure based on bounded relative error for time series forecasting. *Plos One*, 12(3).

[19] Chen, Y.H., He, K.J., & Tso, K.F.G. (2017). Forecasting crude oil prices: a deep learning based model. *Procedia Computer Science*, 122, 300–307.

[20] Dickey, A.D. (n.d.). Stationarity issues in time series models. [online]. Available from: `https://pdfs.semanticscholar.org/a37e/31ed2969afce603128b4c871b1f1e332bf6f.pdf` [accessed 10 July 2019].

[21] Ederington, L.H., et al. (2011). Factors influencing oil prices: a survey of the current state of knowledge in the context of the 2007-08 oil price volatility. U.S. Energy Information Administration. [online]. Available at: `https://www.eia.gov/workingpapers/pdf/factors_influencing_oil_prices.pdf` [Accessed 15 Jun. 2019].

[22] Elder, J., & Serletis, A. (2010). Oil price uncertainty. *Journal of Money, Credit and Banking*, 42(6), 1137–1159.

[23] Engle, F.R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.

[24] Engle, R.F., & Lee, G.G. (1992). A permanent and transitory component model of stock return volatility. [e-book]. California: University of California. Available from: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5848` [accessed 20 June 2019].

[25] Fattouh, B. (2010). The dynamics of crude oil price differentials. *Energy Economics*, 32, 334–342.

[26] Gabralla, L., & Abraham, A. (2013). Computation modeling of crude oil price forecasting: a review of two decades of research. *International Journal of Computer Information Systems and Industrial Management Applications*, 5, 729–740.

[27] Gulen, S.G. (1999). Regionalization in the world crude oil market: further evidence. *The Energy Journal*, 20(1), 125–139.

[28] Hamdi, M., & Aloui, C. (2015). Forecasting crude oil using artificial neural network: a literature survey. *Economics Bulletin*, 35(2), 1339–1359.

[29] Hamilton, D.J. (1983). Oil and the macroeconomy since World War II. *Journal of Political Economy*, 91(2), 228–248.

[30] Herrera, G.P., et al. (2019). Long-term forecast of energy commodities price using machine learning. *Energy*, 179, 214–221.

[31] Ho, S.L., & Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. *Computers & Industrial Engineering*, 35(1-2), 213–216.

[32] Holt, C.C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20, 5–10.

[33] Hooker, A.M. (1996). 'What Happened to the oil price-macroeconomy relationship?' *Journal of Monetary Economics*, 38, 195–213.

[34] Kajitani, Y., Hipel, K.W., & Mcleod, I.A. (2005). Forecasting nonlinear time series with feed-forward neural networks: a case study of Canadian Lynx data. *Journal of Forecasting*, 24, 105–117.

[35] Kang, S.H., Kang, S.M., & Yoon, S.M. (2009). Forecasting volatility of crude oil markets. *Energy Economics*, 31(1), 119-125.

[36] Kulkarni, S., & Haidar, Imad. (2009). Forecasting model for crude oil price using artificial neural networks and commodity futures prices. *International Journal of Computer Science and Information Security*, 2(1).

[37] Lanza, A., Manera, M., & Giovannini, M. (2005). Modeling and forecasting cointegrated relationships among heavy oil and product prices. *Energy Economics*, 27, 831-848.

[38] Lee, Y.H., Hu, H.N., & Chiou, J.S. (2010). Jump dynamics with structural breaks for crude oil prices. *Energy Economics*, 32, 343-350.

[39] Lux, T. (2012). The Markov-Switching Multifractal model of asset return. *Journal of Business & Economic Statistics*, 26(2), 194-210.

[40] Maradiaga, D., Pujula, A., & Zapata, H. (2013). Exploring time series data properties in SAS. *SAS Global Forum 2013*. [online]. Available from: http://support.sas.com/resources/papers/proceedings13/456-2013.pdf [accessed 05 July 2019].

[41] Minimol, M.C. (2018). Relationship between spot and future prices of crude oil: a cointegration analysis. *Theoretical Economics Letters*, 8, 330-339.

[42] Natarajan, G.S., & Ashok, A. (2018). Multivariate forecasting of crude oil spot prices using neural networks. *Computer Science, Cornell University*. [online]. Available from: https://arxiv.org/abs/1811.08963 [accessed on June 16, 2019].

[43] Nomikos, K.N., & Pouliasis, K.P. (2011). Forecasting petroleum futures markets volatility: the role of regimes and market conditions. *Energy Economics*, 33, 321-337.

[44] Nyangarika, M.A., & Tang, B.J. (2018). Oil price factors: forecasting on the base of modified ARIMA model. *Earth and Environment Science*, 192.

[45] Pagan, A. (2002). Cointegration, causality and forecasting: A Festschrift in Honor of Clive W. J. Granger. *The Economic Journal*, 112(480).

[46] Panas, E., & Ninni, V. (2000). Are oil markets chaotic? A nonlinear dynamic analysis. *Energy Economics*, 22, 549-568.

[47] Pindyck, R.S. (2004). Volatility in natural gas and oil markets. *The Journal of Energy and Development*, 30(1), 1-19.

[48] Rey, T., Kordon, A., & Wells, C. (2012). Applied data mining for forecasting using SAS. [online]. Cary: SAS Institute Inc. Available from: https://www.sas.com/store/prodBK_63533_en.html?storeCode=SAS_US [accessed 01 July 2019].

[49] Rodriguez, A.E., & Williams, M.D. (1993). Is the world oil market 'one great pool'? a test. *Energy Studies Review*, 5(2), 121-130.

[50] S. Moshiri, & F. Foroutan. (2006). Forecasting nonlinear crude oil future prices. *The Energy Journal*, 27(4), 81-95.

[51] Sadorsky, P. (1999). Oil price shocks and stock market activity. *Energy Economics*, 21, 449-469.

[52] Salisu, A.A., & Fasanya, I.O. (2013). Modelling oil price volatility with structural breaks. *Energy Policy*, 52, 554-562.

[53] Scheid, B., & Gordon, M. (2018). STEO highlights: EIA cuts Brent, WTI 2019 forecasts nearly $11/b amid supply glut. [online]. Available from: https://www.spglobal.com/platts/en/market-insights/latest-news/oil/121118-steo-highlights-eia-cuts-brent-wti-2019-forecasts-nearly-11-b-amid-supply-glut [accessed 21 June 2019].

[54] Schubert, S., & Lee, T.Y. (2011). Time series data mining with SAS Enterprise Miner. *SAS Global Forum 2011*. [online]. Available from: http://support.sas.com/resources/papers/proceedings11/160-2011.pdf [accessed 27 June 2019].

[55] Seeking Alpha. (2016). Forecasting Crude Oil Price with an Autoregressive Integrated Moving Average (ARIMA) Model. [online]. Available from: https://seekingalpha.com/article/4014343-arima-modeling-forecast-future-oil-prices [accessed 25 July, 2019].

[56] Selvi, J.J., Shree, R.K., & Krishnan, J. (2018). Forecasting crude oil price using ARIMA models. *International Journal of Advance Research in Science and Engineering*, 07(05), 334–343.

[57] Slim, C. (2015). Improved crude oil price forecasting with statistical learning methods. *Journal of Modern Accounting and Auditing*, 11(1), 51–62.

[58] Tang, L., Wang, C., & Wang, S. (2013). Energy time series data analysis based on a novel integrated data characteristic testing approach. *Procedia Computer Science*, 17, 759–769.

[59] Tularam, G.A., & Saeed, T. (2016). Oil-price forecasting based on various univariate time-series models. *American Journal of Operations Research*, 6, 226–235.

[60] Wang, J., & Wang, J. (2016). Forecasting energy market indices with recurrent neural networks: Case study of crude oil price fluctuations. *Energy*, 102, 365–374.

[61] Wang, Y.D., & Wu, C.F. (2012). Forecasting energy market volatility using GARCH models: Can multivariate models beat univariate models?. *Energy Economics*, 34, 2167–2181.

[62] Wang, Y.D., Wu, C.F., & Yang, L. (2016). Forecasting crude oil market volatility: A Markov switching multifractal volatility approach. *International Journal of Forecasting*, 32, 1–9.

[63] Weiner, J.R. (1991). Is the world oil market 'one great pool'/?. *The Energy Journal*, 12(3), 95-107.

[64] Wilmot, A.N. (2013). Cointegration in the oil market among regional blends. *International Journal of Energy Economic and Policy*, 3(4), 424–433.

[65] Xiong, Y., et al. (2016). Application of structural breakpoint test to the correlation analysis between crude oil price and U.S. weekly leading index. *Open Journal of Business and Management*, 4, 322–328.

[66] Yasin, H., et al. (2016). Prediction of crude oil prices using support vector regression (SVR) with grid search-cross validation algorithm. *Global Journal of Pure and Applied Mathematics*, 12(4), 3009–3020.

[67] Yusof, N.M., Sulong, R., & Mohamed, R.Z. (2010). Malaysia crude oil production estimation an application of ARIMA model. International conference on social research. Kuala Lumpur, Malaysia. 5th-7th December 2010.

[68] Yusof, S.F., & Samsudin, R. (2018). Comparison of artificial neural network (ANN) and autoregressive integrated moving average (ARIMA) models for WTI crude oil price forecasting. *Innovations in Computing Technology and Applications*, 3.