International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

Crude Oil Price Forecasting using ARIMA model

Jessin Shah P A[1], Dr. G Kiruthiga[2]

[1]Department of Computer Science & Engineering, IES College of Engineering and Technology, Chittilappilly
[2] HOD, Dept. of Computer Science & Engineering, IES College of Engineering and Technology, Chittilappilly

---

Abstract – *Oil plays a vital role in the production of many items that we use in regular basis. The increase and decrease of the price of oil influence many sectors of the economy. There are many methods to model and forecast time series data. In this paper, we examine the time series and non-linear feature of oil price, and use ARIMA model to make prediction. The results show that ARIMA model yields best performance.*

Key Words: oil price forecast, Time –series, AR, MA, ARIMA model.

1. INTRODUCTION

Crude oil is an unprocessed thick liquid fuel source extracted directly from the oil reservoirs. Oil is used for transportation, petroleum products and plastics. Oil is the base for over 6000 items. There are different types of oil. West Texas Intermediate crude oil is of very high quality. It is the major benchmark of crude oil in the Americas. Brent Blend is a combination of crude oil from fifteen different oil fields in North Sea and is the primary benchmark for crude oils in Europe and Africa. Oil is one of the primary sources of enegy consumption, therefore the price of oil affect the world economy. Accounting for a large part of exports for some countries, a sudden changes of crude oil price can bring significant economic consequences, with crashes of crude oil prices bringing about slower economic activity[1] and booms of crude oil prices causing significant inflation as well. The crude oil prices are affected by supply and demand and also depends economic activity, politics and more[2]. Accurate forecasting is important for both government and industries to make decision in economic strategy. But due to its chaotic and unpredictable nature it is difficult to predict.

Crude oil price prediction is a challenging research topic. Different machine learning models have been applied to address this problem including EEMD-SBL-ADD[3],neural network model, ARIMA model[4], support vector machines(SVM)[5] and error correction models(ECM)[6], etc.. Xie et al. compared the performance of SVM with other popular time series forecasting models, and concluded that SVM outperforms other tested models. Though Lanza At el. (2005) utilized ECMs to investigate crude oil prices, due to the irregular nature of crude oil market, method often proved to be inefficient. In[3] a novel model integrating EEMD(Ensemble Emperical Mode Decomposition), SBL(Sparse Bayesian Learning) and addition for forecasting crude oil prices were proposed and concluded that EEMD-SBL-ADD outperforms any other competitive models in terms of the evaluation criteria. A literature review on crude oil price prediction showed that for making predictions, most existing work relied on ARIMA, fully connected neural networks and recurrent neural networks[7].

Although many factors effect the price of oil, in practice it is hard to consider all the contributing factors. Therefore in this work we consider oil as a pure time series problem. SVM and ARIMA model is considered in this research and compare to figure out a best fit for this prediction.

In the following part of this paper, we will first introduce arima model which we apply in this study, next to fit time series data into models, then perform data processing. Next section contain results and comparisons. Then we conclude our study about oil price forecasting.

2. ARIMA MODEL

In time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. Both these models are used in time series data to understand data or to predict the future points in series that is to forecast. ARIMA model is the combination of autoregression (AR) and moving average(MA). AR uses the dependent relationship between an observation and some number of lagged observations. integrated is the use of differencing of raw observations in order to make the time series stationary. MA is a model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations. Standard notation used for ARIMA model is ARIMA(p, d, q) where p is the lag order that is order of AR, d is the degree of differencing to make the time series stationary, q is the size of moving average window that's the order of MA. The parameters take integer values.

ARIMA model is applied to the time series data with non-stationarity property by applying differencing on raw observations to make the time series data stationary. After differencing the ARIMA model for a serially correlated data can be expressed in the form of

$$y_t = \mu + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 e_{t-1} \cdots - \theta_q e_{t-q} \quad (1)$$

where y_t is the value at time t, e_t random error and the mean of the series. $\phi_t (t=1, 2, \cdots, p)$ and $\theta_t (t=1, 2, \cdots, q)$ are model parameters that needs to be determined from data. p and q define orders of the model.

3. PREPARE THE DATA

3.1 Data Selection

To value the commodity, three crude oil benchmarks are set, they are Brent, West Texas Intermediate(WTI) and Dubai/Oman oil benchmarks. WTI crude oil price data refers to crude oil extracted in the U.S. is used in this study.

WTI oil price from July 1987 to March 2020 is collected for the study. We can observe, the oil price increases in the period between 2000 and 2013 and then dropped sharply as shown in figure1. The drastic variations make it difficult for either traditional ARIMA models or SVM models to be able to capture those changes.

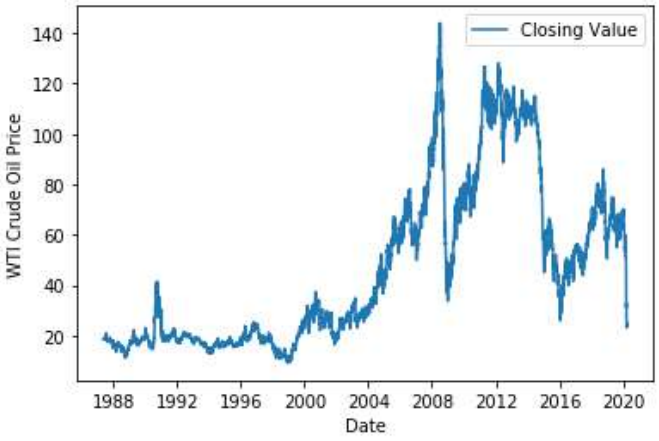

Figure 1: WTI price from 1987 to 2020

3.2 Preliminary Processing of Collected Data

Oil price have a large difference when considering the highest and lowest. To reduce the variation of original WTI data, we take log of the original data so that data can be better fit into the model. The processed log data is shown in figure 2. After taking the log , we can observe that the processed data keeps an overall increasing trend while the difference in price between early years and recent years are reduced significantly, which can facilitate data analysis and model training. Following data analysis will be based on the log data. The whole dataset is split into training and testing set. Training set accounts for 80% of the whole data and testing set accounts for 20% of the whole data set, to verify the model accuracy. This is shown in Figure 3.

Figure 2. Log value of WTI oil price from 1987 to 2020

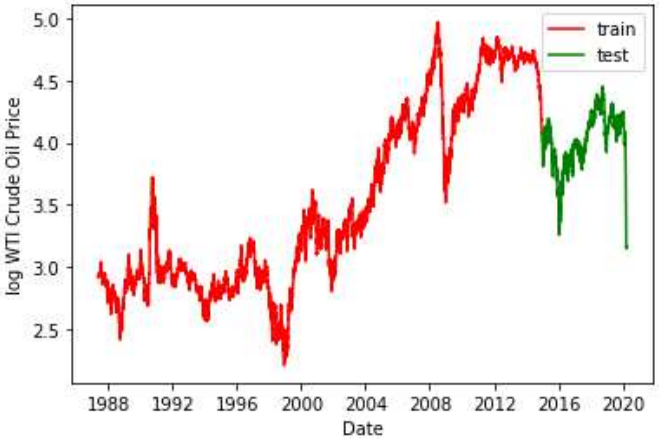

Figure 3. Train and Test Data

3.3 Data preprocessing for ARIMA Model

Standard time series requires the analysis of the stationarity of the data. The distribution of our interest value $X_{t=}$, in this study is the log oil price is independent of time t. We claim that a time series data is stationary if given any T period and time t, $(X_0, X_{1, \ldots.}, X_T)$ and $(X_t, X_{t+1, \ldots.}, X_{t+T})$ have the same distribution. From figure 2, the log WTI oil price data does not satisfy the definition of stationary data, mainly because price is independent of time, still we perform a rigorous non-stationary time series data test.

Stationarity can be examined by running Dickey – Fuller test[8] whose null hypothesis is time series data being non-stationary. We set the critical p-value to be 0.05, meaning if the test p-value is greater than 0.05, we will accept the Dickey-Fuller null hypothesis which states that data is stationary. In this case the test p-value is 0.275 proving that the WTI oil price is non stationary.

As shown in figure 4, wee further decompose log oil price data into trend, seasonality and residuals using moving average by setting the seasonality frequency to 360 days. After taking trend and seasonality, the residuals shows characteristics of the stationarity, further validated by

conducting Dickey- Fuller test with result p-value smaller than 0.05. As the residual term shows to be stationary after a single decomposition, we can conclude that number of nonseasonal difference d required in the ARIMA model is 1.

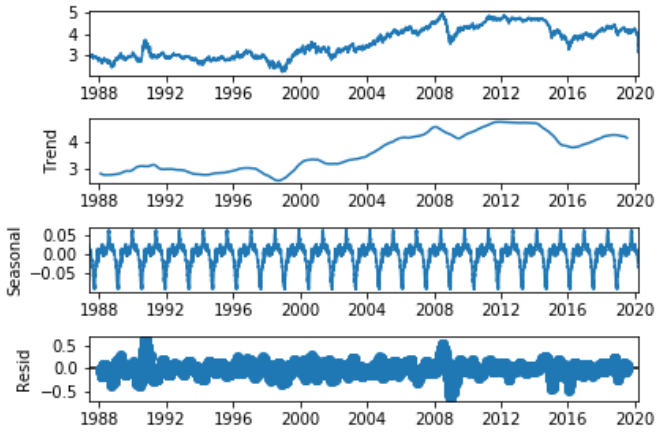

Figure 4. Trend, seasonality and residuals decomposition of log price data

4. ARIMA RESULT

In section 3, we determined the order d to be 1 and instead of performing autocorrelation and partial autocorrelation to determine the order p and q by grid search and determined p=0 and q=4 for the log transformed crude oil price data. The forecast result for the test period is shown in fiure 5. The mean_Squared value is 1.606.

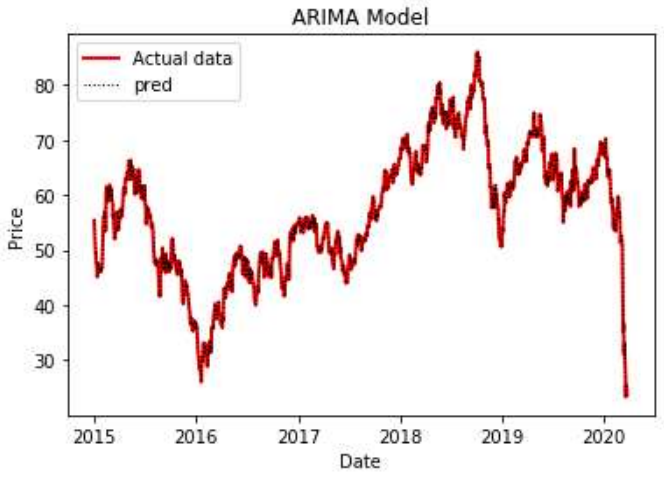

Figure 5: Arima forcast results on test data period

CONCLUSIONS

Crude oil is one of the most important chemical and energy resources. The crude oil and its price affect the economic and social activities, therefore crude effective crude oil forecast can help stabilize economic development and prevent energy crisis. The WTI crude oil data from the period of 1987 to 2020 were analyzed and showed the non stationarity of data. We conducted data analysis and preprocessing prepared for ARIMA.

In future, we will explore several other different models and will apply the techniques explored in this study to other time series problems including Stock price prediction.

ACKNOWLEDGEMENT

This work would not be possible without the support from my parents, husband's family, tutors and HOD of IES Engineering College and my friends. I am grateful for my parents and my husband, Mr. Mohammed Ganeesh, for their financial support in providing necessary hardware in this research. I would specially thank Dr. Kiruthiga who guided me through the whole project.

REFERENCES

[1] International Monetary Fund(IMF), The impact of Higher oil prices on the global economy, A report prepared by the research department of IMF, 2000

[2] E. Panas, & V. Ninmi, "Are oil markets chaotic? A non-linear dynamic analysis," Energy economics, 22(5), 549-568, 2000.

[3] Taiyong Li, Zhenda Hu, Yanchi Jia , Jiang Wu and Yingrui Zhou , "Forecasting Crude Oil Prices Using Ensemble Empirical Mode Decomposition and Sparse Bayesian Learning," article published in energirs on 19 july 2018.

[4] Junhui Guo. "Oil price forecast using depp learning and ARIMA", Shanghai American School, in International Conference on Machine Learning, Big Data and Buisiness Intelligence, 2019

[5] W. Xie, L Yu, S. Xu & S. Wang, " A new method for crude oil price forecasting based on support vector machines," In Internationa conference on Computational Science (pp. 444-451). Springer Berlin, Heidelberg, May 2006

[6] A. Lanza, M. Manera & M. Giovannini, "Modeling and forecasting cointegrated relationships among heavy oil and product prices", Energy Economics, 27(6), 831-848, 2005.

[7] Chiroma, Haruna, et al. "A review on artificial intelligence methodologies for the forecasting of crude oil price." Intelligent Automation & Soft Computing 22.3(2016):449-462.

[8] X. FRANCIS, " On the power of Dickey- Fuller tests against fractional alternatives," Business Cycles: Durations, Dynamics and Forecasting, p.258, 1999