

Volatility of daily and Weekly Crude Oil Prices – A Statistical Perspective

Kanthala Sampath Kumar^{1*}, V. Swapna², S.A. Jyothi Rani³ and V.V. Hara Gopal⁴

¹ Department of Applied Statistics, Telangana University, Nizamabad, India

² Department of H& S, CVR College of Engineering, Mangalpally (v), Hyderabad, India

³ Department of Statistics, Osmania University, Hyderabad, India

⁴ Department of Mathematics, BITS-PILANI, Hyderabad Campus, Hyderabad, India

*Corresponding Author: swapnacvr82@gmail.com,

Available online at: www.isroset.org

Received: 29/Sept/2018, Accepted: 14/Oct/2018, Online: 31/Oct/2018

Abstract— Crude oil is a key commodity for the economy of any country. The changes in Crude oil prices are very complex and therefore, seem to be unpredictable. The effect of increasing price and its daily fluctuations affects not only the economies and financial markets but extends to reach individuals. This is because an increase in oil prices has a direct effect on petrol prices, in addition, it also affects the prices of other goods and services. Therefore, forecasting crude oil is a very important task to reduce the impact of price fluctuations, and help investors, hedgers, and individuals to make decisions when dealing with energy markets. However, one of the main challenges of the econometric models is to forecast such a seemingly unpredictable economic series. The econometric models have not been promising when used for forecasting, particularly in the case of complex series such as oil prices. Although linear and nonlinear time series models have performed much better job in forecasting oil prices, there is yet room for an improvement. If the data generating process is nonlinear, applying linear models could result in large forecast errors. Model specification in nonlinear modeling can also be very case dependent and time-consuming. This article proposes a novel technique for forecasting crude oil price based on Neural Networks. The study adopts the data on crude oil price of West Texas Intermediate (WTI). To evaluate the performance of the model, the study employs three measures, RMSE, MAE and MAPE. In this study, the forecasting accuracy of two models of neural networks, BPN and RNN are compared with the traditional econometric model. The results reveal that the proposed method outperforms the other in terms of forecast accuracy

Keywords— Neural networks, root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), Recurrent Neural Networks (RNN), Multi-Layer Perceptron (MLP), Oil Price and ARIMA.

I. INTRODUCTION

Crude oil is a key commodity for the economy of any country. However, a major characteristic of crude oil markets involves significant price fluctuations. This volatility of oil prices could be attributed to three main factors:

- Supply/demand imbalance, possibly caused by:
 - Economic growth.
 - Oil producing countries behaviors.
- Exogenous events (war, extreme weather...etc).
- Endogenous factors (speculation in the markets).

The effect of increasing price and its daily fluctuations affects not only the economies and financial markets but extends to reach individuals. This is because an increase in oil prices has a direct effect on petrol prices, in addition, it also affects the prices of other goods and services. Therefore,

forecasting crude oil is a very important task to reduce the impact of price fluctuations, and help investors, hedgers, and individuals to make decisions when dealing with energy markets. But forecasting crude oil is not an easy task. (Haidar, Imad, Siddhivinayak Kulkarni, and Heping Pan, 2008.)

A brief review of neural networks is given in Section 2. Section 3 explains the methodology of forecasting the crude oil price using neural networks and traditional econometric model. Results and discussions are given in Section 4, finally conclusions are collated in Section 5.

II. RELATED WORK

The importance of crude oil to the economy is reflected by a number of studies in this area. There is large and rich literature related to every aspect of crude oil. However, until recently, the majority of studies were based on either

analytical or linear models. While analytical models has failed completely in producing good forecast for crude oil, linear models such as Box-Jenkins model (ARMA) have not provided better results . Nevertheless, it is been believed that crude oil prices are nonlinear time series, hence nonlinear models such as ANN should be superior to other linear models.

Among many and different forecasting models that have been developed to predict the "black gold" price, the traditional statistical and econometric methods are the first ones to be applied by academic researchers.

Moshiri, and Foroutan [1] compared linear and nonlinear models for forecasting crude oil futures prices. The authors compared ARMA and GARCH, to ANN, and found that ANN is superior and produced a statistically significant forecast.

Xie, *et al* [2] proposed SVM model for monthly crude oil prices prediction, the authors claimed that SVM outperformed feed forward network with back prorogation (BPNN) and ARIMA for out-of-sample. However, their results were not consistent, as BPNN outperformed SVM for two of the four sub periods tested. Nonetheless, both BPNN and SVM outperformed ARIMA for all four periods.

Liu *et al* [3] presented hybrid model based on fuzzy neural model to forecast Brent crude oil prices. Three forecasting models were used viz. Redial Based Network, Markov chain based semi parametric model and wavelet analysis based forecasting model. The output of the three methods was used as an input to fuzzy neural network, while the target was the actual Brent crude oil price. The authors concluded that the nonlinear combination outperformed any single model tested. However, the authors based this conclusion on one performance metric only, the root mean square error.

Wang *et al* [4] presented a hybrid methodology to forecast crude oil monthly prices. The model consists of combination of three separate components, Web mining from which the authors extract rule based system, in addition to ANN, and ARIMA models. These three components are made to work disjointedly, and then intergraded together to get the final results. They claimed that nonlinear integration of these three models has outperformed any single one. However, in our opinion there are several issues in this system. For example, the rule base system of the text mining model2 depends on the knowledge base which was developed by human experts. This process is not only controversial, but also unreliable, because experts' opinion vary on the same problem. Moreover, neither the rules nor the knowledge base was made available to the public.

Most of the above studies were based on monthly prices. On one hand using monthly prices reduces the noise in the data,

on the other hand, it also limits the data size significantly, and to use old data from the 70's which could be irrelevant to the current economical situations. Besides, limited data also will affect the conclusion as the testing set is barely statistically significant. Most of the studies were concentrated on developing new techniques, but little attention was paid in testing different inputs. The output of any model (linear or nonlinear) is affected significantly by how much information the input contains. In this context, the present model is based on univariate inputs to forecast the short term-direction of crude oil prices aiming to cover a gap in non linear forecasting of crude oil.

III. THE DATA AND METHODOLOGY

We implemented the methodology to the data series with daily and weekly crude oil prices starting from 1986 until 2018. These data sets were retrieved from US Department of energy: Energy Information Administration: <http://www.eia.doe.gov/>. The data was divided chronologically into training and testing sets, 90% of the data was used for training and 10% for out-of-sample for testing. Further, an early stopping was used to control the training process, the training data was divided into 60% training and 20% for testing and 20% for validation.

Box and Jenkins (1970) developed a coherent, versatile four stage iterative cycle for time series identification, estimation, checking and forecasting, rightly known as the Box-Jenkins methodology (De Gooijer, and Hyndman, 2006). The Box-Jenkins methodology deals with the non-stationary models such as autoregressive integrated moving average (ARIMA) models. In an ARIMA model, the future value of a variable is assumed to be a linear function of past observations and random errors. ARIMA models are, in theory, the most general class of models for forecasting a time series which can be stationarized by transformations such as differencing.

Let $\{Z_t\}$ be a non-stationary time series with stabilized variance. Where $\nabla=1-B$, choose d such that the mean of W_t is stabilized, then $\{W_t\}$ will be a stationary time series. In practice, the differencing operator d is usually 0, 1 or at most 2. The ARMA model for stationary series $\{W_t\}$ is given by $\phi(B)\tilde{w}_t = \theta(B)a_t$ or this can be written as

$g(B)\check{z}_t = \theta(B)$ where $g(B)=\phi(B)\nabla^d$. Since $\nabla^d \check{z} = \nabla^d Z_t$, this model can be written as

$$g(B)Z_t = \theta(B)a_t \text{ -----(1)}$$

The model in equation (1) is a non-stationary model in Z_t and is known as autoregressive integrated moving average model of order (p, d, q) i.e. ARIMA (p, d, q). The general ARIMA process may be generated by summing or "integrating" the stationary ARMA process W_t , d times. In

this model $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ is a polynomial in B of order p and is known as AR operator. $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ is a polynomial in B of order q and is known as MA operator. $g(B) = \phi(B)\nabla^d$ is a polynomial in B of order in (p+d) and is known as non-stationary operator (Chatfield, 1991). Model parameters are identified based on the autocorrelation and partial auto correlation function. Once a tentative model is specified, the parameters are estimated using nonlinear optimization procedures. Diagnostic checking of model accuracy is basically to check if the model assumptions about the errors are satisfied. If the model is not adequate, the above three steps are repeated until an adequate model is obtained. Diagnostic information helps in suggesting alternative models. The selection of the model is based on the principle of parsimony. The final selected model can be used for forecasting purposes. The model performance measured using different error measures such as mean absolute percentage error (MAPE) and Index of agreement (Willmott, 1981). In this paper, ARIMA modeling is implemented via SPSS and MS Excel is used for computations and charts.

Building ARIMA Model

Building ARIMA Model for Daily crude oil prices

In this Section, we discuss the modeling of daily crude oil prices in global market using, Box-Jenkins methodology. The data considered is daily crude oil prices from 2nd January, 1986 to 14th May, 2018 consisting of 8162 observations in which 8000 daily observations are used for modeling, and 162 daily observations are used for forecasting. The development of ARIMA model for any variable involves mainly four steps: Identification, Estimation, Diagnostic checking, and Forecasting.

Model identification

Time plot of the daily crude oil prices (Figure 1) reveals that the data is non seasonal and non-stationary.

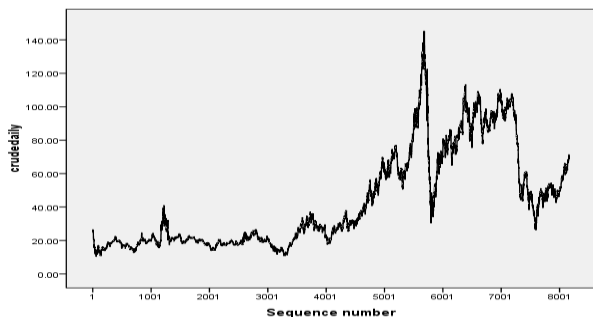


FIGURE 1: Time plot of daily crude oil prices

From the Time plot Figure 1, one can observe that the given crude oil price is non stationary, and an autoregressive integrated moving average (ARIMA) model can be fitting

well for the given data. Non stationarity in variance is corrected through natural logarithm transformation and non stationarity in mean is corrected through appropriate differencing of the data. In this case, non seasonal difference of order 1 (i.e. d=1) is sufficient to achieve stationarity in mean and variance.

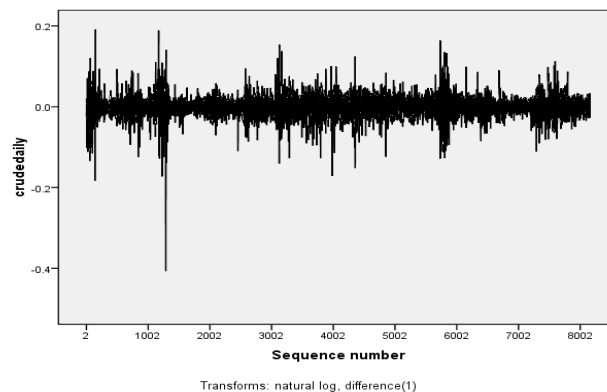


FIGURE 2: Time plot of log transformed the daily crude oil price with non-seasonal difference 1.

The graph (Figure 2) of W_t is stationary in mean and variance. The next step is to identify the values of p, q. Autocorrelations and partial autocorrelations for 25 lags of W_t are computed for the identification of the parameters of ARIMA model. All ARIMA models are considered for different possible values of (p, d, q) and the adequate models are selected and listed in Table 1. For these adequate models BIC, MAPE, RMSE, MAE are computed and are given in Table 2. Among these three models ARIMA (2, 1, 1) Model is having lowest values of these measures and hence, it is the suitable model for forecasting daily crude oil price.

Table 1: Model Description

Model ID	Model Description		
	Crude-Daily Price	Model_1	Model Type
			ARIMA(2,1,1)

Table 2: Adequate ARIMA Models for Forecasting daily Crude Oil Price.

Box –Jenkins models	BIC
ARIMA(1,1,2)	0.273
ARIMA(2,1,1)	0.270
ARIMA(2,1,2)	0.274

From the Table 2, it is observed that only ARIMA (2, 1, 1) is significant with respect to parameters as well as adequacy of the model. So the most suitable model is ARIMA (2, 1, 1) as this model has the lowest BIC values in the significant models. The estimated parameters of the selected model are presented in the following table. The fitted model for the forecasting crude oil prices is $Z_t = 0.704Z_{t-1} - 0.026Z_{t-2} + (1 + 0.722B)a_t$. Diagnostic checking is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders.

Table 3: Forecasts of daily crude oil price

Model		Forecasts									
		8163	8164	8165	8166	8167	8168	8169	8170	8171	8172
Crude daily price-Model_1	Forecast	71.01	71.02	71.03	71.04	71.06	71.08	71.10	71.13	71.15	71.18
	UCL	74.57	76.04	77.11	77.99	78.76	79.46	80.10	80.70	81.27	81.81
	LCL	67.59	66.25	65.31	64.56	63.93	63.38	62.88	62.42	62.00	61.60

Building ARIMA Model for Weekly crude oil prices

Similar to the previous section, here, we discuss the modeling of weekly crude oil price using Box-Jenkins methodology. The data is weekly crude oil price from 3rd January, 1986 to May, 2018 consisting of 1692 observations. As stated before, the development of ARIMA model for any variable involves mainly four steps: Identification, Estimation, Diagnostic checking and Forecasting which are implemented for the data on weekly oil prices.

Model identification

Time plot of the weekly crude oil price (Figure 3) reveals that the data is non seasonal and non-stationary



FIGURE 3: Time plot of the weekly crude oil price.

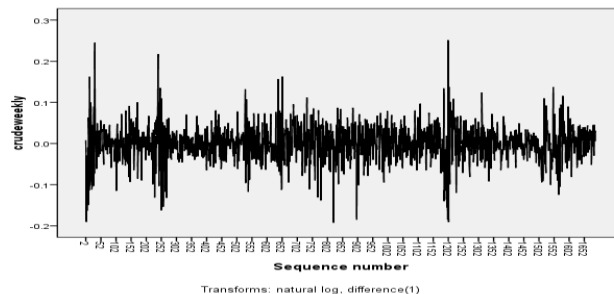


FIGURE 4: The time plot of the transformed series.

From Figure 4, it is observed that W_t is stationary in average and variance. The next step is to identify the values of p and q. The following tentative ARIMA models are fitted and chosen a model, which has minimum AIC and BIC values. The Ljung-Box adequacy tests and their significant probability are considered for identifying a suitable model for the given data.

Time Series Modeler

Table 4: Time Series Model Selection

Model Description			
Model ID	Crude-Weekly	Model_1	Model Type
			ARIMA(1,1,1)

Model parameters (without constant term in the model) are estimated using SPSS for selected model. Results of estimation of parameters are given below.

Table 5: Model Selection

Box -Jenkins models	BIC
ARIMA(1,1,1)	1.348
ARIMA(1,1,2)	1.356
ARIMA(0,1,1)	1.351

From the Table 5, it is observed that only ARIMA (1,1,1) is significant with respect to parameters as well as adequacy of the model. So the most suitable model is ARIMA (1,1,1) as this model has the lowest BIC values in the significant models. The estimated parameters of the selected model are presented in the following table.

The fitted model for the forecasting crude oil prices is

$$Z_t = -0.379Z_{t-1} + (1 - 0.507B)a_t + .001$$

Diagnostic checking is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders.

The selected ARIMA model is used to forecast the one-step ahead future weekly oil prices and the forecasts are presented in the following table:

Table 6: Forecasts

Model		Forecasts										
		1690	1691	1692	1693	1694	1695	1696	1697	1698	1699	1700
Crude weekly price-Model_1	Forecast	70.94	70.96	71.12	71.22	71.35	71.47	71.59	71.71	71.83	71.96	72.08
	UCL	77.09	80.39	82.86	85.01	86.96	88.75	90.43	92.03	93.56	95.03	96.46
	LCL	65.15	62.38	60.65	59.16	57.92	56.81	55.82	54.92	54.09	53.31	52.59

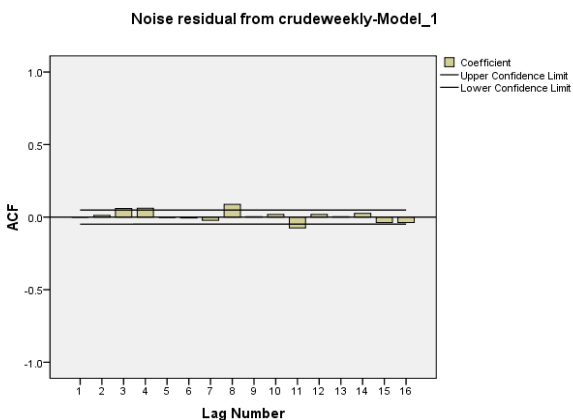


Figure 5: ACF plot of Transformed series

Neural Networks Model

Neural Networks comprise one of the five main computational intelligence paradigms, and are known as universal approximates. The first (and most popular) network is the MLP network; the second is the RNN.

Building Feed forward and Recurring Neural Networks Models

The MLP network is a member of the feed forward network architecture, and is the simplest of the networks under investigation. In this network, there are 3 layers, each composed of neurons. The 3 layers are the input, hidden, and output layers. In this Section, building of a feed forward and recurring neural networks model for forecasting of daily and weekly crude oil price is discussed. SPSS package is used to train and test the feed forward neural networks.

Structure of the network:

The model is a three-layer feed forward neural network and it consists of an input layer, a hidden layer and an output layer. In this model only one output unit is needed and it indicates the forecasts of daily and weekly crude oil prices. There is no easy way to determine the optimum number of hidden units without training and testing. The best approach to find the optimal number of hidden units is trial and error. In practice, one can use either the forward selection or backward selection to determine the hidden layer units. Forward selection method is applied, in which, a small number of hidden neurons are selected then the network performance is recorded by computing the MAE, MAPE and RMSE. Next the hidden neurons are increased one by one, trained and tested until the error is acceptably small or no significant improvement is noted.

A back propagation algorithm is used for the network training with the following parameters.

Building BPNN Model

Architecture: MLP

Training Algorithm: BP algorithm (gradient-descent rule)

Learning rate: 0.6

Momentum rate: 0.4

Number of epochs: 1700

Number of observations: 8162(daily data)

Activation/ transfer function: Sigmoid

Building BPNN Model

Architecture: MLP

Training Algorithm: BP algorithm (gradient-descent rule)

Learning rate: 0.6

Momentum rate: 0.5

Number of epochs: 1800

Number of observations: 1689(weekly data)

Activation/ transfer function: sigmoid

IV. RESULTS AND DISCUSSION

Results for Forecasting Daily Crude oil price

The two neural network models BPN (static model) and RNN (dynamic model) are used to forecast the daily crude oil price and the results obtained of the neural network models are compared with the results of traditional econometric model. The forecasting accuracy of these three models is tested by using RMSE, MAE, and MAPE values.

In BPN and RNN models, 80% of the data is used as training set to estimate the models and the remaining 20% of the data is used as testing set for assessing the forecast performance based on out-of-sample forecasts. The results obtained by

neural network and traditional models are shown in Table 9 as follows:

Table.7: The prediction result of daily data using BPNN and ARIMA

MODEL	Observations	RMSE	MAE	MAPE
ARIMA(2,1,1)	8162	1.142	0.707	1.730
MLP (FFNN) (with one hidden layer)	8162	1.004	0.302	1.426
RNN (FBNN) (with one hidden layer)	8162	0.627	0.103	1.002

Results for Forecasting Weekly Crude oil price

The two neural network models BPN (static model) and RNN (dynamic model) are used to forecast the weekly crude oil price and the results obtained are compared with the results of traditional econometric model. The forecasting accuracy of these three models is tested by using RMSE, MAE, and MAPE values.

Table.8: The prediction result of weekly data using BPNN and ARIMA

MODEL	Observations	RMSE	MAE	MAPE
ARIMA(1,1,1)	1689	1.949	1.264	3.114
MLP (FFNN) (with one hidden layer)	1689	1.306	1.082	2.648
RNN (FBNN) (with one hidden layer)	1689	0.746	0.831	1.473

V. CONCLUSION

The above Empirical evaluation shows that RNN model is performing better than BPN and traditional models to forecast daily and weekly Crude oil price globally. Also, the performance of BPN is better than the traditional econometric model. It shows that neural network models are very successful in predicting nonlinear relationships. For the application studied in this paper, it is concluded that the RNN model (dynamic model) is significantly more accurate than the BPN model (static model) and the traditional model. Since the RNN has the internal feedback which allows it to capture the dynamic structure of the data, it has been able to outperform the BPN model and the traditional model. As most of the economic series are dynamic, therefore the RNN model can be a good alternative to the time series and economic series data (Rani, SA Jyothi, 2017). Though, the above procedures show good performance, one can try the data sets considered here by using SVM and GARCH models for a better model fit for the future study. The results obtained in this study show that while all neural networks investigated, have the potential for crude oil price prediction daily and weekly, the best results are generally obtained from

the recurrent neural architectures, especially on data outside the training range.

REFERENCES

- [1] Bashar Al-hnaity and Maysam Abbod (2016), Predicting Financial Time Series Data Using Hybrid Model, © Springer International Publishing Switzerland.
- [2] Box, G.E.O. and G.M. Jenkins (1976), Time Series Analysis, Revised Edition, Holden-Day San Francisco.
- [3] Box, G. E. P. and Jenkins, G. M. (1970), Time Series Analysis Forecasting and Control, Pearson Education.
- [4] Chatfield, C., 1991. The Analysis of Time Series. Chapman and Hall, New York.
- [5] De Gooijer, Jan G., and Rob J. Hyndman. "25 years of time series forecasting." *International journal of forecasting* 22.3 (2006): 443-473.
- [6] Fatimah Mohd. (1986), Crude Palm Oil Price Forecasting: Box-Jenkins Approach, Pertanika, Vol. 9 No.3.
- [7] Haidar, Imad, Siddhivinayak Kulkarni, and Heping Pan. "Forecasting model for crude oil prices based on artificial neural networks." *Intelligent Sensors, Sensor Networks and Information Processing, 2008. ISSNIP 2008. International Conference on*. IEEE, 2008.
- [8] Hamdi, Manel, and Chaker Aloui. "Forecasting crude oil price using artificial neural networks: a literature survey." *Econ Bull* 3.2 (2015): 1339-1359.
- [9] Koutroumanidis, Theodoros, Konstantinos Ioannou, and Garyfallos Arabatzis. "Predicting fuelwood prices in Greece with the use of ARIMA models, artificial neural networks and a hybrid ARIMA-ANN model." *Energy Policy* 37.9 (2009): 3627-3634.
- [10] Maditinos, Dimitrios, and Prodromos Chatzoglou. "The use of neural networks in forecasting." *Review of Economic Sciences* 6 (2004): 161-176.
- [11] Mohamed A. Hmilad (2015), Regression analysis to forecast Malaysia's imports of crude material, International Journal of Management and Applied Science, ISSN: 23947926, Volume-1, Issue-4.
- [12] Oancea, B., & Ciucu, S. C. (2014). Time series forecasting using neural networks. *arXiv preprint arXiv:1401.1333*.
- [13] Rana Abdullah Ahmed (2014), **Daily** crude oil price forecasting model using ARIMA, Generalised autoregressive Conditional Heteroscedastic and support vector Machines, American Journal of Applied Sciences, 11 (3): 425-432
- [14] Rani, SA Jyothi, V. V. Haragopal, and M. Krishna Reddy. "Forecasting Inflation Rate of India using Neural Networks." *International Journal of Computer Applications* 158.5 (2017).
- [15] Rani, SA Jyothi, and V. V. Haragopal. "Forecasting Exchange Rates using Neural Networks."
- [16] S. Moshiri and F. Foroutan. "Forecasting nonlinear crude oil futures prices " *The Energy Journal*, vol. 27, pp. 81-95, 2005.
- [17] Yadav, S., & Meshram, S. (2013). Brain tumor detection using clustering method. *International Journal of Computational Engineering Research*, 3(4), 11-14.