PREDICTION OF CRUDE OIL PRICES USING FEED FORWARD BACK PROPAGATION

Sourabh Mahajan*

Student, Master of Business Administration (Finance Management), Department of Management Studies,

Christ (Deemed to be University)

ABSTRACT

"Crude oil" is consider to be as Black gold. Crude oil is produce by small numbers of companies in very few countries. There is often mismatch between the point of consumption and oil producing countries. Generally, Organization of Petroleum exporting countries (OPEC) and Middle East countries hold nearly 50% of the world oil's reservoirs. It leads to trading in oil globally which results in healthy and successful trading between countries, private traders and suppliers and investors. This research paper explored a prediction model for crude oil price using Artificial Neural Network (ANN) with feed forward back propagation. Crude oil is one of the commodity which has large number of future contracts been traded. Oil prices are generally affected by US dollar index, S&P Index, Gold and Silver prices. The three performance measures, the coefficient of determination (R2), Mean Square Error (MSE), and Sum Squared Error (SSE) are implemented to evaluate the performances of model developed. The results shows that predicted values are nearly accurate with target value. With the proper prediction of crude oil price investors and participants like exporters, marketers, stakeholders can apply risk management technique which will improve efficiencies and competitiveness of price mechanism.

Keyword – Crude oil price, ANN, Prediction, Feed forward back propagation

Introduction

Crude oil plays a very significant role in global economic system. It is also very important global development and growth for industrialized. Crude oil is one of the highly traded commodity and has large volume of future contracts. Geopolitical factors, weather conditions, changes in financial markets are major factors of the crude oil market which creates volatility in price of oil market. Crude oil prices as wells as other assets class are sometimes closely correlated. A number commodities and their prices are interrelated. The link between Gold and crude oil prices are usually positive correlated where any of the related factor can move the price in same direction. On the other hand, the equity stock and crude oil prices are often negative correlated. For example the price of crude oil decreases the oil company stocks move up (valid only for related industry).These effect of changes in the oil prices creates direct impact on business, traders, and investors and on the economy as whole. The price fluctuation can have both positive and negative impact on individual perceptive basis. To eliminate and mitigate risk and to prevent one from bankruptcy, it is necessary to predict the movement of oil prices which can be predicted with artificial neural network model using machine learning algorithm.

ANN model is used for better accuracy and management of crude oil investment calls. Commodity market requires a large amount to be invested, so an accurate prediction model is required to predict the change in price trend so that risk factor can be mitigate in the future. Moreover, it is very difficult to generate ANN algorithm using traditional mathematical models and these models are based on certain assumptions and prior understanding of inputs statistical parameters. The main aim of this paper is to analyze the capability of ANN model to predict the crude oil price with high degree of accuracy. The model is evaluated according to three performance measures, which includes the coefficient of determination (R2), Mean Square Error (MSE), and Sum Squared Error (SSE). The remaining part of the paper is organized as follows- Section (ii) summaries the Literature Review, Section (iii) provides a brief on Artificial Neural network, Section (iv) presents about the analyses of prediction result, Section (v) shows the brief conclusion and future scope.

Literature Review

There has been a large number of studies conducted to predict the crude oil prices. In Sehgal, K. Pandey [2001], Artificial intelligence methods for oil price forecasting: a review and evaluation with an objective to use various methods for oil price forecasting using Artificial Intelligence contributed an evidence indicating demand from China and India as a leading driver in the world oil price system.

Sompui, Wongsinlatam [2014], Prediction Model for Crude Oil Price Using Artificial Neural Networks, contributed that prediction of crude oil prices using ANN is better compared to least Square Method.

Mombeini, Yazdani-Chamzini [2015], Modeling Gold Price via Artificial Neural Network, and Multi-layer feed-forward perceptron (MLP) with a back propagation learning algorithm is employed to model gold price with 200 training data sets and 20 points for validation. It is measured using three performance parameters coefficient of determination (R^2), root mean square error (RMSE), and Mean absolute error (MAE).

Mahdiani, Khamehchi [2016], a modified neural network model for predicting the crude oil price, the paper gives insight about prediction of the crude oil price based on the historical similarities of points. Genetic algorithm is been used for prediction, which greatly improves neural network estimation.

Chittedi [2012], Do Oil Prices Matters for Indian Stock Markets? An Empirical Analysis, provides the relationship between oil prices and stock prices in India, and the purpose of the paper is to understand the level of susceptibility of stock prices in emerging economies to movement in global oil prices and there is high significant impact of change in stock prices to oil prices.

All the above mentioned studies indicated that prediction through artificial neural network will provide a better accuracy than the traditional mathematical models.

Artificial Neural Network

Artificial Neural Network (ANN) has a long history since decades. ANN mechanism is same as biological neural network. ANN provides a model to predict the targeted or desired value for that variable. There are two types of ANN model are available with single neural network layer and multiple neural network layer. Single neural network has single input and single output layer, whereas multiple neural network has more than one input layer and output. In between input and output layer there are hidden layer present. Hidden layer is connected to each input layer and it is present in between. The certain weight is assigned to each layer connected between input and hidden layers. The tan sigmoid activation function is used. Activation function is used to normalize the value in the range from -1 to 1.

Nature of data and Methodology

The study revolves around the investigation of the long run relationship between data of US Dollar Index, S&P 500 Index, Gold Price, Silver price and Crude Oil future prices. The targeted crude oil price data has been collected from Petroleum Planning and Analysis Cell, Gold prices is collected from gold.org, Silver price data is collected from kitcosilver.com, US Dollar Index and S&P 500 has been collected from investing.com.

There are several methods available to predict the crude oil prices. In this study we adopt ANN as a mapping model to predict the crude oil price. ANN has appeared as a powerful tool in recent times with wide range of fields such as pattern recognition, forecasting, data mining and analyze. Here, ANN is viewed as non-parametric, non-linear, without any assumptions. ANN does not make any prior assumption about problem statement, data learns by itself. Further, Feedforward backpropagation network with nonlinear function is employed and the studies conducted prior related in same field shows the higher degree of success with same network.

Results and Discussion

The output of the feed forward back propagation neural network can be explained using parameters like mean square error, coefficient of Regression (R^2) and sum of squared error. Finally the predicted value is compared with the targeted value to check the error. The model will divide data in such a manner that 70 percent of the data is used for training purpose, 15 percent for testing and rest is for validation purpose. Fig 1 shows the prediction model with feed forward back propagation. This model has 4 input layer as discussed earlier. Each input layer is connected to hidden layer using appropriate weight and bias value with tan sigmoid as activation function. Activation function is used to normalize the value in the range from -1 to 1. The output is generated using simulation model in MATLAB R2017b. If the predicted output is not in line with targeted value then weight that is assign between input and hidden layer need to change accordingly and repeat till we get the in line predicted value.

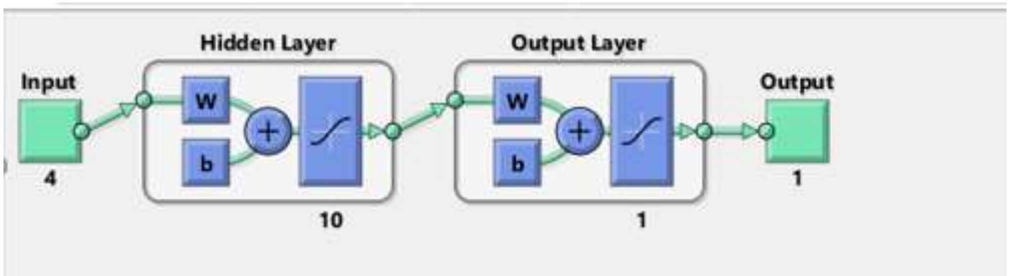

Fig 1- Prediction model (Author's work)

Theoretically, performance measure were calculated Mean square error, coefficient of determination (R^2) is calculated using the formula given by equation (i) and (ii) respectively.

$$MSE = \sum_{i=1}^{N} \frac{(Ti - Pi)^2}{N} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (i)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(Ti-Pi)^2}{\sum_{i=1}^{N}(Ti-\mu)^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (ii)$$

where, T_i= Targeted value,

P_i= Predicted value,

N= Number of datasets

$\mu = average\ of\ observation\ data$

Mean square error of predicted model should be low as possible and it provides the difference between the target and predicted value.

Coefficient of determination (R^2) explains how much the variability in dependent variable can be explain by all the independent variable combine together. Generally, R^2 ranges from one to zero. One or close to it represent a good fit for prediction model and zero or close to it represent a poor fit. Performance measures are explained below with graphs.

(i) MSE (Mean Squared Error)

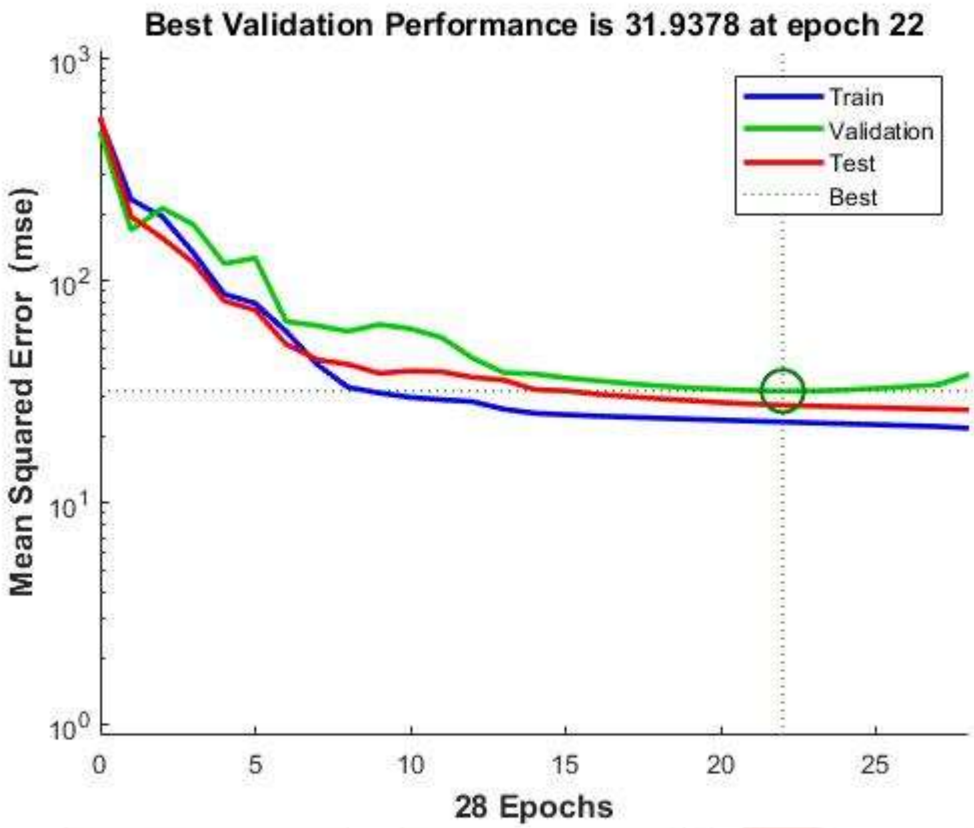

Fig 2- Mean Square Error

The above graph is performance for Mean square error. Mean square error is the difference between the expected value and the expector value. The value of the mean square error should be low. The ideal value should be zero, but it is practically not possible. Here, in the prediction of crude oil prices mean square occurs at 28 epochs. The network has four input layer with weight assign to each input layer.

First the complete data set is divided into training, testing set and validation set. Since the neural network is self -learning tool, for training the data Levenberg Marquardt backpropagation is used so that it continuously optimize the weight and bias value. Blue line depicts the train data. Further, validation line is in green line. Dotted line shows the best validation performance is seen at 31.93 at 22 epochs.

(ii) Coefficient of Determination(R^2)

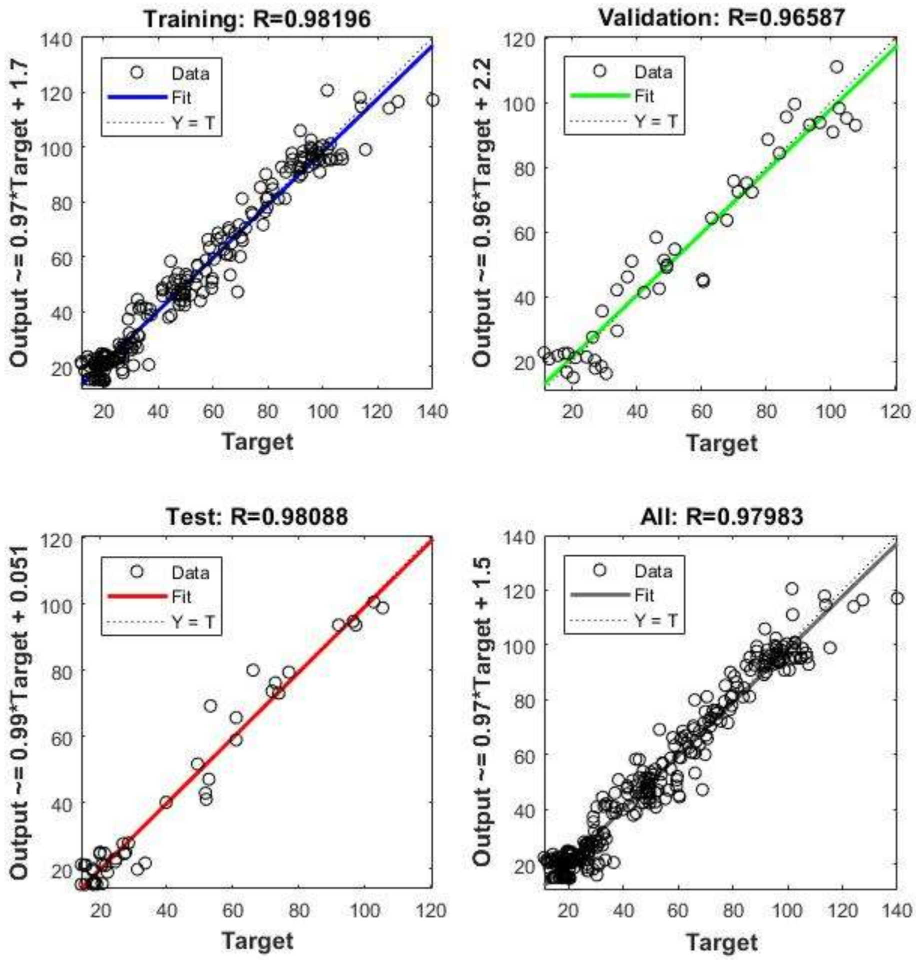

Fig 3- Coefficient of Determination (R^2)

Coefficient of Determination (R^2) can be best explained by the best fit of the line. In all the four graphs it is clearly seen that while training the data sets nearly 0.9819(98.19%) of the training data lies on the line. Also, for testing and validation most of the data lies close to one. R-Squared clearly explains that out of total 302 data points 97.98% of the change in dependent variable of prediction model can be explained by combining all the independent variable of the model.

(iii) Sum of Squared Error

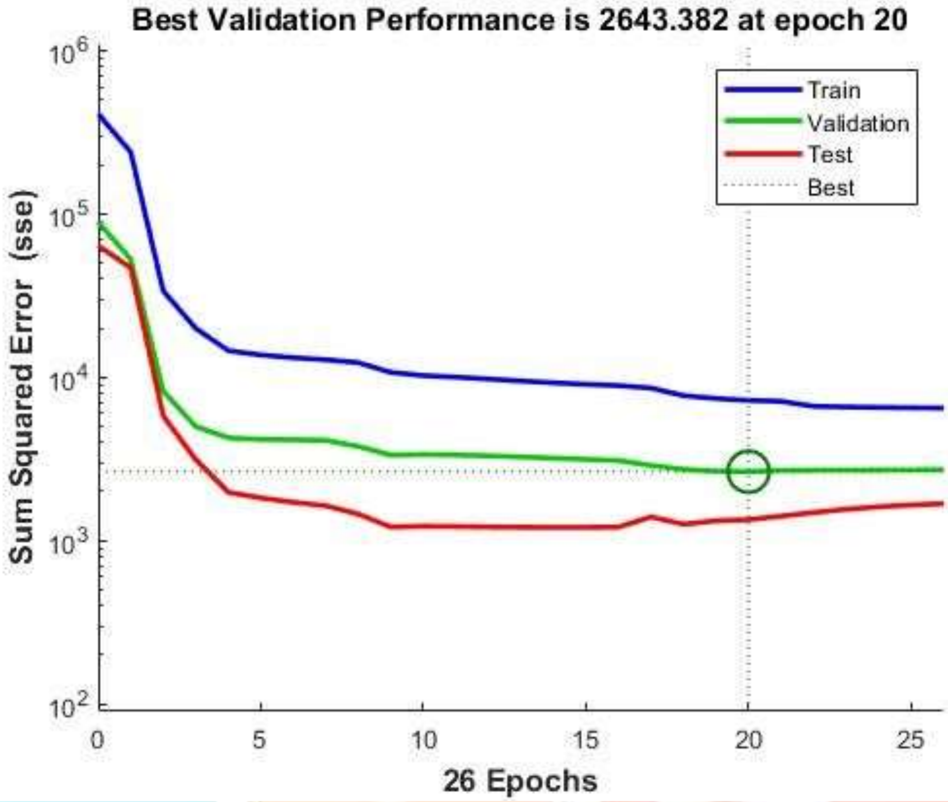

Fig 4- Sum of squared Error

Sum of squared error is another term for Residual sum of squares. It is sum of all the error value originating from the difference between target value and predication value. A small value for this indicate the best fit for the model. From the above graph it is indicates that training, testing and validation have the same curve. Validation data set point curve is between training and testing curve. The sum squared error performance is best measured at 20 epochs and value is found to be 2643.382.

The data was selected from time horizon of 1993 to 2017 (302 data points) on monthly basis was analyses for input variable and target variable. The table 1 shows the overview of few sample data output point. Predicted value is calculated using the developed feed forward back propagation model using MATLAB R2017b which gives the output in terms of absolute error which is difference between target and predicted value.

The table 1 provides the absolute error for few data points. Both the values are move in tandem with small residual value.

Targeted Value Crude Oil Price($)	Predicted value Crude Oil Price($)	Error
65.07	**63.0711**	-1.9989
64.73	**64.9776**	0.2476
60.42	**58.1312**	-2.2888
57.4	**56.5197**	-0.8803
54.38	**53.9143**	-0.4657
51.67	**54.491**	2.821
47.23	**53.5828**	6.3528
50.17	**54.6249**	4.4549
46.04	**51.5053**	5.4653
48.32	**48.9654**	0.6454
49.33	**47.267**	-2.063
50.6	**50.1642**	-0.4358
54.01	**50.1475**	-3.8625
52.81	**50.3798**	-2.4302
53.72	**52.963**	-0.757
49.44	**49.4992**	0.0592
46.86	**43.8593**	-3.0007
48.24	**48.7226**	0.4826
44.7	**47.347**	2.647

Table -1 Target vs Predicted Value

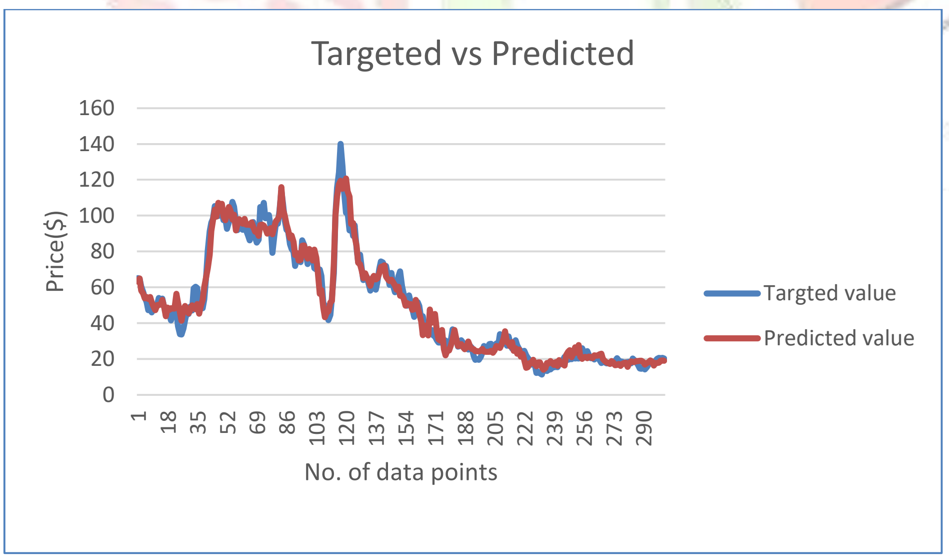

Fig 5 – Targeted value vs Predicted Value

The figure 5 shows the line graph between the targeted and predicted value. Both the line follows the same trend with small amount of residual value between them.

Conclusion and Future Scope

In this study a prediction model for crude oil prices using artificial neural network was developed using machine learning tool MATLAB R2017b. The data for crude oil price as target variable was selected from 1993 to 2018 on monthly basis. The four input variable was selected and each connected to their corresponding variable. The data was divided into training, testing and validation sets. The output was measured using performance measures such as Mean square error (MSE), Coefficient of determination (R^2), sum of squared error (SSE). The output shows that with ANN model the predicted values are nearly accurate with targeted value.

Future implication for the model could be the forecasting model can be developed which can forecast the value in the near term which can be beneficial for investor and marketers.

References

1. Sompui, Mayuree, and Wullapa Wongsinlatam. "Prediction Model for Crude Oil Price Using Artificial Neural Networks." *Applied Mathematical Sciences*, vol. 8, 2014, pp. 3953–3965., doi:10.12988/ams.2014.43193
2. Mombeini, Hossein, and Abdolreza Yazdani-Chamzini. "Modeling Gold Price via Artificial Neural Network." *Journal of Economics, Business and Management*, vol. 3, no. 7, 2015, pp. 699–703., doi:10.7763/joebm.2015.v3.269.
3. Sehgal, Neha, and Krishan K. Pandey. "Artificial Intelligence Methods for Oil Price Forecasting: a Review and Evaluation." *Energy Systems*, vol. 6, no. 4, 2015, pp. 479–506., doi:10.1007/s12667-015-0151-y.
4. Live Gold Price. (n.d.). Retrieved February 1, 2018, from http://www.kitco.com/charts/livegold.html
5. Silver News, Price Charts & Quotes | KITCO Silver. (n.d.). Retrieved February 1, 2018, from http://www.kitcosilver.com/