International Journal of Conceptions on Computing and Information Technology
Vol. 5, Issue. 1, July' 2017; ISSN: 2345 - 9808

Comparative study on demand forecasting by using Autoregressive Integrated Moving Average (ARIMA) and Response Surface Methodology (RSM)

Nummon Chimkeaw, Yonghee Lee, Hyunjeong Lee and Sangmun Shin
Department of Industrial and Management System Engineering
Dong-A University
Busan, Republic of Korea
mon943055@gmail.com, yonghee@dau.ac.kr, gpfapt1223@gmail.com, sshin@dau.ac.kr

Abstract— **Forecasting of crude oil price has been received attention from both public and private sectors. Oil price is a significant factor that can affect the economy of the country and its changes can affect the planning of government as well as commercial sectors. The prediction of oil prices can assist the planners to organize their planning. In spite of this, a number of methods were proposed but they are complex to understand for new practitioner whoever interested in the forecasting field. The main objective of this paper is to apply response surface methodology (RSM) to forecast crude oil price more effectively and conveniently. Monthly crude oil prices of West Texas Intermediate (WTI) are forecasted by using the proposed methodology. In this study, data from January 2013 to December 2015 are used as a training data set. Afterward, forecast the monthly oil price from July 2015 to August 2016. To attain the aim of the proposed methodology, this paper is organized in the following steps. Firstly, Autoregressive integrated moving average (ARIMA) is interpreted as a conventional method. Secondly, RSM is proposed as an alternative crude oil price forecasting methodology. Thirdly, a case study is conducted based on the WTI oil price data series to demonstrate the performance of the proposed methodology. Finally, the result of the proposed methodology is compared with single ARIMA model and combined model by using two statistical criteria i.e. mean absolute percentage square error (MAPE) and direction statistics (D-stat).**

Keywords- combine forecasting, crude oil price, ARIMA, RSM, variance-covariance

I. INTRODUCTION

Crude oil price has been played a significant role in world economy because of its huge worldwide production, demand, and consumption. It has consequential effects in economic circumstances of numerous countries. Various researchers and intellectual personalities have been put forward their interests in modeling and forecasting crude oil prices. High value of oil prices is often affect the oil importing countries with inflation whereas low prices, as result causes economic recession in exporting countries. So, it is very crucial to forecast the oil price to maintain the overall GDP of the country. Although, the dramatic raised up and dropped down of prices in several intervals has lead the forecasting practitioners often in trouble to envision the crude oil prices. Different researchers have proposed various methodologies to predict the crude oil prices in an optimum way. Historical data are very important to foresee upcoming prices of crude oil.

Several models have been developed and implemented for the forecasting of crude oil prices. Mainly the crude oil price forecasting literature can be categorized into two groups- quantitative methods and qualitative methods time series models, artificial models, support vector machines are quantitative methods whereas, Delphi method, belief networks, fuzzy logic and expert systems, web text mining methods are qualitative methods (Niaz & Jose) [1]. The quantitative methods have been extensively used and famous for time series data. The numerical techniques in these models also helped to estimate and predicts the data more accurately. The qualitative methods can be determined by using qualitative factors such as political, geological, natural, and other environmental factors. Although the quantitative factors are more important and are more decisive to determine the world oil prices. The crude oil is basically determined by its supply and demand, but more strongly influenced by many irregular past/present/future events like weather, stock levels, GDP growth, political aspects and so on. (Xie) [2] has employed a support vector machine (SVM) to forecast the oil price with production, inventory and consumption used as factors affecting the prices. The complicated data set can be decomposed into finite small 'intrinsic mode functions' by using an 'empirical mode decomposition' (EMD) based on neural network for world oil price forecasting (Yu) [3]. The decomposition of the data into smaller units could be beneficial for more accurate forecasting. The methodology developed in this paper is based on the crude oil price of the West Texas Intermediate (WTI). Several methodologies are

developed and applied to find the variation of crude oil price of the WTI. Soft computing approach is introduced with novel algorithm to forecast WTI crude oil spot price variations in a specified future duration (Ghaffari & Zare) [4]. (Michael & Ye) [5], has developed a model for forecasting monthly crude oil spot prices under the normal market circumstances using readily available data with only one independent variable. This paper presents a short-term forecasting by using 36-month data with the 4 factors namely- world crude oil production, Organization for Economic Co-operation, and Development (OECD) crude oil inventory, U.S crude oil inventory and worldwide demand affecting the price.

A number of literature were mentioned in the paper to forecast crude oil prices, however, there is ample room for improvements. Forecasting is a complex issue and no method is purely correct. So, a very simple and effective method is proposed in this paper to predict the WTI crude oil price more precisely with four factors. So, a well-known method called response surface methodology (RSM) has proposed to investigate the interdependencies of these variable and the scenarios of price variation. Initially, the regression analysis helps to investigating the effect of the 4-independent variable (input variables) to the one dependent variable (output/response variable). Using regression analysis along with RSM for building a second order model for response variable. Secondly, an advanced estimation method for oil price forecasting by using RSM is then proposed, which is powerful statistical tool for modeling and analyzing the situations where the crude oil prices are affected by several factors, such as world crude oil production, OECD inventory, U.S inventory and worldwide demand. Finally, a case study for WTI has conducted and tested in order to find the efficiency of the proposed model.

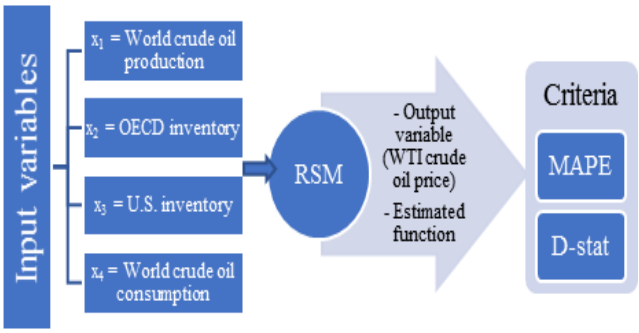

Figure 1. The proposed crude oil price forecasting procedure

II. Conventional Method

Autoregressive Integrated Moving Average (ARIMA) is one of the most popular time series forecasting method. It is the generalization of AR (Auto regressive) and MA (Moving Average). (Box & Jenkins)[6] had first proposed this model with some iterative procedure for modeling a time series. The iterative procedure consists of three steps- Identification, Estimation, and Diagnostic checking. Generally, non-seasonal ARIMA models are denoted by ARIMA (p, d, q) where p, d and q are the order of the autoregressive term, the degree of non-seasonal differences needed for stationarity, and the number of lagged forecast errors in the prediction equation respectively. From the above-mentioned procedure, the appropriate values and orders of p, d and q can be obtained by using identification step. The estimation procedure helps to estimate the model. And lastly, the diagnostic checking check the model to find out its efficiency.

In an autoregressive integrated moving average model, the future value of a variable is assumed to be a linear function of several past observations and random errors (Zhang)[7]. The general forecasting equation in terms of y is in Eq. (1):

$$\hat{y} = \mu + \emptyset_1 y_{t-1} + .. + \emptyset_p y_{t-p} - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q} \quad (1)$$

where y_t and ε_t are the actual value and random error at time period t, respectively; $\emptyset_i (i = 1,2, \ldots\ldots, p)$ and $\theta_j (j = 0,1,2, \ldots\ldots, q)$ are model parameters, p and q are integers and often referred to as orders of the model. Random errors, ε_t are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2.

III. Proposed Methodology

Response surface methodology, or RSM, is a collection of mathematical and statistical techniques useful for the modeling and analysis of problems in which a response of interest is influenced by several variables and the objective is to optimize the response. (Myers & Montgomery) [8]. In general, the relationship between the input variable and a response of interest (y) is unknown but can be approximated by low-degree polynomial model in the form

$$y = f'(x) + \varepsilon \quad (2)$$

where $x = (x_1, x_2, .., x_k)'$, f(x) is a vector function of p elements that consists of powers and cross-products of power, β is a vector of p unknown constant coefficients referred to as parameters, and ε is a random error (Behmiri & Manso)[9]. The generally used second-degree model is

$$\hat{y} = \beta_0 + \sum_{l=1}^{k} \beta_i x_i + \sum \sum_{i<j} \beta_{ij}\, x_i x_j + \sum_{l=1}^{k} \beta_i\, x_i^2 \quad (3)$$

where $\hat{y}$ is the response and β denote the coefficients of an estimated function. The goodness of fit was expressed by the correlation coefficient (R^2). In this study, the regression model of the WTI oil price has been constructed by using 4 independent variables and the results were analyzed by using analysis of variance (ANOVA) in the Minitab software. The proposed crude oil price forecasting procedure is illustrated in Figure 1.

IV. Case Study

A. Data

There are five empirical data that were used in this study are defined as follow:

- Response variable (y) is the monthly crude oil sport prices of West Texas Intermediate (WTI), the observations from January 2013 to December 2015, as illustrated in Figure 2.

- Input variables are - the world crude oil production (million barrels per month) used as x_1, the inventory of crude oil from countries in the Organization for Economic Co-operation and Development (OECD) used as x_2, the U.S. crude oil inventory used as x_3 and the average world crude oil consumption (million barrels per month) used as x_4. All data of the input variables were observed from January 2013 to December 2015, as illustrated in Figure. 3, Figure. 4, Figure. 5 and Figure. 6, respectively.

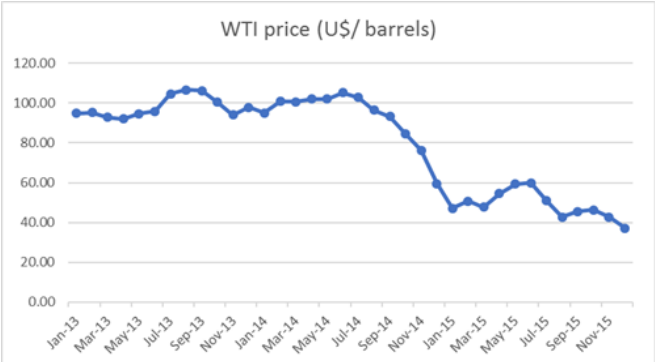

Figure 2. The monthly oil price for the period 2013 – 2015.

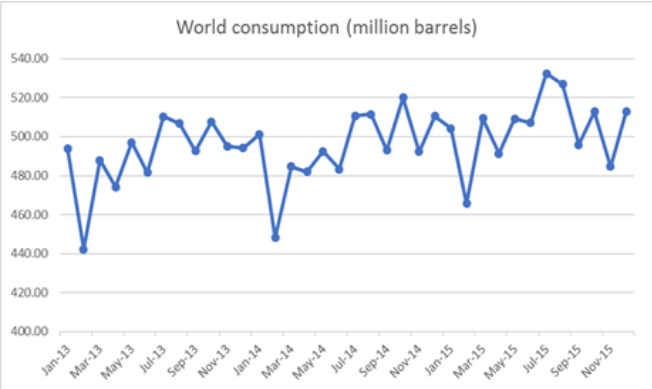

Figure 3. Crude oil world consumption

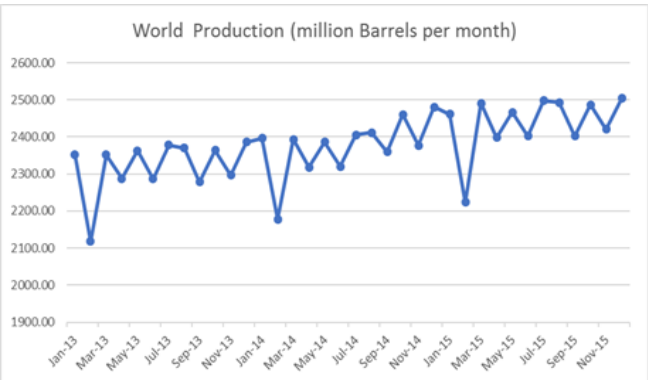

Figure 4. Crude oil world production

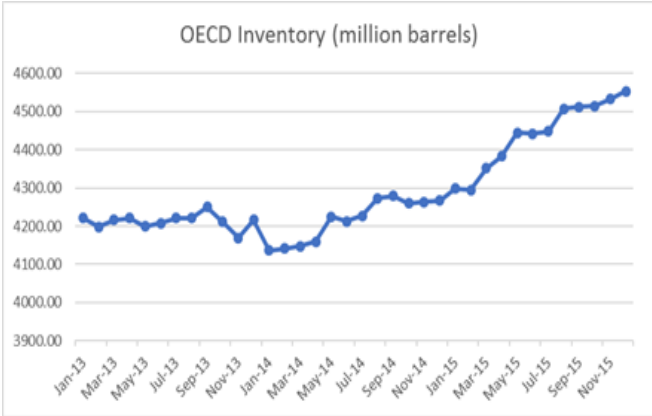

Figure 5. OECD crude oil inventory.

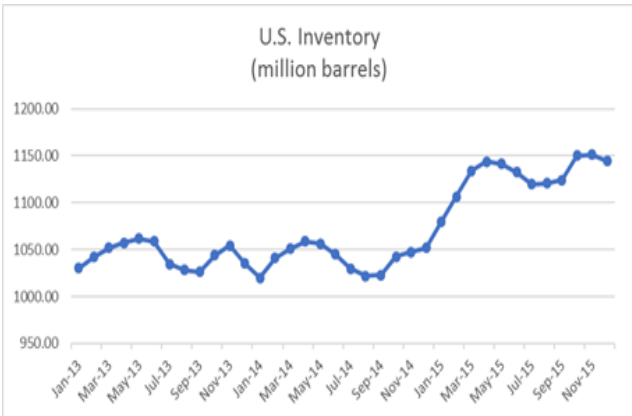

Figure 6. U.S. crude oil inventory.

B. ARIMA

ARIMA (1,1,0) is used in this paper. In order to utilize the ARIMA model, a first difference is necessary (Xie *et al.*) [2]. The prediction equation for ARIMA (1, 1, 0) is a first order autoregressive model with one order of non-seasonal difference and a constant term which can be expressed as Eq. (4):

$$\hat{Y}_t = \mu + Y_{t-1} + \emptyset_1(Y_{t-1} - Y_{t-2}) \quad (4)$$

C. Response Surface Methodology

The relationship between the response variable and four input variables were studied and development mathematical equations and calculated as the sum of a constant, four first-order effects (terms in x_1, x_2, x_3 and x_4) four second-order effects (terms in $x_1^2 x_2^2 x_3^2$ *and* x_4^2) and six interaction effects (terms in x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4 and x_3x_4) according to the Eq.(1). Using a MINITAB software, the analysis of variance (ANOVA) and estimated function were calculated and presented in Table 1 and Eq. (5), respectively.

TABLE 1: ANOVA TABLE OF ALL MODELS

Source	DF	Adj SS	Adj MS	F-value	p-value
Regression	14	29366.3	2097.59	40.50	0.000
x_1	1	288.3	288.32	5.57	0.025
x_2	1	0.4	0.35	0.01	0.935
x_3	1	476.6	476.57	9.20	0.005
x_4	1	123.0	123.01	2.38	0.134
x_1^2	1	244.9	244.90	4.73	0.038
x_2^2	1	27.1	27.05	0.52	0.476
x_3^2	1	43.2	43.19	0.83	0.369
x_4^2	1	7.6	7.57	0.15	0.705
x_1x_2	1	104.7	104.74	2.02	0.166
x_1x_3	1	321.7	321.72	6.21	0.019
x_1x_4	1	42.4	42.38	0.82	0.373
x_2x_3	1	68.0	67.98	1.31	0.261
x_2x_4	1	68.3	68.31	1.32	0.260
x_3x_4	1	102.6	102.61	1.98	0.170
Error	29	1501.9	51.79		
Total	43	30868.2			
Model summary					
	s	R-sq	R-sq (adj)	R-sq (pred)	
	7.19649	95.13%	92.79%	86.08%	

$$y = 3371 + 2.92x_1 + 0.1x_2 - 8.77x_3 - 7.38x_4 - 0.001419x_1^2 - 0.000289x_2^2 - 0.0025x_3^2 - 0.005x_4^2 - 0.000862x_1x_2 + 0.00474x_1x_3 + 0.00465x_1x_4 + 0.00225x_2x_3 + 0.0039x_2x_4 - 0.0142x_3x_4 \quad (5)$$

According to ANOVA in Table 1, it can be seen that the p-value of the regression model is less than 0.05 indicating that this model can shows a significant relationship between the input variables of the data set. Furthermore, the quality of the polynomial regression model is expressed in terms of the value of the correlation coefficient (R^2). Therefore, a value of R^2 in ANOVA indicates that 95.13% of the response variability is explained by the model. Furthermore, input variable x_1, x_3, x_1^2, and x_1x_3 have a low p-value 0.025, 0.005, 0.038 and 0.019 respectively. That mean, these input variables have a high relationship with the response variable than others.

D. Combine forecasting results

In this study, the variance-covariance method has been used to combine the forecasting results from RSM and ARIMA model. This method calculates the weights by using the historical performance of the single forecasts model into consideration.

The error of the combined forecast is:

$$e_c = y_t - f_{ct} = w_1 e_{1t} + w_2 e_{2t} \quad (6)$$

The weight (w_i) of the combined forecast can be calculates as:

$$w_1 = (\sum e_{2t}^2 - \sum e_{1t}\, e_{2t})/(\sum e_{1t}^2 + \sum e_{2t}^2 - 2\sum e_{1t}\, e_{2t}) \quad (7)$$

$$w_2 = (\sum e_{1t}^2 - \sum e_{1t}\, e_{2t})/(\sum e_{1t}^2 + \sum e_{2t}^2 - 2\sum e_{1t}\, e_{2t}) \quad (8)$$

E. Result and discussion

In this section, the forecasting result that is obtained from the regression model using Eq. (5) is compared with ARIMA (1,1,0) model and combine model. The forecasting results of the monthly WTI crude oil price from July 2015 to August 2016 are demonstrated in Figure 7.

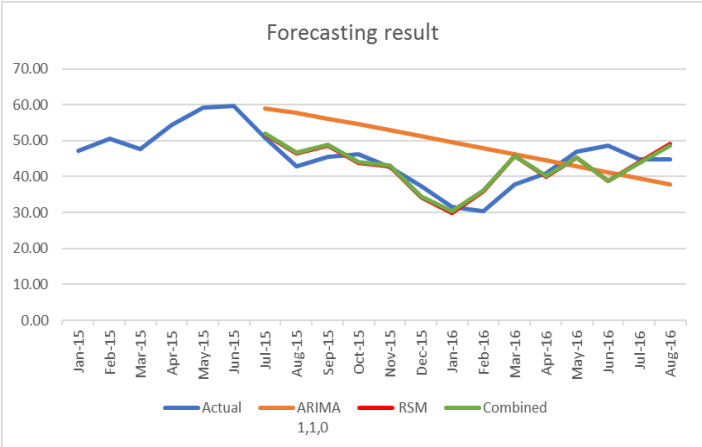

Figure 7. Forecasting result of WTI monthly crude oil price

Several criteria are used to evaluate the forecasting performance. In this research, two main evaluation criteria, mean absolute percentage error (MAPE) and direction statistic (D-stat), are used and defined as follows:

The MAPE represents the error of the forecasting values divided by actual values.

$$\text{MAPE} = \frac{\sum_{t=1}^{n} \frac{|e_t|}{y_t}}{n}$$

where e_t and y_t are forecast error and actual values of WTI crude oil price, respectively, and n is the number of observations. The accuracy of the proposed model for oil price forecasting is very important but on the other side, the trend of price changes is also as significant as the accuracy of forecasting. Direction statistic (D-stat) is used in this study to calculate the change in trend.

$$\text{D-stat} = \frac{1}{N}\sum_{t=1}^{N} a_t$$

where $a_t = 1$ if $(y_{t+1} - y_t)(\hat{y}_{t+1} - y_t) \geq 0$, and $a_t = 0$ otherwise.

The result of forecasting performance from two criteria are illustrated in Table 2.

TABLE 2. PERFORMANCE OF FORECASTING MODELS

Criteria	Proposed method	ARIMA (1,1,0)	Combined method
MAPE	0.081	0.250	0.080
D_{stat} (%)	71.43	57.10	71.43

The MAPE value in Table 2 has indicated that the forecasting performance of the proposed methodology for monthly WTI crude oil price can be acceptable, which is 0.081 when compared with the traditional method ARIMA with 0.250. The RSM has presented the better performance than the ARIMA model. Meanwhile, the value of D-stat demonstrates that the model can predict the direction of change of a time series from one time period to the next. Therefore, when comparing D-stat, the RSM still presented better performance than ARIMA, which are 71.4% and 57.1%, respectively. However, combined result from both methods has a slightly improve in term of MAPE but the results from D_{stat} is unchanged.

V. CONCLUSION AND FURTHER STUDY

In this paper, the RSM has been used as a new alternative methodology to forecast crude oil price. The results were evaluated by using two criteria, i.e. MAPE and D-stat. The results obtained from the case study has confessed that the proposed methodology has moderate forecasting power for the WTI crude oil price. However, the forecast value has a large error at some point on the time series data when compared with the actual value. That means the proposed model as shown is insufficient to describe the nonlinear dynamics. Meanwhile, the results of D-stat are satisfactory for predict the trend of oil price. Despite, the model itself is very simple and easy to understand and the good fitting of the model with high R-squared value and low p-value, it cannot work well to capture the complex dynamic behaviours of crude oil price. Furthermore, the forecasting results can be improved by using best curve fitting function after long-range data dividing into several short-range data and the result can be improved by combining the proposed methodology with Exponential smoothing method or Support vector machine (SVM).

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2014R1A2A1A11049845).

REFERENCES

[1] B. B. Niaz and P. M. Jose, "Crude oil price forecasting techniques: a comprehensive review of literature," *Alternative Investment Analyst Review*, pp. Volume 2, issue 3, 30-49, 2013.

[2] W. Xie, L. Yu, S. Xu and S. Wang, "A New Method for Crude Oil Price Forecasting Based on Support Vector Machine," *Computational Science-ICCS*, pp. 444-451, 2006.

[3] L. Yu, "Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm," *Energy Economics Volume 30, Issue 5,*, pp. 2623-2635, 2008.

[4] S. Z. Ali Ghaffari, "A novel algorithm for prediction of crude oil price variation based on soft computing," *Energy Economics*, pp. Volume 31, Issue 4, 531-536, 2009.

[5] J. Z. J. S. Michael Ye, "A monthly crude oil spot price forecasting model using relative inventories," *International Journal of Forecasting*, pp. Volume 21, Issue 3, 491-501, 2005.

[6] G. E. P. Box and G. M. Jenkins, "Time Series Analysis: Forecasting and Control," *An Unexpected Route to Time Series*, p. 575, 1976.

[7] G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing Volume 50*, p. 159–175, 2003.

[8] M. H. Myers, D. C. Montgomery and C. M. Anderson-Cook, Response Surface Methodology: Process and Product Optimization Using Designed Experiments, Third Edition, NY: A John Wiley and Sons Inc., 2009.

[9] S. M. Andre I. Khuri, "Response Surface Methodology," *Wires Computational Statistics*, pp. Volume 2, Issue 2, p 128-149, 2010.

[10] W. Xie, "A new method for crude oil price forecasting based on support vector machines," *Computational Science-ICCS*, pp. 444-451, 2006.