Iranian Journal of Business and Economics
Volume5,Issue4,2018
http://www.scijour.com/ijbe
SCIJOUR

Modeling and Forecasting Crude Oil Price with an Autoregressive Integrated Moving Average (ARIMA) Model

Ramezan Rezaeyan[1,*], Milad Taghizadeh[2]

[1,2] Department of the Financial Mathematics, Islamic Azad University, Nour Branch, Iran

ABSTRACT

Oil is very important as the largest energy consumer in the world due to its role in the economies of producing countries. Therefore, it is necessary to determine different parameters affecting the oil market for these countries. In this context, this paper forecasts the global market for oil prices as an important variable using ARIMA approach. It should be noted that this forecast is more dynamic. In this study, using ARIMA model, time series and a 7-year period from January 2009 to December in 2015, price of a single shipment of Iranian light crude oil was used.

KEYWORDS

Monthly forecast of crude oil prices, ARIMA, time series,modeling

INTRODUCTION

Today, oil plays an important role in the world changes as a commodity in political economy and as long as it is not generally replaced by other energy sources, it effects on the world economy and almost all human artifacts in the process of production and distribution energy, costs to transport are dependent to it. The importance of oil revenues on the economy and its impact on GDP is undeniable, so that the oil sector, not only is one of the most important economic activities impact on other economic variables, but also revenues from oil is a major supplier for government spending.

Therefore, determining the parameters affecting the oil market as oil prices in the present and the past may be necessary in understanding the proper way to correct the policies and strategies for economic,. By definition, prediction is a quantitative estimate (or estimations) about the likelihood of future events that are based on present and past. The predictions are used as a guide public and private policy because no programming knowledge is not possible to predict. ARIMA model is used as one of the Iranian economy and a tool for forecasting time series data. Thus, in this study among the models of the ARIMA, the model which has the lowest RMSE is used as the best model to predict monthly price of oil. Also in this paper, 5 years data (60 data) is used for training and 2 years (24 data) for prediction [6].

METHOD

1. Autoregressive Integrated Moving Average (ARIMA)

Predicting the daily time series the In this study for modeling and cross over recurrent average model is used. This model by the frequency of w and the levels of recurrent average model and seasonal subtracting of P,Q,D and non_seasonal of p,q,d which is showed as ARIMA(p,d,q)(P, D, Q) $_w$ can be considered as the following example. (Eq 5)

$$\Phi(B^w)\phi(B)(1-B^w)^D(1-B)^d x_t = \Theta(B^w)\,\theta(B)\varepsilon_t \quad (5)$$

In which x normalized time series Φ, Θ seasonal parameters and recurrent moving average model and ϕ , θ recurrent parameters and average of non-seasonal moving and B is the backward operator ، $B^m(x_t) = x_{t-m}$. And ε is a symbol for a pure random process which is used with a zero average and variance of $\sigma^2{}_\varepsilon$ [1].

Parametric modeling of time series by using the linear parameter model is divided into Initial analysis, identification and parameter estimation, Suitability test, Choosing the best model and prediction [1, 2].

- **Initial analysis:**

In this phase the ratio of static data is checked with the variance and the average. First the ratio of static data to the variance and then the ratio of static data to the average shall be examined. In this study the Box- Cox transformation is used for making data to be static. (Eq 6)

*Corresponding Author: Ramezan Rezaeyan
E-mail r:R.rezaeyan@gmail.com
Telephone Number r: Fax. Number r:

Iranian Journal of Business and Economics
Volume5,Issue4,2018
http://www.scijour.com/ijbe
SCIJOUR

$$\mathcal{Y} = \frac{x^{\lambda} - 1}{\lambda} \quad ; \quad x > 0 \quad , \quad \lambda \neq 0$$
$$\mathcal{Y} = Ln(x) \quad ; \quad x > 0 \quad , \quad \lambda = 0 \qquad (6)$$

And x is a data which shall be normalized and $\mathcal{Y}$ is the transformed amount and λ is a real value that should be set so that the distribution of the data as much as possible closer to the normal distribution [8].

- **Identify the basic model and parameter estimation:**

The basic model and parameters for modeling tools such as the autocorrelation function (ACF) and partial autocorrelation function (PACF) and the amount of Parameters are identified to be significantly associated with zero. Since the theory of time series developed based on static data Hence, In addition to the variance data must be static to the ratio of average. In case of non-stationary data to the average subtraction operation will be used [3]. (Eq7)

$$\nabla^{d} = (1 - B)^{d} \quad , \qquad \nabla^{D}{}_{w} = (1 - B^{w})^{D} \qquad (7)$$

D: Seasonal subtracting amount
d: non seasonal subtracting amou
w: Seasonal frequency period, B: backward operato

- **Suitability test:**

In this phase residuals being normal and independent shall be examined. To study normality of residuals correlation coefficient test is used. (PPCC).this test is offered by Vogel [4] in 1981.Also the Histogram diagram for residuals is used. In order to evaluate the residuals independency Box-Pierce test is used [3].

- **Choosing the best model:**

In this phase the model that has the least amount of errors MSE (Eq 8), yet it has the least number of parameters is chosen as the best model. In this study, an information measure called AIC that is provided by Akaike in 1974 (Eq 9). This test is based on estimates of variance. By using this model the model which has the lowest coefficient is selected as the best model [5, 7].

$$AIC = n_2 Ln(MSE) + 2(p + q) \qquad (9)$$
$$MSE = \frac{\sum_{t=1}^{n_1} (\hat{x}_t - x_t)^2}{n_1} \qquad (8)$$

n_2:The number of data, n_1: Number of observations
MSE: Mean square error
$\hat{x}_t$: The predicted amount, x_t: The real amount

- **Prediction:**

After the final model selection and forecasting models based on the root mean square error criterion (RMSE) shall be considered to examine (Eq 10).

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (\hat{x}_t - x_t)^2}{n}} \qquad (10)$$

n: number of observations
$\hat{x}_t$: predicted amount , x_t: The real observation amount

Results

As mentioned before, the first phase in modeling time series consists of stationary time series. In this study, stationary time series of monthly price of oil is performed using Box-Cox transformation. Studying Figure 1 illustrates the necessity of conversion actions and as Figure (2) shows, after applying the Box-Cox transformed, data is stationary relative to the variance [3].

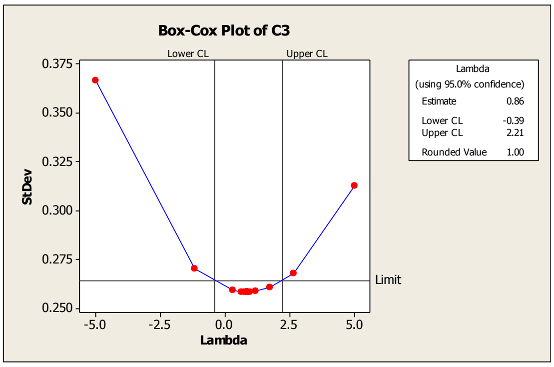

Fig.1.Diagram of Box-Cox before transformations

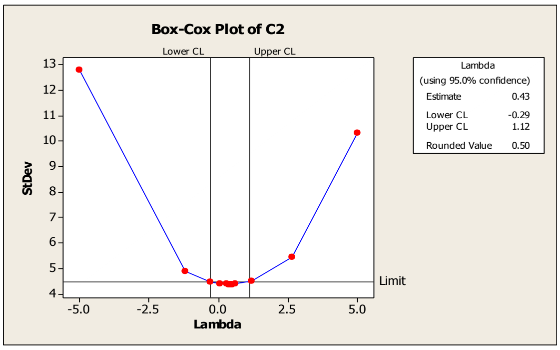

Fig.2. Diagram after Box-Cox transformation

The static data will be reviewed than the average; evaluating autocorrelation function (ACF) has been shown in figure 3, displays non-static, so there is a need to apply a simple subtraction.

As shown in Figure 4, after a simple subtraction, data are static compared to the average and static data after performing a simple subtraction indicates that data are non-seasonally.

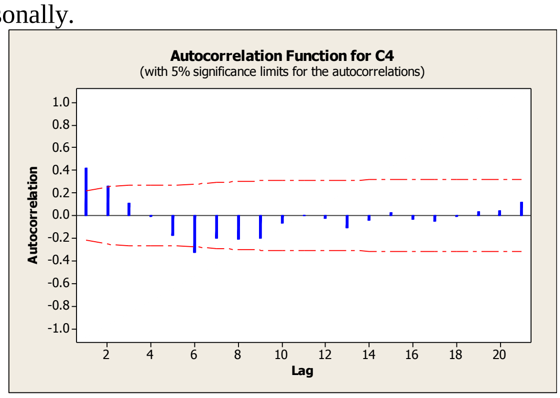

Fig.3. ACF plot before performing a simple subtraction

Iranian Journal of Business and Economics
Volume5,Issue4,2018
http://www.scijour.com/ijbe
SCIJOUR

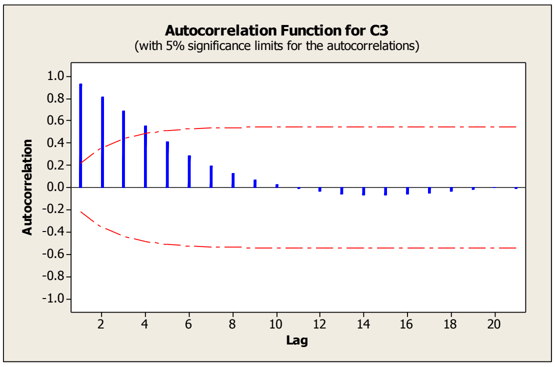

Fig.4.ACF plot after a simple subtraction

ACF and PACF functions are identified by the initial model; it is possible to review the choice of model. Mira functions of ACF and PACF, generally require the presence of non-seasonal moving average components returned to his modeling but it is better to identify the most appropriate model of all models, in this study may be considered the same for all models, ARIMA (p, 1, q) with different times returned to his non-seasonal moving average (p, q = 0, 1,2,3).

In estimating the model parameters, they are significantly different from zero. Thus, comparison of t is equal to the estimated standard deviation divided by the estimated parameters for each of the t-student test estimated for testing the significance of parameters than the null hypothesis of (parameter zero equivalent) mean the 5 is used and therefore the parameters of the models were significantly greater than 5 percent of the competing models excluded. At the end of the model (0,1,1), (2,1,0), (1,1,0), which is a significant test parameters tests, including normality and independence of residuals hearts fit well spent model (1,1,0) who are the lowest MSE was selected as the best model in table 1.

Tab.1.Comparison of MSE

Model	(1,1,0)	(0,1,2)	(0,1,1)
MSE	0.020	0.021	0.018

In order to ensure selection of the best model, the AIC information criterion was used in Table 2 and finally re-model (1,1,0) ARIMA time series studied were suitable for predicting the AIC criterion, resulting in the ability to identify the best model proves to be.

Tab.2.Comparison of AIC

Model	(1,1,0)	(0,1,2)	(0,1,1)
AIC	-234.72	-231.79	-241.04

Note that the test root mean square error (RMSE) as the ultimate test is to select the best model for prediction of these tests were used to compare the performance of prediction models and re-model ARIMA (0,1,1) have is the lowest RMSE was selected as the best model for predicting the RMSE values for the three selected models in table 3.

Tab.3.Comparison of the RMSE

Model	(1,1,0)	(0,1,2)	(0,1,1)
RMSE	0.0143	0.0147	**0.0137**

Note that the model (1,1,0) ARIMA was selected as the best model among all the models in Table 4 are given the values of each of the selected model parameters. As you can see below the 5% significance level corresponding t-statistics which show that effective participation in modeling parameters.

Table 4: The parameters

Parameters	Coef	SE Coef	T	P-Value
β	0.3519	0.1229	-2.86	0.006

Box-Pierce test and the values obtained by this test shows that all these values and these values represent more than 5% of the residues being independent of the table (5) show is a 5 the autocorrelation function (ACF) shows the residuals confirms the independence of the residuals as well.

Tab.5. Box-Pierce test values

Lag	12	24	36	48
P-Value	0.092	0.444	0.674	0.769

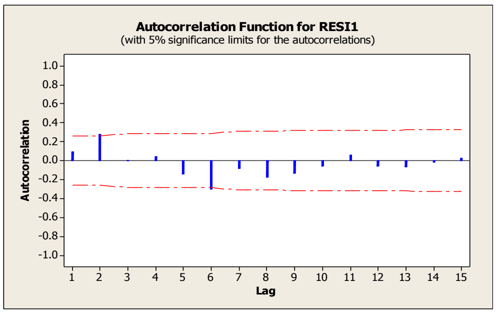

Fig. 5. ACF function residuals

Figure 6 shows the normal probability plot points located roughly along a straight line, and it shows that the residuals of the fitted model, ARIMA (0,1,1) are normally distributed (Figure 7) values predicted by the model (1,1,0) ARIMA compared to historical values .

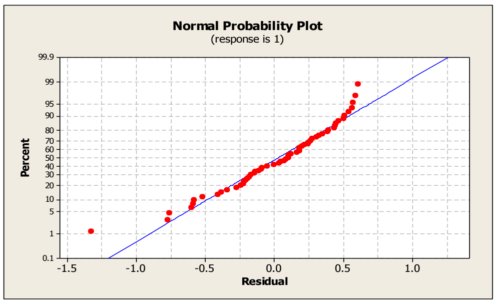

Fig.6.normal probability plot

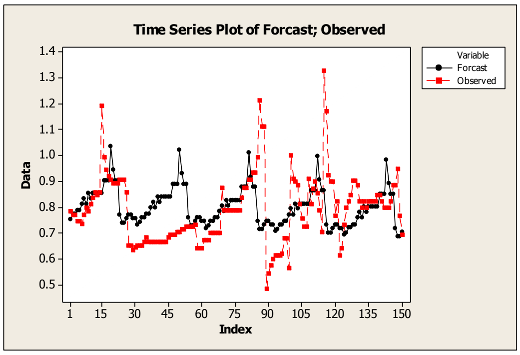

Fig.7.Comparison of the observed and predicted values

CONCLUSIONS

1. Comparing the values obtained by the standard RMSE indicated that the ARIMA (0,1,1) with the lowest error (RMSE) is the best model for prediction.
2. Selecting Model ARIMA (0,1,1) as the best model by the information criteria AIC indicates the ability of this criterion in the diagnosis of the best model.
3. ARIMA forecasts indicate that this method is not very accurate method of predicting the non-seasonal ARIMA feature.

REFERENCES

[1] ***Dargay, J.M., Gately, D. (2010).*** *World oil demand's shift toward faster growing and less price-responsive products and regions. Energy Policy 38, 6261–6277.*

[2] ***Fatemi Qomi. M. T., (1992)*** *.Forecasting and time series analysis, Amirkabir University Press,*

[3] ***Hossinpour k,M,, (1388),*** *Dynamic prediction of crude oil price using artificial neural networks and using petroleum storage OECD countries, Journal of Economic Research, vol 88, pp 25-46,*

[4] ***Hipel ,K .W and Mcleod, A. L (1994)*** *Time series modeling of water resources and environmental systems Elsevier New York*

[5] ***Niroumand. H. A. V.,*** *(1991). Introduction to time series analysis(translation), Ferdowsi University of Mashhad Press.*

[6] ***Suleiman, S., Alabi, M. A, Issa, Suleman, U. Usman, Umar Adamu, (2015),*** *Modeling and Forecasting the Crude Oil Price in Nigeria, International Journal of Novel Research in Marketing Management and Economics, Vol. 2, Issue 1, pp: (1-13).*

[7] ***Salas, J .D., Delleur ,J .W., Yevjevich ,V and Lane ,W .L (1988)*** *Applied modeling of hidrologic time series Water Resources Publications Littleton Colorado 498p*

[8] ***Vogel, R. M (1986)*** *The probability plot correlation coefficient test for normal lognormal and Gumbel distributional hypotheses Water Resources Research 20 pp 727-732*