Developing a new approach for forecasting the trends of oil price

Hossein Mombeini
Faculty of finance, Islamic Azad University, Dubai branch, Dubai, UAE

Abdolreza Yazdani-Chamzini
South Tehran Branch, Islamic Azad University, Tehran, Iran

Keywords
ARIMA, ANFIS, Oil price forecasting, Time series

Abstract

According to the importance of oil in economy of the world, different models have been developed for formulating the behavior of oil price. It is due to the fact that the models established based on the new techniques are more reliable and accurate on account of their ability in taking into account either linear or non-linear structures involved in the process of oil pricing. Auto-Regressive Integrated Moving Average (ARIMA) method is one of the most common time series models applied in forecasting over the resent decades. These researches indicate that two major limitations are found in past models: (1) ARIMA approach assume that there is a linear relationship between the future values of a time series with current and past values as well as a white noise; so that estimations obtained by ARIMA cannot be appropriate for modeling the nonlinear problems; and (2) a large number of historical data are required to satisfy the results. On the other hand, adaptive neuro-fuzzy inference system (ANFIS) is a powerful tool for modeling the non-linear structures. In this paper, ARIMA and ANFIS are used to overcome the limitations of conventional models, thus obtaining more accurate results. Empirical results of oil price forecasting indicate that the proposed model outperforms other methods and exhibits the accuracy of oil price forecasting is improved; so that, the proposed model can be applied as a proper option to forecast financial time series.

1. Introduction

Crude oil is one of the most significant production factors in many economies. It is mainly because of the key role of oil in the world economy. Forecasting the future price of this commodity and managing the risks associated with oil prices became essential for governments and businesses because the change of oil prices significantly affects other markets. However, high oil prices often lead to an increase in inflation and subsequently hurt economies of oil-importing countries and low oil prices may result in economic recession and political instability in oil-exporting countries since their economic development can get retarded (Zhang et al, 2008).

Autoregressive integrated moving average (ARIMA) models are one of the most common univariate linear models for time series forecasting over the recent decades that have demonstrated their potential applications in forecasting social, engineering, economic, stock, and management problems. These models are applied by different researchers for modeling of oil price changes (Arshad, Ghaffar, 1986). The generality of the ARIMA model is due to its high performance and robustness. In ARIMA analysis, an identified underlying process is generated based on observations to a time series for generating a fitted model that shows the process-generating mechanism precisely (Khashei, Bijari, 2011).

ARIMA models have a major drawback although they have the some advantages of accurate forecasting over a short period of time. These models assume that future values of a time series have a linear relationship with current and past values as well as with white noise, so estimations by ARIMA models may not be adequate for complex nonlinear real-world problems (Khashei et al, 2009). Therefore, it is not reasonable to assume that model is pre-assumed linear; whereas, the real-world time series are rarely pure linear combinations (Valenzuela et al, 2008). In particular, the political events and many complicated factors contributed to the change of the oil price during the last three decades have made oil prices appear highly nonlinear and even chaotic (Ghaffari, Zare, 2009). To overcome this limitation in modeling oil price, several nonlinear models have been developed in the literature, such as artificial neural networks (Yu et al, 2008; Tehrani, Khodayar, 2011; Movagharnejad et al, 2011; Jammazi, Aloui, 2012), hidden Markov model (Silva et al, 2010), and GARCH-type models (Morana, 2001; Kang et al, 2009; Li et al, 2010; Wei et al, 2010; Mohammadi, Su, 2010; Zhu, Ling, 2011; Hou, Suardi, 2012). However, the mentioned models are not able to handle simultaneously both complexity and uncertainty connected with oil price.

Adaptive neuro-fuzzy inference system (ANFIS) is a powerful tool for modeling and forecasting time series over the last decade that is applied by different researchers. This method is found to be a viable contender to various conventional time series models (Aznarte et al, 2007; Vairappan et al, 2009; Moreno, 2009; Tektaş, 2010; Ho, Tsai, 2011; Wang et al, 2011).

The unique forecasting features of ANFIS make this technique more popular in comparison with the traditional forecasting techniques. These can be due to the existing advantages in two methods artificial neural network (ANN) and fuzzy inference system (FIS) that form the ANFIS structure. However, ANN is capable to model all types of existing complexity and nonlinearity in the structure of the data under consideration. Likewise, FIS is successful in face of uncertain data and can take into account the human knowledge in modeling. ANFIS is recently applied by different researchers in many different areas of time series forecasting.

A study based on ANFIS for modeling photovoltaic power supply system by Mellit and Kalogirou (2011) was accomplished. Melin et al (2012) proposed a new approach for time series prediction using the integration of the ensembles of ANFIS models, including integrator by average and the integrator by weighted average. They demonstrated that the performance obtained with this architecture overcomes several standard statistical approaches and neural network models. An ANFIS approach is used by Firat and Güngör (2008) to construct a hydrological time-series forecasting system. Despite request of ANFIS models in different forecasting problems, there are few studies in literature which applied ANFIS for oil price forecasting.

The main objective of this study is to propose a new hybrid model based on ARIMA and ANFIS to investigate the dynamics and volatility of prices of West Texas Intermediate (WTI) crude oil markets in order to obtain a more accurate, precise, and sure model. In the proposed model, a time series is taken into account as function of a linear and a nonlinear component, so, in the first step, an ARIMA model is utilized to identify the linear patterns in the data set. The residuals of the ARIMA model will then comprise only the nonlinear structures. Hence, in the second step, an ANFIS is applied as a nonlinear tool in order to recognize the nonlinear patterns in the pre-processed data and to forecast the future value of time series. To show the capability and effectiveness of the proposed model to forecast time series, a real data set from crude oil price is evaluated.

2. Time series forecasting models

Several different methods are developed to time series forecasting, which can be classified into two main categories linear and nonlinear methods. Linear methods, including autoregressive integrated moving average (ARIMA) and exponential smoothing, forecast the future values as a linear function of past observations. The rest of the methods such as the autoregressive conditional heteroscedastic (ARCH) model, general autoregressive conditional heteroscedastic (GARCH), artificial neural networks, fuzzy inference system (FIS), adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM) are classified into nonlinear models. In this section, the basic principles of the two components of the proposed model (ARIMA and ANFIS) are illustrated.

2.1. ARIMA model

The autoregressive integrated moving average ARIMA(p,d,q) model is well-known as an important forecasting tool that formalizes a problem by defining three steps: identification of the model, estimation of the coefficients and verification of the model (Box and Jenkins, 1976). Forecasts by this method are under the assumption that the past history can be translated into predictions for the future (Koutroumanidis et al, 2009). The ARIMA model is used when time series is non-stationary and so stationary AR, MA or ARMA processes cannot be directly applied. One general way of handling non-stationary series is to use differencing so as to convert them into stationary series. An ARIMA model combines three different processes including an auto-regressive (AR) process, integration part, and a moving average (MA) process (Melin et al, 2012). The former indicates that the time series it is a weighted linear sum of the past p values plus a random shock. The latter indicates that the time series is a weighted linear sum of the last q random shocks. If the original data series is differenced d times before fitting an ARMA(p,q) process, then the model for the original undifferentiated series is said to be an ARIMA(p,d,q) process where the letter 'I' in the acronym stands for integrated and d denotes the number of differences taken (Chatfield, 2000). The ARIMA can be mathematically written as:

$$\phi(B)(1-B)^d X_t = \theta(B)Z_t \tag{1}$$

where Z_t denotes a purely random process with zero mean and variance σ_z^2, and B shows the backward shift operator ($BX_t = X_{t-1}$).The first differences, expressed as $(X_t - X_{t-1}) = (1-B)X_t$, may themselves be differenced to give second differences, and so on; so that, the dth differences can be addressed as $(1-B)^d X_t$.

2.2. ANFIS modeling

ANNs are one of the most popular forecasting methods in presence of non-linearity of the data set, while a fuzzy inference system applying fuzzy if-then rules can handle the qualitative aspects of human knowledge and are expected to obtain better results than ANNs in facing with uncertainty caused by less and/or even lack of information and lack of clarity. Consequently, the integration of these two useful techniques, well-known as neuro-fuzzy approach, is a powerful tool for modeling under the sophisticated environments with many different effective parameters.

A fuzzy inference system applied in the form of a neuro-fuzzy system with crisp functions in consequents is the Takagi-Sugeno-type fuzzy system, which is well-known as ANFIS. ANFIS was initially developed by Jang (1993). The ANFIS can be trained to tune its parameters and learn the existing structures in the data set. The relationships between input and output variables are represented by means of fuzzy if–then rules with unclear predicates. The ANFIS model is established by adapting the antecedent parameters and consequent parameters; so that, a specified objective function (usually a difference between the model output and the actual output) is minimized (Alizadeh et al, 2012). Several methods are developed for learning rules (Jang, 1993; Mizutani, 1996; Tang et al, 2005). In this paper, the hybrid learning algorithm introduced by Jang (1993) that is a combination of least square estimation and back-propagation algorithms is applied.

To show the ANFIS scheme for improving the forecasting capability of the oil price model, for example, two if-then rules can be taken into account as follows:
Rule 1 : If x is A_1 and y is B_1 then $f_1 = p_1 x + q_1 y + r_1$
Rule 2 : If x is A_2 and y is B_2 then $f_2 = p_2 x + q_2 y + r_2$

where x and y are two inputs; A_i and B_i are the terms which are represented by fuzzy sets. f_i are the output variables within the fuzzy region specified by the fuzzy rule whose membership function parameters are premise parameters. p_i, q_i and r_i are designing parameters which are obtained during the learning process (Mehrabi et al, 2011). The ANFIS architecture with two rules is depicted in Fig. 1. Output of each node in every layer is indicated by O_i^l (ith node output in lth layer). The performance of each layer can be described as follows:

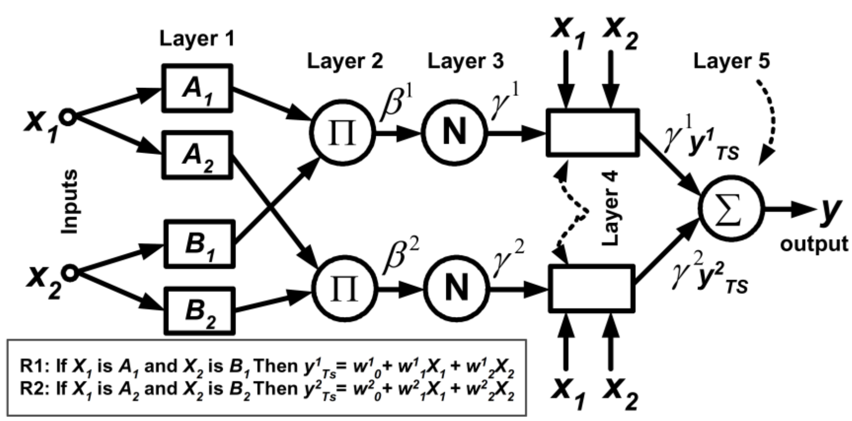


Fig. 1. ANFIS architecture with two rules

The first layer is the fuzzifying layer in which A_i and B_i are the linguistic terms. The output of the layer is the membership functions of these linguistic terms are given as:

$$O_i^l = \mu_{A_i}(x) \tag{2}$$

$$O_i^l = \mu_{B_i}(y)$$

In the second layer, the rules' firing strengths are calculated by multiplying each node with each other as presented in the following equation:

$$w_i = \mu_{A_i}(x)\mu_{B_i}(y), \qquad i = 1,2 \tag{3}$$

where μ_{A_i} and μ_{B_i} are the membership functions of the input variables x and y, respectively. In the third layer, the firing strengths obtained in the previous layer of the nodes are normalized. Every node in the layer computes the ratio of the ith rule's firing strength to the sum of all rules' firing strengths by using Eq. (4):

$$\bar{w}_i = \frac{w_i}{w_1 + w_2}, \qquad i = 1,2 \tag{4}$$

In the fourth layer, every node calculates a linear function where the function coefficients are adapted by using the back propagation algorithm of the artificial neural networks (Jeong et al, 2012). The output of this layer is derived from multiplication of normalized firing strength obtained in the third layer by first order of Sugeno fuzzy rule.

$$\bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i), \qquad i = 1,2 \tag{5}$$

In the fifth layer, the overall output is calculated as a summation of all the incoming signals through Eq. (6):

$$\sum_{i=1}^{2} \bar{w}_i f_i = \frac{\sum_{i=1}^{2} w_i f_i}{\sum_{i=1}^{2} w_i} \tag{6}$$

The first and fourth layers in ANFIS structure are adaptive layers. The consequent coefficients (p_i, q_i and r_i) are continuously adjusted by using fuzzy membership functions in order to minimize the errors between the model outputs and the observations (Jeong et al, 2012).

3. The proposed model

The accuracy of time series forecasting is a critical problem to many decision makers; so that, a number of time series models are developed by different researchers to improve the efficiency of forecasting models. The results of various studies show that predictive performance improves in hybrid models (Taskaya, Casey, 2005; Khashei et al, 2009; Khashei, Bijari, 2011). The main aim of using hybrid models is to decrease the risk of failure by integrating different models to obtain more accurate and precise results. Generally, this is due to the existing some difficulties connected with identifying whether a time series under study is generated from a linear or nonlinear underlying process as well as a single model may not be capable to capture different patterns equally well (Zhang, 2003).

Hybrid models that decompose a time series into its linear and nonlinear form have been demonstrated their capability and effectiveness to model complex structures in the real world problems. Nourani et al (2011) introduced two hybrid AI-based models which are reliable in capturing the periodicity features of the process for watershed rainfall–runoff modeling, including the SARIMAX (Seasonal Auto Regressive Integrated Moving Average with exogenous input)-ANN and wavelet-ANFIS models. Zhang (2003) suggested a hybrid methodology that combines both ARIMA and ANN models to take advantage of the unique strength of ARIMA and ANN models in linear and nonlinear modeling. Rojas et al (2008) presented a new procedure to predict time series using paradigms such as: fuzzy systems, neural networks and evolutionary algorithms. Che and Wang (2010) proposed a hybrid model that combines both

SVR and ARIMA models to take advantage of the unique strength of SVR and ARIMA models in nonlinear and linear modeling. Khashei and Bijari (2010) proposed a novel hybrid model of ANNs using ARIMA models in order to yield a more accurate forecasting model than ANNs. A new hybrid approach combining Elman's Recurrent Neural Networks (ERNN) and ARIMA models is proposed by Aladag et al (2009).

Tseng et al (2002) proposed a hybrid forecasting model, which combines the seasonal time series ARIMA (SARIMA) and the neural network back propagation (BP) models. A hybrid forecasting scheme which combines the classic SARIMA method and wavelet transform (SW) is proposed by Choi et al (2011). Koutroumanidis et al (2009) used a hybrid model based on ARIMA-ANN to predict the future selling prices of the fuel wood produced by Greek state forest farms. A novel hybrid method to forecast day-ahead electricity price is proposed by Shafie-khah et al (2011). This hybrid method is based on wavelet transform, ARIMA models and Radial Basis Function Neural Networks (RBFN). Valenzuela et al (2008) proposed a hybridization of intelligent techniques such as ANNs, fuzzy systems and evolutionary algorithms, so that the final hybrid ARIMA–ANN model could outperform the prediction accuracy of those models when used separately. Egrioglu et al (2009) proposed a hybrid approach based on partial high order bivariate fuzzy time series forecasting model in order to analyze seasonal fuzzy time series. A novel hybridization of ANNs and ARIMA model is proposed by Khashei and Bijari (2011) in order to overcome mentioned limitation of ANNs and yield more general and more accurate forecasting model than traditional hybrid ARIMA-ANNs models.

On the other hand, in the problems of real world there is rarely a complete knowledge of the process. This causes to the inherent complexity be involved in the problems that leads to uncertainty and imprecision in the process of modeling. Fuzzy models can face with the existing complexity in the systems by a flexible way. A model which mimics the target system without needing to detailed first principles knowledge can be constructed by using the concepts of fuzzy sets and approximate reasoning. However, the above-mentioned hybrid models are not able to handle simultaneously both complexity and uncertainty connected with time series.

In this paper, a new hybridization of ARIMA and ANFIS models is introduced in order to overcome the above-mentioned limitations of conventional hybrid models and also limitations of linear and nonlinear models by using the unique features of ARIMA and ANFIS models in linear and nonlinear modeling, respectively. According to the structure of hybrid models, a time series can be taken into account to be composed of a linear autocorrelation structure and a nonlinear component (Khashei and Bijari, 2011). It is reasonable to take into consideration a time series to be composed of the function of a linear and a nonlinear component as follows:

$$Y_t = L_t + N_t \tag{7}$$

where L_t and N_t are the linear and the nonlinear components, respectively. Therefore, the key idea is to employ in the first place an ARIMA model, and next apply an ANFIS in order to model the residuals from the linear model. Let e_t address the residual at time t from the linear model, then

$$e_t = Y_t - \hat{L}_t \tag{8}$$

where $\hat{L}_t$ is the value forecasted by the ARIMA model for time t. Residuals are important to recognize the efficiency of linear models; so that, a linear model is not efficient if there are still linear correlation structures left in the residuals (Zhang, 2003). However, any nonlinear pattern left in the residuals indicates the drawbacks of the ARIMA model. By applying ANFIS, the existing nonlinear relationships in the pre-processed data can be extracted. The ANFIS model with n input nodes for the residuals can be defined as

$$e_t = f(e_{t-1}, e_{t-2}, \ldots, e_{t-n}) + \varepsilon_t \tag{9}$$

where f is a nonlinear function yield by the ANFIS and ε_t is the random error. The fitted model selection is important because the error term may be not essentially random. If the forecast from Eq. (7) be displayed as $\hat{N}_t$, the hybrid model can be written as

$$\hat{Y}_t = \hat{L}_t + \hat{N}_t \tag{10}$$

4. Forecasting model

In this section, the proposed model based on ARIMA and ANFIS is implemented in order to demonstrate the appropriateness and capability of the model. The ARIMA is modeled to extract the linear relationship between the oil price data set by using Eviews Software and the ANFIS is handled to recognize the non-linear relationship between the pre-processed data set by using the MATLAB Fuzzy Logic Toolbox. We analyze West Texas Intermediate (WTI) crude oil spot prices (in US dollars per barrel). The datasets consist of monthly prices over the period from April 1996 to November 2005 and contains 116 observations for WTI crude oil market. The plot of the time series is depicted in Fig. 2. Basic descriptive statistics is listed in the Table 1.

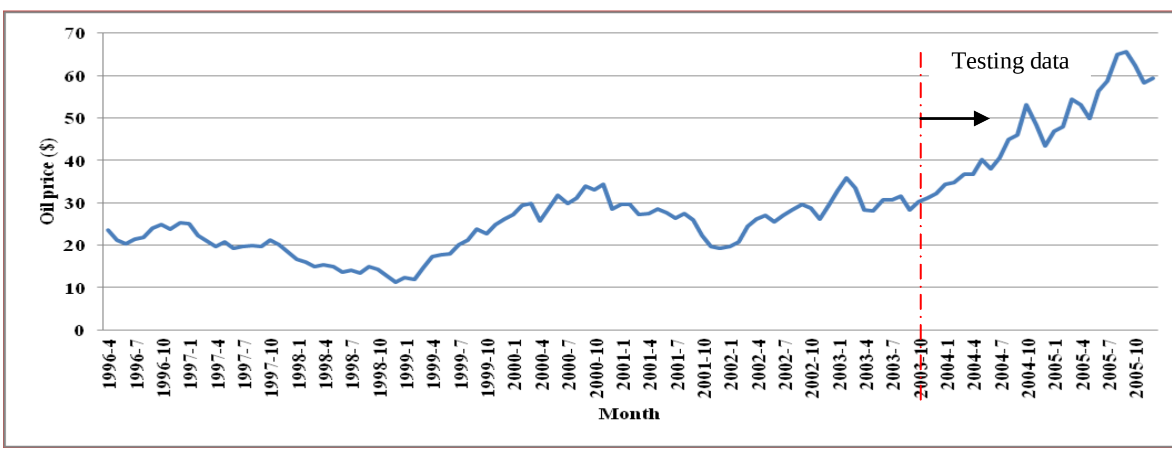


Fig. 2. WTI oil price in the period 1996/04-2005/11

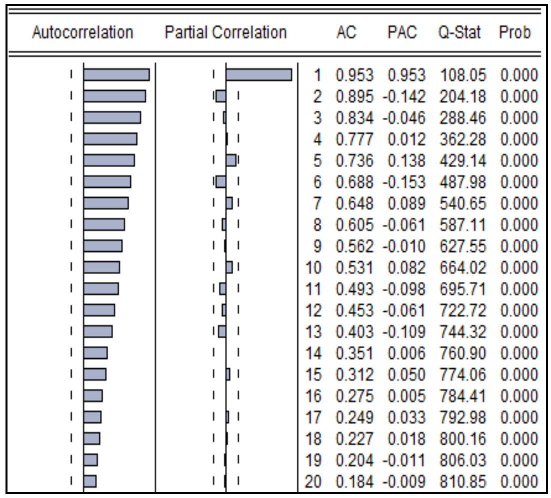
Table 1. Basic descriptive statistics

Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	Jarque-Bera
28.7444	26.98	65.57	11.28	12.03937	1.185727	4.115927	33.20061

4.1. Constructing the ARIMA model

From an inspection of the time plot graph, it is clear that the level of oil price is a non-stationary time series. Likewise, according to correlogram plotted in Fig. 3, there is a linear decrease for the autocorrelation functions (ACFs) as well as only one significant spike in period

1 can be seen for partial correlation function (PACFs). Hence, the time series can be stationary by removing the trend using first differenced data. The result of the unit root test is presented in Table 2 and the Augmented Dickey-Fuller (ADF) test confirms that the first difference is stationary.



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.953	0.953	108.05	0.000
		2	0.895	-0.142	204.18	0.000
		3	0.834	-0.046	288.46	0.000
		4	0.777	0.012	362.28	0.000
		5	0.736	0.138	429.14	0.000
		6	0.688	-0.153	487.98	0.000
		7	0.648	0.089	540.65	0.000
		8	0.605	-0.061	587.11	0.000
		9	0.562	-0.010	627.55	0.000
		10	0.531	0.082	664.02	0.000
		11	0.493	-0.098	695.71	0.000
		12	0.453	-0.061	722.72	0.000
		13	0.403	-0.109	744.32	0.000
		14	0.351	0.006	760.90	0.000
		15	0.312	0.050	774.06	0.000
		16	0.275	0.005	784.41	0.000
		17	0.249	0.033	792.98	0.000
		18	0.227	0.018	800.16	0.000
		19	0.204	-0.011	806.03	0.000
		20	0.184	-0.009	810.85	0.000

Fig. 3. Correlogram of oil price

Table 2. Unit root test

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.916741	0.0000
Test critical values:	1% level	-3.488585	
	5% level	-2.886959	
	10% level	-2.580402	

Then, the best ARIMA model is identified by comparing different ARIMA models based on their performance as shown in Table 3. From Table 3, it is clear that the indices to judge for the best model demonstrate ARIMA (1,1,0) has relatively small of AIC and HQC; therefore, is a relatively best model. The results of the ARIMA model in the training and testing data set on forecasting performance is listed in the last row of Table 4.

Table 3. Comparison of ARIMA models

ARIMA model	Akaike info criterion (AIC)	Schwarz-Bayesian criterion (BIC)	Hannan-Quinn criterion(HQC)
(1,1,0)	**4.191684**	4.247987	**4.214367**
(2,1,0)	4.193781	4.250469	4.216608
(3,1,0)	4.220855	4.277933	4.243826
(1,1,1)	4.204639	4.289094	4.238664
(1,1,2)	4.192717	4.277172	4.226742
(0,1,1)	4.192016	**4.24794**	4.214558
(0,1,2)	4.198566	4.254491	4.221108
(0,1,3)	4.206697	4.262621	4.229238

4.2. ANFIS model

In this step, the existing nonlinear relationships in the pre-processed data in the previous step can be extracted with the help of the ANFIS model. In order to generate fuzzy rules in the ANFIS system, each input parameter can be classified into several class values, and each fuzzy rule would be constructed using two or more membership functions (MFs) (Noori et al, 2010). There are different methods to classify the input data and generate the fuzzy rules. Grid partition is one of the most popular methods, especially when there are a few input variables. In this paper because of few number of input variables, a grid partition fuzzy clustering was applied to generated a knowledge base according to relationship between the input and output variables. Before constructing the ANFIS model, all variables were normalized to the interval of 0 and 1 through following equation:

$$X_{norm} = (X - X_{\min}) / (X_{\max} - X_{\min}) \quad (11)$$

The ANFIS algorithm should be trained using the training data to the parameters of Sugeno-type fuzzy inference system based on the hybrid algorithm be identified; so that the output of the network be fitted with actual values. Training should be evaluated by testing data set to the network be not overfitted. Therefore, the testing data set are interred into the trained network to the responses of the network be evaluated by comparing with the actual values for the performance measurement. Forming the architecture of network plays an important role in modeling the ANFIS for a time series, which this task is implemented by selecting a proper number of membership functions (q) and the element of input vector (p, the lagged observations). However, there is no theory that can be applied to guide the selection of q and p; therefore, the best results can only be found by a process of trial and error to select an appropriate value for p and q. Table 4 presents the result for the effect of different numbers of input and MFs in the training and testing data set on forecasting performance of the ANFIS model for selecting the best-fitted model.

According to there is not a unique and more appropriate unbiased estimators applied to see how far the model is able to forecast the values of oil price, several measures of accuracy are employed. For this reason, the models are evaluated by four estimators contain of the coefficient of determination (R^2), the persistence index (PI), the square root of the mean square error (RMSE), and the mean absolute percentage error (MAPE).

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (A_i - P_i)^2}{\sum_{i=1}^{N} (A_i - \bar{A}_i)^2} \quad (12)$$

$$MAPE = 100 \times \frac{1}{N} \sum_{i=1}^{N} \left| \frac{A_i - P_i}{A_i} \right| \quad (13)$$

where P_i is predicted values, A_i is observed values, $\bar{A}_i$ is the average of observed set, and N is the number of the observations of the validation set, and PI is calculated by (Kitanidis, Bras, 1980; Anctil, Rat, 2005; Díaz-Robles et al, 2008):

$$PI = 1 - \frac{\sum_{i=1}^{N} (A_i - P_i)^2}{\sum_{i=1}^{N} (A_i - A_{i-1})^2} \quad (14)$$

From Table 4, it is evident that the ANFIS model with 1 input variable (the first-lagged observation) and 7 MFs outperforms others models. The forecasted value of ARIMA and the proposed model for test data are plotted in Fig. 4.

Table 4. Comparison among performances of different models

Lags	MFs	In sample			Out of sample		
		R^2	MAPE	PI	R^2	MAPE	PI
1	3	0.945	7.231	0.083	0.632	22.73	-5.45
1	4	0.951	6.786	0.071	0.57	23.46	-5.51
1	5	0.953	7.094	0.074	0.672	17.42	-4.46
1	6	0.941	6.563	0.069	0.72	19.86	-2.33
1	7	**0.975**	**3.542**	**0.041**	**0.873**	**5.642**	**0.200**
1	8	0.955	4.231	0.053	0.82	6.981	0.252
1	9	0.961	7.432	0.075	0.853	7.344	0.312
2	3	0.958	6.752	0.069	0.785	10.56	0.342
2	4	0.943	5.872	0.042	0.796	11.27	-2.35
2	5	0.969	6.21	0.083	0.704	15.46	0.289
2	6	0.893	9.342	0.087	0.812	14.77	0.312
2	7	0.927	8.697	0.079	0.807	13.56	0.336
2	8	0.947	6.254	0.058	0.69	20.34	-3.39
2	9	0.965	7.876	0.063	0.847	31.28	-4.42
3	3	0.899	9.328	0.122	0.769	27.36	0.268
3	4	0.966	8.231	0.113	0.683	29.57	0.383
3	5	0.943	5.871	0.065	0.811	16.98	0.413
3	6	0.969	5.653	0.057	0.588	35.67	-5.42
4	3	0.953	6.233	0.063	0.678	38.12	-1.39
4	4	0.952	7.346	0.076	0.822	7.36	0.285
4	5	0.964	6.892	0.071	0.764	22.16	0.454
5	3	0.971	6.435	0.065	0.783	31.23	-2.38
5	4	0.959	7.568	0.059	0.847	7.69	0.278
6	3	0.896	9.322	0.062	0.834	8.23	0.265
6	6	0.957	6.547	0.056	0.817	16.84	-2.27
ARIMA		0.973	6.467	0.061	0.85	16.24	-4.59

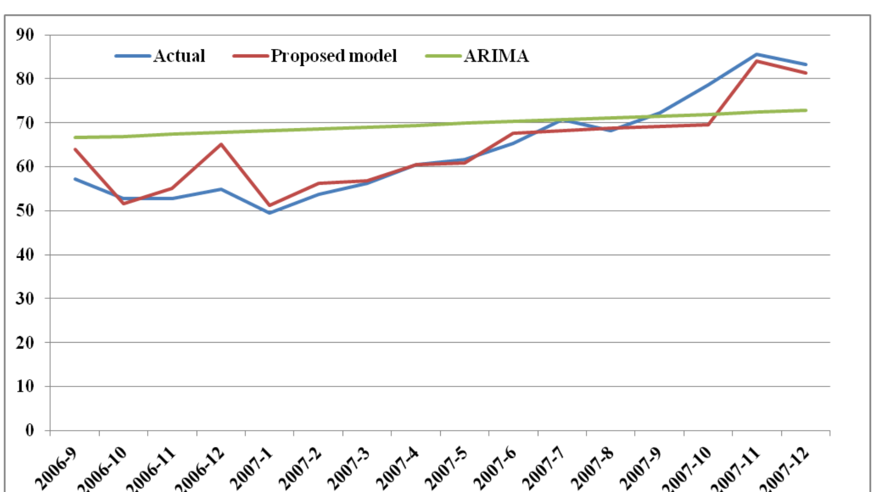


Fig. 4. Actual and forecasted values during validation by ARMA and the proposed model for oil price

5. Conclusion

In this paper a hybrid method is proposed to model the oil price trend. The unique feature of the proposed method lies in integration of well-known and robust ARIMA and ANFIS models to find the appropriate method with most accurate assessments. This model is capable to take the advantages of both the ARIMA and the ANFIS models. For achieving the aim, different statistical analyses using MAPE, R^2, and PI tests are performed on the results of ARIMA and the proposed model. The proposed model can be easily used to complex and uncertain environments. The proposed model receives the data filtered using the ARIMA model and then applies ANFIS system for these pre-processed data to model the existing non-linearity in the time series involved. The monthly oil price data from West Texas Intermediate (WTI) were employed to test the proposed model in comparison with the ARIMA model. The results demonstrate the best performance can be yield by the proposed model during the training and validation phases. The forecasting results of the proposed model during both phases outperform the ARIMA model. Therefore, the results of the study indicate that the proposed model has a high potential in modeling the trend of oil price, and this may provide valuable suggestion for researchers to apply this method for modeling the trend of time series.

References

Aladag, C.H., Egrioglu, E., Kadilar, C., (2009). Forecasting nonlinear time series with a hybrid methodology. Applied Mathematics Letters 22: 1467-1470.

Alizadeh, M., Jolai, F., Aminnayeri, M., Rada, R., (2012). Comparison of different input selection algorithms in neuro-fuzzy modeling. Expert Systems with Applications 39: 1536-1544.

Anctil, F., Rat, A., (2005). Evaluation of neural network stream flow forecasting on 47 watersheds. Journal of Hydrologic Engineering 10: 85-88.

Arshad, F. M., Ghaffar, R. A., (1986). Crude Palm Oil Price Forecasting: Box-Jenkins Approach. Pertanika 9(3): 359 - 367

Aznarte, J.L.M., Sánchez, J.M.B., Lugilde, D.N., Fernández, C.L., Guardia, C.D., Sánchez, F.A. S., (2007). Forecasting airborne pollen concentration time series with neural and neuro-fuzzy models. Expert Systems with Applications 32: 1218-1225.

Chatfield, Ch., (2000). Time-series forecasting. Chapman & Hall/CRC.

Che, J., Wang, J., (2010). Short-term electricity prices forecasting based on support vector regression and Auto-regressive integrated moving average modeling. Energy Conversion and Management 51(10): 1911-1917.

Choi, T.M., Yu, Y., Au, K.F., (2011). A hybrid SARIMA wavelet transform method for sales forecasting. Decision Support Systems 51: 130-140.

Díaz-Robles, L. A., Ortega, J. C., Fu, J. S., Reed, G. D., Chow, J. C., Watson, J. G., Moncada-Herrera, J. A., (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. Atmospheric Environment 42: 8331-8340.

Egrioglu, E., Aladag, C.H., Yolcu, U., Basaran, M.A., Uslu, V.R., (2009). A new hybrid approach based on SARIMA and partial high order bivariate fuzzy time series forecasting model. Expert Systems with Applications 36: 7424-7434.

Firat, M., Güngör, M., (2008). Hydrological time-series modelling using an adaptive neuro-fuzzy inference system. Hydrol. Process. 22: 2122-2132.

Ghaffari, A., Zare, S., (2009). A novel algorithm for prediction of crude oil price variation based on soft computing. Energy Economics 31: 531-536.

Ho, Y.Ch., Tsai, Ch.T., (2011).Comparing ANFIS and SEM in linear and nonlinear forecasting of new product development performance. Expert Systems with Applications 38: 6498-6507.

Hou, A., Suardi, S., (2012). A nonparametric GARCH model of crude oil price return volatility. Energy Economics, in press.

Jammazi, R., Aloui, Ch., (2012). Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. Energy Economics, in press.

Jeong, H., Hwang, W., Kim, E., Han, M., (2012). Hybrid modeling approach to improve the forecasting capability for the gaseous radionuclide in a nuclear site. Annals of Nuclear Energy 42: 30-34.

Kang, S. H., Kang, S. M., Yoon, S. M., (2009). Forecasting volatility of crude oil markets. Energy Economics 31: 119-125.

Khashei, M., Bijari, M., (2010). An artificial neural network (p,d,q) model for time series forecasting. Expert Systems with Applications 37: 479-489.

Khashei, M., Bijari, M., (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Applied Soft Computing 11: 2664-2675.

Khashei, M., Bijari, M., Ardali, Gh.A.R., (2009). Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and Artificial Neural Networks (ANNs). Neurocomputing 72: 956-967.

Kitanidis, P.K., Bras, R.L., (1980). Real time forecasting with a conceptual hydrological model. 2: applications and results. Water Resources Research 16: 1034-1044.

Koutroumanidis, T., Ioannou, K., Arabatzis, G., (2009). Predicting fuel wood prices in Greece with the use of ARIMA models, artificial neural networks and a hybrid ARIMA–ANN model. Energy Policy 37: 3627-3634.

Koutroumanidis, T., Ioannou, K., Arabatzis, G., (2009). Predicting fuel wood prices in Greece with the use of ARIMA models, artificial neural networks and a hybrid ARIMA–ANN model. Energy Policy 37: 3627-3634.

Li, H., Wang, Sh., Shang, W., (2010). Modeling volatility in crude oil market: a comparison between GARCH and extreme-value-based av model, Dynamics of Continuous, Discrete and Impulsive Systems, Series B: Applications & Algorithms 17: 67-82.

Mehrabi, M., Pesteei, S.M., Pashaee,, T. G., (2011). Modeling of heat transfer and fluid flow characteristics of helicoidal double-pipe heat exchangers using Adaptive Neuro-Fuzzy Inference System (ANFIS). International Communications in Heat and Mass Transfer 38: 525-532.

Melin, P., Soto, J., Castillo, O., Soria, J., (2012). A new approach for time series prediction using ensembles of ANFIS models. Expert Systems with Applications 39: 3494-3506.

Mellit, A., Kalogirou, S.A., (2011). ANFIS-based modelling for photovoltaic power supply system: A case study. Renewable Energy 36: 250-258.

Mohammadi, H., Su, L., (2010). International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. Energy Economics 32: 1001-1008.

Morana, C., (2001). A semiparametric approach to short-term oil price forecasting. Energy Economics 23: 325-338.

Moreno, J., (2009). Hydraulic plant generation forecasting in Colombian power market using ANFIS. Energy Economics 31: 450-455.

Movagharnejad, K., Mehdizadeh, B., Banihashemi, M., Kordkheili, M. Sh., (2011). Forecasting the differences between various commercial oil prices in the Persian Gulf region by neural network. Energy 36: 3979-3984.

Nourani, V., Kisi, Ö., Komasi, M., (2011). Two hybrid Artificial Intelligence approaches for modeling rainfall–runoff process. Journal of Hydrology 402: 41-59.

Rojas, I., Valenzuela, O., Rojas, F., Guillen, A., Herrera, L.J., Pomares, H., Marquez, L., Pasadas, M.., (2008). Soft-computing techniques and ARMA model for time series prediction. Neurocomputing 71: 519-537.

Shafie-khah, M., Moghaddam, M. P., Sheikh-El-Eslami, M.K., (2011). Price forecasting of day-ahead electricity markets using a hybrid forecast method. Energy Conversion and Management 52: 2165-2169.

Silva, E. G. S., Legey, L. F. L., Silva, E. A. S., (2010). Forecasting oil price trends using wavelets and hidden Markov models. Energy Economics 32: 1507-1519.

Taskaya, T., Casey, M.C., (2005). A comparative study of autoregressive neural network hybrids. Neural Networks 18: 781-789.

Tehrani, R., Khodayar, F., (2011). A hybrid optimized artificial intelligent model to forecast crude oil using genetic algorithm. African Journal of Business Management 5(34): 13130-13135.

Tektaş, M., (2010). Weather Forecasting Using ANFIS and ARIMA MODELS; A Case Study for Istanbul. Environmental Research, Engineering and Management 1(51): 5-10.

Tseng, F.M., Yu, H.Ch., Tzeng, G.H., (2002). Combining neural network model with seasonal time series ARIMA model, Technological Forecasting & Social Change 69: 71-87.

Vairappan, C., Tamura, H., Gao, Sh., Tang, Zh., (2009). Batch type local search-based adaptive neuro-fuzzy inference system (ANFIS) with self-feedbacks for time-series prediction. Neurocomputing 72: 1870-1877.

Valenzuela, O., Rojas, I., Rojas, F., Pomares, H., Herrera, L.J., Guillen, A., Marquez, L., Pasadas, M., (2008). Hybridization of intelligent techniques and ARIMA models for time series prediction. Fuzzy Sets and Systems 159: 821-845.

Wang, F.K., Chang, K.K., Tzeng, Ch.W., (2011). Using adaptive network-based fuzzy inference system to forecast automobile sales. Expert Systems with Applications 38: 10587-10593.

Wei, Y., Wang, Y., Huang, D., (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. Energy Economics 32: 1477-1484.

Yu, L., Wang, Sh., Lai, K. K., (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Economics 30: 2623-2635.

Zhang, G. P., (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50: 159-175.

Zhang, X., Lai, K.K., Wang, Sh. Y., (2008). A new approach for crude oil price analysis based on Empirical Mode Decomposition. Energy Economics 30: 905–918.

Zhu, K., Ling, Sh., (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA–GARCH/IGARCH models. The Annals of Statistics 39(4): 2131-2163.

Zou, H.F., Xia, G.P., Yang, F.T., Wang, H.Y., (2007). An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. Neurocomputing 70: 2913-2923.