Indian Journal of Science and Technology, *Vol 10(31), DOI: 10.17485/ijst/2017/v10i31/117465, August 2017*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

A Hybrid GA-FEEMD for Forecasting Crude Oil Prices

Nuramirah Akrom and Zuhaimy Ismail*

Faculty of Sciences, Department of Mathematical Sciences, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia; nuramirah2@live.utm.my, zuhaimyutm@gmail.com

Abstract

Forecasting crude oil prices is essential but usually involves a difficult process. In this paper, we proposed a hybrid Genetic Algorithm and Fast Ensemble Empirical Mode Decomposition (GA-FEEMD) for forecasting crude oil price time series data. The proposed GA-FEEMD basically involves three steps. Firstly, we decomposed the original crude oil price time series data into two Intrinsic Mode Functions (IMFs) using FEEMD algorithm. Then, we forecasted the second IMF which basically is the intrinsic trend of the crude oil prices. Then, we applied Genetic Algorithm (GA) to obtain the stopping criterion in the FEEMD process. The hybrid GA-FEEMD forecasting model was compared with ARIMA and Artificial Neural Network methods. The results showed that the proposed GA-FEEMD model improved the forecasting accuracy of the crude oil price time series data.

Keywords: Crude Oil Price Forecast, Fast Ensemble Empirical Mode Decomposition (FEEMD), Genetic Algorithm (GA) and Hybridization

1. Introduction

Crude oil has become one of the powerful exchanged assets in the world[1]. Crude oil price trend and pattern are among the factors that affect the fluctuation of a country's economy and influence the assets of the country[2]. Recent decades have seen more instability in the crude oil price trend, thus compelling the government and market participants to forecast crude oil prices[2]. Accurate crude oil price predicting is one of the key to strategic planning by the government in ensuring economic stability of the country[3].

The intrinsic instability of crude oil price trend and pattern necessitates a huge effort from various scholars in predicting crude oil price time series data[4]. Previous literatures have shown many efforts dedicated in improving the forecasting accuracy of crude oil prices. Basically there are three major methods for crude oil price forecasting. The first method can be referred to as the traditional statistical methods such as ARIMA and GARCH models[5], Bayesian Vector Autoregressive model[6], Qualitative Vector Autoregressive model[7], and Grey Wave Forecasting method[8]. For example, Mohammadi and Su[5] employed ARIMA-GARCH models for modeling and forecasting the conditional mean and volatility of weekly crude oil spot prices for eleven international markets; Gupta and Kotze[6] proposed the Bayesian vector autoregressive approach for forecasting real oil price data for South Africa; Gupta and Wohar[7] applied Qualitative Vector Autoregressive (QualVar) to forecast West Texas Intermediate (WTI) monthly oil and (S&P500) stock returns over the period of September 1884 to August 2015; and Chen et al[8] suggested grey wave forecasting method in conducting multistep-ahead forecasting for daily crude oil price. However, there are a lot of drawbacks to the traditional statistical methods, one of which is it cannot capture crude oil price data in a nonlinear pattern.

The second method, consisting of Artificial Intelligence (AI) methods such as Artificial Neural Network (ANN)[9] and Support Vector Machines[10], was proposed to improve the traditional statistical models for predicting crude oil prices. Movagharnejad et al[9] applied a neural network model for forecasting different commercial oil prices in the Persian Gulf region. Xie et al[10] implemented Support Vector Machine method for forecasting WTI oil from January 1970 to December 2003. Gao and Lei[11] also investigated changes in the pattern of oil prices using stream

**Author for correspondence*

learning forecasting model. Nevertheless, the disadvantages of AI methods should be taken into account as well, such as their slow local minima and convergence in the forecasting procedures.

The last method, i.e., empirical-decomposed and ensemble-decomposed methods, was famously introduced for predicting crude oil price data. Xiong et al[12] hybridized Empirical Mode Decomposition (EMD), Feed-Forward Neural Network (FNN), and Slope-based Method (SBM) in capturing the complex dynamic of crude oil prices' predicting process. Yu et al[13] utilized the ensemble-decomposed method with data-characteristic-driven reconstruction for crude oil price forecasting. Yu et al[14] used ensemble-decomposed and Extended Extreme Learning Machine (EELM) in predicting the complex pattern of crude oil prices. Just like the first two methods, the empirical-decomposed and ensemble-decomposed methods also have their drawbacks in that they are time-consuming, complex, and costly, where each Intrinsic Mode Function (IMF) from the empirical-decomposed and ensemble-decomposed algorithm requires another forecasting model to be built.

In our work, we propose a hybridization of Genetic Algorithm (GA) and Fast Ensemble Empirical Mode Decomposition (FEEMD) for forecasting crude oil prices. Basically, forecasting crude oil prices with the GA-FEEMD model involves three main steps. Firstly, the crude oil price data sets are decomposed into two IMFs using FEEMD algorithm. Then the second IMF is assumed as the intrinsic trend for the crude oil price time series data. Lastly, GA is implemented as the stopping criterion for FEEMD, and the Mean Absolute Percentage Error (MAPE) is the objective function in GA. Our proposed GA-FEEMD model is also compared with ARIMA and ANN models to estimate the accurateness of the forecasting performance of the proposed model.

The rest of this paper is organized as follows: section 2 provides a brief account of the methodology of ARIMA, ANN, and GA-FEEMD models; section 3 presents the empirical analysis and results of the forecasting crude oil price data; and section 4 concludes the findings of the study.

2. Methodology

2.1 ARIMA Model

Basically, an ARIMA (p,d,q) model can be expressed as[15]:

$$\phi(B)(1-B)^d Y_t = \mu + \theta(B)a_t \tag{1}$$

where

$$\phi(B) = 1-\phi_1 B-\phi_2 B^2-\phi_p B^p$$

$$\theta(B) = 1-\theta_1 B-\theta_2 B^2-\theta_q B^q$$

Here, Y_t and a_t are the actual observation and white noise series at time t, respectively, and B is the backshift of the lag operator; p and q are the model orders; ϕ and θ are the parameter values for AR and MA models; d represents the degree of ordinary differencing; and μ is the constant value.

2.2 ANN model

Mathematically, an ANN algorithm can be written as[16]:

$$y(x) = F(\sum_{i=1}^{L} w_i(t).x_i(t)+b) \tag{2}$$

where $x_i(t)$ is the predictor input variable in time t, $y(x)$ is the forecasted data set, L is the hidden neuron, $w\ (t)$ is the weight that connects the i th neuron in the input layer to the output layer j th, b is the neuronal bias, and $F(.)$ is the hidden transfer function.

2.3 Empirical Mode Decomposition (EMD)

As an adaptive data decomposition method, the EMD can decompose any time series data into several simple modes, namely IMFs and a residue[17]. The EMD algorithm behaves as follows:

1. All local maximum and local minimum of the data series, $x(t)$, are determined.
2. All the local maximum and local minimum are connected by using cubic spline interpolation in forming envelops $e_{\max}(t)$ and $e_{\min}(t)$.
3. The mean envelop, $m(t)$, is obtained where:

$$m(t) = \frac{e_{\max}(t)+e_{\min}(t)}{2} \tag{3}$$

4. The difference, $d(t)$, between $x(t)$ and $m(t)$ is calculated, where:

$$d(t) = x(t) - m(t) \tag{4}$$

5. $d(t)$ is judged whether it satisfies the IMF conditions. If it does, it is regarded as the first IMF; otherwise, it is considered as the original sequence. After obtaining the mean envelop, $m_{11}(t)$, whether or not $d_{11}(t)$ meets the conditions of IMF is determined. If it cannot meet the conditions, the cycle is repeated k times until $d_{1k}(t)$ does meet the IMFs conditions. In case $c_1(t) = d_{ik}(t)$, $c_1(t)$ is the first IMF of $x(t)$.
6. After separating $c_1(t)$ from $x(t)$, the residue $r_1(t) = x(t) - c_1(t)$ is regarded as the original series. Then, steps 1-5 are repeated so that m IMFs and one residue $r(t)$ can be obtained after meeting the stopping criterion of standard deviation [0.2, 0.3].

However, EMD has one serious drawback, which is the mode-mixing problem. Wu and Huang[18], therefore, proposed Ensemble Empirical Mode Decomposition (EEMD). The EEMD algorithm is as follows:

1. A number of Gaussian white noises, $n_i(t)$, are introduced into the data series, $x(t)$, $n_i(t) \sim N(0, \sigma^2)$:

$$x_i(t) = x(t) + n_i(t) \tag{5}$$

2. EMD is conducted on each $x_i(t)$, obtaining a set of the IMFs' $c_{ij}(t)$ and a residue $r_i(t)$, where $c_{ij}(t)$ is the j th IMF decomposed by EMD after adding Gaussian white noise for the i th time.
3. The above-mentioned steps are repeated. The ensemble average of the corresponding IMFs is considered the final decomposition result:

$$c_j(t) = \frac{1}{N} \sum_{i=1}^{N} c_{ij}(t) \tag{6}$$

2.4 GA-FEEMD Model

The computational process involved in the GA-FEEMD model consists of the following steps:

1. The steps in the previous EEMD algorithm are followed, but the m IMFs in Step 6 of the EMD algorithm is targeted as 1 and with one residue, $r(t)$.
2. Once the targeted $IMF = 2$, the second IMF acts as the intrinsic trend of the prediction series.
3. The stopping criterion of the standard deviation is calculated using GA. The stopping criterion for the IMFs and residue will vary and may no longer be [0.2, 0.3].
4. The MAPE is set as the objective function of GA. The MAPE can be written as:

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{A_t - F_t}{A_t} \right| \times 100 \tag{7}$$

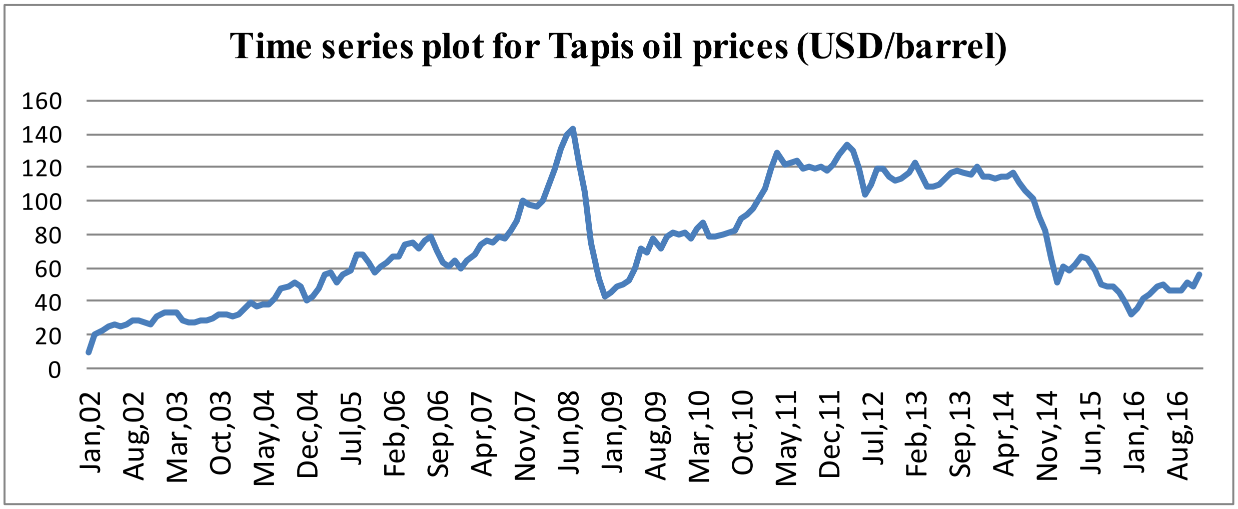

Figure 1. Tapis crude oil prices from January 2002 to December 2016.

where A_t is the actual data, F_t is the forecasted data, and N is the number of samples in the data series.

3. Empirical Results

3.1 Data

This study employed monthly Tapis crude oil price data which were provided by the Ministry of Finance Malaysia (MOF) up till the recent months. The data sets span from January 2002 to December 2016. Figure 1 indicates that the Tapis crude oil price fluctuated frequently and the fluctuation ranges were irregular. From the plot, the crude oil price gradually increased to 104.30 USD per barrel from January 2002 to September 2008 and then suddenly dropped to 74.90 USD per barrel the next month. Then from October 2008 to December 2016, the crude oil price pattern gradually climbed up and then fell back in October 2014.

3.2 Forecasting Modeling

This section explains the forecasting modeling of the Tapis crude oil prices using ARIMA, ANN and GA-FEEMD models. The data were initially divided into two groups, which are in-sample data and out sample data. The in-sample data contained 168 observations, which were from January 2002 to December 2012, while the out-sample data contained 12 observations, which were from January 2016 to December 2016. For the already-established ARIMA model, the data was differenced once $(d = 1)$ only since the level of crude oil prices is nonstationary. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were used to determine the model orders p and q. As a result, ARIMA (1,1,0) was proved to be the most proper model specification for this crude oil price data set.

A Multilayer Preceptor (MLP)-with three layers-feedforward network was developed to predict the crude oil price time series data. We carried out a trial-and-error process for choosing the best Transfer Function (TF) for the hidden layer and output layer of neurons. The best TF for the hidden layer was hyperbolic tangent sigmoid function while linear function was the best for the output layer of neurons. The best network structure for the MLP model was [2-1-1], with the number of weights totaling 5.

For the GA-FEEMD modeling process, we used Matlab 2010 to run GA and FEEMD algorithms. Figure 2 shows IMF 1 and IMF 2 for the crude oil price data. As can be seen from the figure, the first IMF gave unstable frequencies in the crude oil prices. The second IMF, meanwhile, exhibited a smooth intrinsic trend in the crude oil price data series and hence was used for predicting the intrinsic trend of the actual crude oil prices while the first IMF was ignored. Most previous researches stated that building prediction models for IMFs is time-consuming. When we applied the second intrinsic trend for forecasting the actual data, the steps for obtaining the prediction of the intrinsic trend was minimized since we need not build another model for both IMFs. The stopping criterion that gave the minimal value for MAPE was 0.64.

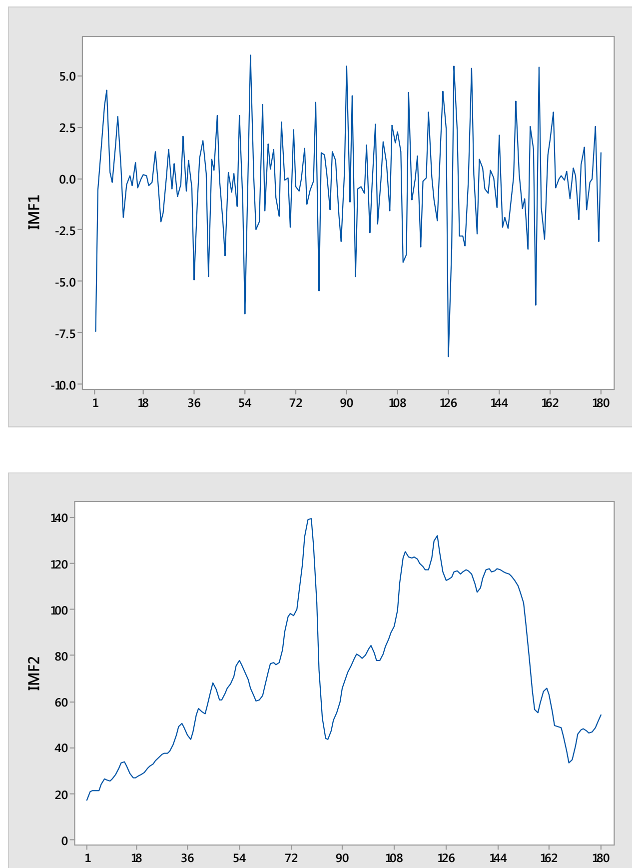

Figure 2. IMFs for crude oil prices.

3.3 Forecasting Results

This section compares the forecasting performances of ARIMA, ANN and GA-FEEMD methods. The results are shown in Table 1. For the in-sample forecasts, the hybrid GA-FEEMD gave the best forecasting performances as compared to ARIMA and ANN methods. For the out-sample forecasts, the GA-FEEMD model, again, gave the best forecasting performance with ARIMA being the second best, followed closely by ANN model. The GA-FEEMD model gave the best forecasting performance because it can smooth the intrinsic trend for the crude oil prices data, thus making the forecasting process easier. Figure 3 demonstrates the time series plot for the in-sample and out-sample forecasts of the crude oil prices using ARIMA, ANN and GA-FEEMD models.

Table 1. Forecasting performances.

Method	In-sample 168 observations (MAPE %)	Out-sample 12 observations (MAPE %)
ARIMA	6.2253	8.1193
ANN	5.6095	8.3747
GA-FEEMD	2.9989	2.5861

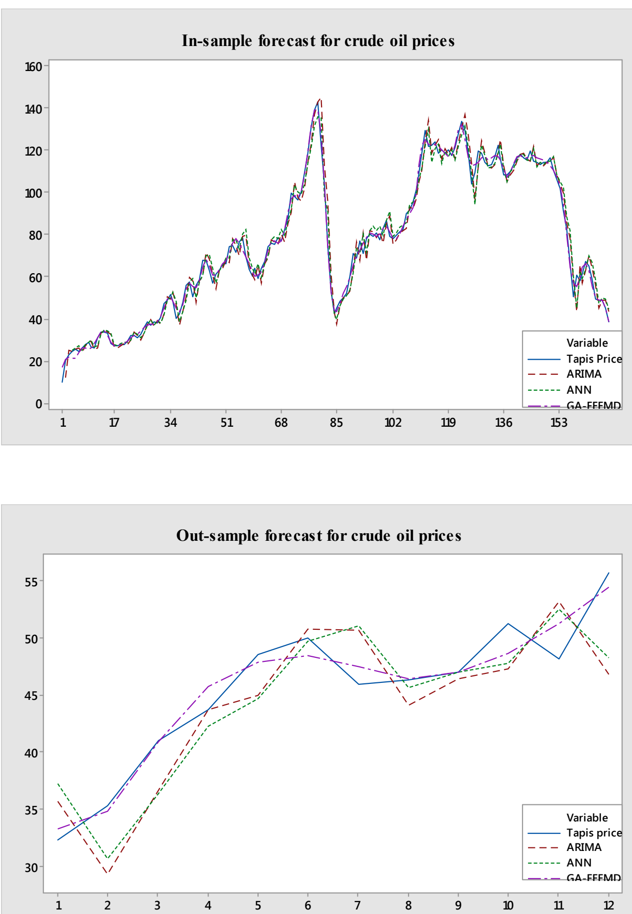

Figure 3. In-sample and out-sample forecasts for crude oil prices.

4. Conclusion

This paper has proposed a hybrid GA-FEEMD method for forecasting crude oil price data. Through the empirical analysis of forecasting the crude oil prices using GA-FEEMD, the results indicated that GA-FEEMD's forecasting performance was superior to ARIMA and ANN methods. This method also minimized the forecasting process in which the second IMF functioned to smooth the intrinsic trend of the crude oil price data. As such, we would recommend forecasting crude oil price data using this GA-FEEMD model.

5. Acknowledgement

This work is supported by UTM Grant Tier 1 (11H89) and MyBrain15 scholarship.

6. References

1. Alvarez-Ramirez J, Rodriguez E, Martina E and Ibarra-Valdez C. Cyclical behavior of crude oil markets and economic recessions in the period 1986–2010. Technological Forecasting Social Change. 2012; 79:47–58. Crossref
2. Zhang YJ. Speculative trading and WTI crude oil futures price movement: an empirical analysis. Applied Energy. 2013; 107:394–402. Crossref
3. Zhang YJ and Wang ZY. Investigating the price discovery and risk transfer functions in the crude oil and gasoline futures markets: some empirical evidence. Applied Energy. 104:220–8. Crossref
4. Bekirosa S, Gupta R and Paccagnini A. Oil price forecasting and economic uncertainty. Economics. Letters. 2015; 132: 125–8. Crossref
5. Mohammadi H and Su L. International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. Energy Economics. 2010; 32(5):1001–8. Crossref
6. Gupta R and Kotze K. The role of oil prices in the forecasts of South African interest rates: A Bayesian approach. Energy Economics. 20174; 61:270–8. Crossref
7. Gupta R and Wohar M. Forecasting oil and stock returns with a Qual VAR using over 150 years off data. Energy Economics. 2017; 62:181–6. Crossref
8. Chen Y, Zou Y, Zhou Y and Zhang C. Multi-step-ahead crude oil price forecasting based on Grey Wave forecasting method. Procedia Computer Science. 2016; 91:1050–6. Crossref
9. Movagharnejad K, Mehdizadeh B, Banihashemi M and Kordkheili MS. Forecasting the differences between various

commercial oil prices in the Persian Gulf region by neural network. Energy. 2011; 36(7):3979–84. Crossref

10. Xie W, Yu L, Xu S. and Wang S. A new method for crude oil price forecasting based on Support Vector Machines. Springer Berlin Heidelberg. 2006; 3996:444–51. Crossref
11. Gao S and Lei Y.A new approach for crude oil price prediction based on stream learning. Geoscience Frontiers.2017; 8(1):183–7. Crossref
12. Xiong T, Bao Y and Hu Z. Beyond one-step-ahead forecasting: Evaluation of alternative multi-step-ahead forecasting models for crude oil prices. Energy Economics. 2013; 40:405–15. Crossref
13. Yu L, Wang Z and Tang L. A decomposition-ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. Applied Energy. 2015; 156:251–67. Crossref
14. Yu L, Dai W and Tang L. A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting. Engineering Applications of Artificial Intelligence. 2016; 47:110–21. Crossref
15. Box GEP and Jenkins GM. Time Series Analysis, Forecasting and Control. 3rd Ed. Holden-Day:California; 1970.
16. Keskin ME and Terzi O. Artificial neural network models of daily pan evaporation. Journal Hydrology Engineering. 2006; 11:65–70. Crossref
17. Huang NE, Shen Z and Long SR. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. Proceedings of the Royal Society of London. 1998; 454:903–95. Crossref
18. Wu Z and Huang NE. Ensemble empirical mode decomposition: A noise-assisted data analysis method. Advances in Adaptive Data Analysis. 2009; 1(1):1–41. Crossref