Energy 154 (2018) 328–336

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

# A novel decompose-ensemble methodology with AIC-ANN approach for crude oil forecasting

Yishan Ding

*Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, 200030, China*

ARTICLE INFO

*Article history:*
Received 27 April 2017
Received in revised form
18 April 2018
Accepted 23 April 2018
Available online 26 April 2018

*Keywords:*
Crude oil price prediction
Ensemble empirical mode decomposition
Akaike's information criterion
Hybrid model
Predicting accuracy
Stability

ABSTRACT

Forecasting international crude oil is a well-known issue. The hybrid modeling principle tells us that combining different methods could take full advantage of all the merits and leave out the shortcomings. Therefore, hybrid methodology has been widely used in current research. In this study, a novel decompose-ensemble prediction process combining the ensemble empirical mode decomposition (EEMD) and artificial neural network (ANN) is proposed. Moreover, this method, i.e., EEMD-ANN-ADD method, adds the decompose-ensemble to the single AI model to further improve the predicting accuracy. The overall process can be divided into four steps: model selection via Akaike's information criterion (AIC), data decomposition via EEMD, individual prediction via ANN and ensemble prediction through addition ensemble method. To verify the results, we use the official data of oil price to conduct the predicting. The result confirms that "decompose-ensemble" models are better than the normal hybrid one, in terms of prediction accuracy (both level and directional measurement) and modified Diebold-Mariano test. What's more, back to the decompose-ensemble models, the EEMD-based one outperforms the empirical mode decomposition (EMD) one. At last but not the least, AIC gives us reasonable and convincing statement about determining the value of lag. Generally speaking, this novel forecasting technique is a prominent insight for the price of crude oil.



## 1. Introduction

According to some recent published articles [1,2], energy market activities, especially the condition of international crude oil have significant effects on the world economic activities. The price of international crude oil fluctuates widely and it will directly affect the economy [3]. The oil price movements might result in inflation, economic depression or even political turmoil. Intuitively, we can draw the conclusion that predicting the crude oil price might become a vital topic in the academic field, which is in accordance with the actual condition. However, an abundance of literature has fully proved that the forecasting of crude oil is an extremely tough and challenge task [4]. It mainly because the affecting factors are various, ranging from economy, politics to sudden events. Both of the above mentioned effects are sometimes hard to measure. For these reasons, research on prediction of oil price is not meaningful but challenging tasks, the discussion about oil price attracts much attentions [5]. The forecasting methods includes the statistical and econometric models, artificial intelligence and hybrid modeling. Hybrid modeling, especially the "decompose-ensemble" method remarkably improves the prediction accuracy. However the original decomposition method itself has its drawbacks, such as mode mixing which would weaken the prediction accuracy. Moreover, researchers ignore to give reasonable statement about how to choose the proper model of crude oil price forecasting.

The objective of this paper is to build an effective decompose-ensemble procedure to forecast the oil price, i.e., AIC-based EEMD-ANN-ADD method. Based on some benchmark models and performance evaluation criteria, we prove the effectiveness and superiority of this method. To better understand the research progress of this problem, Section 2 thoroughly reviews and compares different methods. In Section 3, we execute empirical studies and proves that this novel method are better than the other ones. The materials, process, criteria and discussion are presented in this part. In the end, Section 4 covers the conclusions, limitations and future extensions.

## 2. Literature review

Crude oil is a kind of indispensable energy source, chemical

*E-mail address:* dys90@163.com.

https://doi.org/10.1016/j.energy.2018.04.133
0360-5442/

materials and strategic resource in economic in a country's economic development. The price level related tightly to a country's economic development level, social stability and even political orderliness. The forecasting of international crude oil price is a famous widely discussed issue and there are abundant literature discussing around this topic. The existing method can be classified into three groups: traditional econometric, artificial intelligence (AI) and hybrid techniques [6].

The traditional econometric and statistic models cover many familiar models: Generalized Auto Regressive Conditional Heteroskedasticity (GARCH) family [7–9], Autoregressive Integrated Moving Average Model (ARIMA) [10], co-integration analysis [11], vector auto-regression (VAR) [12], error correction models (ECM) [13] and other methods, like linear regression (LinR), naive random walk (RW) [14], grey model (GM) [15]. Related forecasting literature by means of econometric and statistic models are summarized in Table 1.

The methods listed are based on strict linear assumptions and the price series must be linear or near linear. In fact, the real data is always nonlinear and irregular, under which condition those techniques could not effectively capture the hidden nonlinear features [16]. That's why AI tools were proposed. The typical artificial intelligence tools are listed in Table 2 which include: Artificial neural networks (ANN) [17–21], support vector regression (SVR) [22–24], least squares support vector regression (LSSVR) [25] and other methods like genetic programming (GP) [26] and belief network (BN) [27]. Although the AI models could effectively pick out the nonlinear patterns in the time series, they still have their own shortcomings such as parameter sensitiveness, over-fitting and local minimum [16].

With the development of AI tools, the traditional econometric models, with their own shortcomings, still play irreplaceable role in analyzing the price condition in energy market [28]. Wang et al. [29] proposed model selection with parameter restricted model, which has been proved to improve the predictive accuracy and enrich the policy implication. Baumeister et al. incorporated more inputs into the predictive regression [3,30–34]. Moreover, in recent years other approaches have appeared in this context, including long memory and fractional integration approaches [35–37].

To overcome the disadvantages of single traditional statistical and AI tools, the concept of "hybrid modeling" was proposed. It tells us that a hybrid model gathers all the models' advantages and offsets the drawbacks, which leads to the widely use of hybrid models in the crude oil price prediction. Wang et al. [38] hybrid ANN and rule-based expert system, with web text mining, to forecast the oil price. Amin-Naseri and Gharacheh [39]. combined feed-forward neural networks (FNN), genetic algorithm, and k-means clustering, to predict the crude oil price and obtained great results. Chiroma et al. [26] combined genetic algorithm and neural network to predict the WTI crude oil price. Baruník and Malinská [40] hybridized regression method with ANN to analyze the term structure of oil price and get the lowest errors. In particular, the interactive inner factors would result in some noises which might corrupt the true oil price data and weaken the prediction accuracy. Inspired by integrating denoised methods into AI models to solve economic problems [41–47], Yu et al. [4] hybrid the denoising procedure, ie. compressed sensing based denoising (CSD), with AI models to forecast the crude oil price. Logically and thoroughly, they analyze the effectiveness, superiority and robustness of this novel model. The results of these studies have showed that the hybrid models are better than their respective single models, and intuitively confirmed that the hybrid models are effective.

Particularly, "decompose-ensemble" principle has been considered as an alternative helpful hybrid tool for analyzing the crude oil data which is highly complex and irregular [16,48]. Jammazi and Aloui [49]. combined multilayer back propagation neural network and wavelet decomposition to forecast the price of crude oil. Yu et al. [16] suggested a novel FNN-based decompose-ensemble prediction method for crude oil. He et al. [50] using wavelet-based decompose-ensemble methodology to analyze and predict the price of crude oil. Tang et al. [51] coupled the complementary EEMD and extended extreme learning machine (EELM) to research on the prediction of crude oil price. Yu et al. [6] proposed a novel data-characteristic-driven ensemble method to predict the price of international crude oil. Tang et al. [52] and Zhu et al. [53] integrated EEMD and LSSVR to predict complex price series with high volatility and irregularity. Similarly, Wang et al. [54]developed a two-layer decomposition hybrid model which combined EEMD, firefly algorithm and BP neural network to predict time series.Yang et al. [55] combined the wavelet transform, the kernel extreme learning machine and auto-regressive moving average to predict the electricity price fluctuation. As a specific occasion of hybrid modeling, the methodology of "decompose-ensemble" can remarkably improve the prediction.

According to the literature review, EEMD-based AI methods are very popular in recent years. However, there are many knowledge gaps waiting for us to fill. For example, only a few scholars discussed how to choose the proper model and pick up the most relative input time series [4]. It would be an improvement if we can apply a classical model selection method to the international crude

**Table 1**
Typical literature using the statistic and economic models to forecast crude oil prices.

| Typical literature | Forecasting models | Main results |
|---|---|---|
| Hou and Suardi [7] | Non-parametric GARCH model | The proposed GARCH model outperforms the traditional GARCH model. |
| Li et al. [8] | Component GARCH | Their forecasting model has better predictive accuracy than the GARCH and ARMA models. |
| Morana [9] | semiparametric GARCH model | The forecasting approach can be used to obtain a performance measure for the forward price, in addition to compute interval forecasts for the oil price. |
| Xiang and Zhuang [10] | ARIMA model | Model ARIMA (1,1,1) possessed good prediction effect and can be used as short-term prediction of International crude oil price. |
| Gülen [11] | Cointegertion analysis | Cointegertion analysis was used to confirm the issue of "simple efficiency", which states that the futures price is an unbiased predictor of the spot price, in the case of trading in crude oil futures at NYMEX. |
| Mirmirani and Li [12] | VAR and ANN techniques | This study applies VAR and ANN techniques to make ex-post forecast of U.S. oil price movements, and the analysis suggests that the BPN-GA model noticeably outperforms the VAR model. |
| Lanza et al. [13] | Error Correction model | The comparison of ECM with a naïve (short-run) model suggests that cointegration marginally improves static forecasts in EU. |
| Murat and Tokat [14] | Random Walk | The results showed that (a) both the crack spread futures and the crude oil futures outperformed the RWM; and (b) the crack spread futures are almost as good as the crude oil futures in predicting the movements in spot oil markets. |
| Lin [15] | Grey model | The results show that the model of GM (1,1) is suitable for crude oil prices forecast. |