A CEEMD-ARIMA-SVM Model with Structural Breaks to Forecast the Crude Oil Prices Linked with Extreme Events

Yuxiang Cheng
Peking University School of Economics
Jiayu Yi
Nanyang Technological University College of Humanities Arts and Social Sciences
Xiaoguang Yang
Academy of Mathematics and Systems Science Chinese Academy of Sciences: Chinese Academy of Sciences Academy of Mathematics and Systems Science
Kin keung Lai (mskklai@outlook.com)
Shaanxi Normal University International Business School https://orcid.org/0000-0003-0014-2095
Luis Seco
University of Toronto Department of Mathematics

Research Article

Keywords: Decision support, Machine learning, Volatility forecasting, Mathematical finance, Support vector machine

Posted Date: January 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1029907/v1

A CEEMD-ARIMA-SVM Model with Structural Breaks to Forecast the Crude Oil Prices Linked with Extreme Events

Yuxiang Cheng[a], Jiayu Yi[b], Xiaoguang Yang[c], Kin Keung Lai[d,*] and Luis Seco[e]

[a] *School of Economics, Peking University, Beijing, China;*
[b] *School of Social Science, Nanyang Technological University, Singapore;*
[c] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China;*
[d] *International Business School, Shaanxi Normal University, Xian, China;*
[e] *Department of Mathematics, University of Toronto, Toronto, ON, Canada*

[*] *Corresponding author: E-mail: mskklai@outlook.com*

ABSTRACT: This paper develops an integrated framework to forecast the volatility of crude oil prices by considering the impacts of extreme events (structural breaks). The impacts of extreme events are vital to improving prediction accuracy. Aiming to demonstrate the crude oil price fluctuation and the impacts of external events, this paper employs the Complementary Ensemble Empirical Mode Decomposition (CEEMD). It decomposes the crude oil price into some constituents at various frequencies to extract a market fluctuation, a shock from extreme events, and a long-term trend. The shock from extreme events is found to be the most crucial element in deciding the crude oil prices. Then we combine the Iterative Cumulative Sum of Squares (ICSS) test with the Chow test to get the structural breaks and analyze the extreme event impacts. Finally, this paper combines the structural breaks, the Autoregressive Integrated Moving Average model (ARIMA), and the Support Vector Machine (SVM) to make a forecast of the crude oil prices. The empirical process proves that the CEEMD-ARIMA-SVM model with structural breaks performs the best when compared with the other ARIMA-type models and SVM-type models. The framework offers an insightful view to help decision-makers and can be used in many areas.

Key words: Decision support, Machine learning, Volatility forecasting, Mathematical finance, Support vector machine

1. Introduction

Crude oil is called "industry blood", which shows the great importance and the high value of crude oil to industry. Nowadays, the demand for crude oil is becoming larger, which reflects the greater influence of its market demand on commodity prices. Wei (2019) found that oil supply had significant impacts on China's real import. Rafiq et al. (2009) demonstrated that the volatilities and the fluctuations of the oil prices had substantial impacts on macroeconomic particulars like unemployment. Besides, crude oil is a non-renewable energy resource, which means it owns the advantages of mature mining technology, low exploitation cost and little utilization difficulty against renewable energy resources. Thus, it is of necessity to investigate the fluctuation characteristics of the crude oil prices.

Also, the crude oil market is crucial to global financial markets. Recently, on March 9, 2020, oil prices ended down more than 10%. The huge impact caused the stock market to slump. The Dow Jones plunged over 2,000 points in the worst day since 2008. The trading on Wall was frozen on this day. The forecast of the crude oil price volatility is significantly important to evaluate global financial market stability.

As one of the traditional energy products, crude oil emphasizes its financial attributes and serves more financial purposes in the past few years. Crude oil makes an increasingly big difference in diversified investment portfolios. Xiao et al. (2018) and Hu et al. (2018) discovered that the crude oil price fluctuations exerted asymmetric impacts on the Chinese stock market. Chang et al. (2013) demonstrated the volatility spillovers between the returns of stock index and those of WTI and Brent crude oil futures, forward and spot. In the past few years, the increase of the crude oil price volatility has raised the crude oil investment risk and also changes the investment risks in other capital assets. To investigate the volatility features of the crude oil prices, especially the intrinsic trends and the external shocks to make accurate predictions of price movements is especially important to decrease investment risks.

The extreme events are often ignored when we forecast the volatility. It refers to significant changes of macroeconomic conditions, such as policy changes, natural catastrophes, and oil crises. For instance, the coronavirus happened and has existed for three months. More than 110,000 people

have already been infected by COVID-19 worldwide, which leads to the shutdowns of factories, schools, and stores and cancellations of journeys and other gatherings. Owing to this extreme event, the oil price sank nearly 20% after an escalating oil-market war was intensified by Saudi Arabia and Russia.

The crude oil prices had discontinuous jump variation and high volatility during some periods (Krichene 2006). Gupta and Yoon (2018) proved linear Granger causality tests were misspecified, yielding the unreliable linear oil futures market return results with non-predictability. Therefore, crude oil prices are proved to be non-linear and multi-scale time series. Affected by obvious financial attributes of the crude oil prices, the short-run trend deviates from the mid-run and long-run trends periodically. The mid-term trend is more vulnerable to significant events, keeping in accord with the long-term trend. Ruelke et al. (2011) found that the mid-run prediction is consistent with neither short-term nor long-term forecasts. The crude oil prices consist of the fluctuations of different durations and have all of their characteristics. From macroeconomics theories, crude oil reserves and production play the key roles in determining the long-term price trend. Zhang et al. (2008) found that the long-term trend was the most crucial one among the determinants of the crude oil prices. However, this paper finds as the crude oil financial attribute appears more obvious and the Shale Oil Revolution develops further, the mid-term trend affected by extreme events has become the most important factor determining the WTI spot crude oil prices. Therefore, it is sustantially significant to investigate the future price in the short run, medium run and long run separately when analyzing and predicting the crude oil price volatilities.

This paper predicts the prices of crude oil and also concentrates on the impacts of significant events. Oladosu (2009) found that regardless of whether exogenous events led to actual supply loss, they could affect the short-term and mid-term prices. Martina et al. (2011), Coleman (2012) and Ji and Guo (2015) figured out how macroeconomic events impacted on the crude oil market. Zhang et al. (2009) used EMD to find that, the Gulf War, the Iranian War and the OPEC Oil Output Reduction Agreements rapidly increased crude oil prices. Martina et al. (2011) found that significant events changed the complex structure of the short-term crude oil market but do not change that of the long-term market. To some extent, whether significant events affect price trends and how long the effect lasts depend on the trend length. Therefore, significant events would affect the short-run and mid-

run trends of the crude oil prices respectively. However, few works of literatures have analyzed the impacts of significant events on the crude oil price volatilities in a quantitative view and took them as dummy variables of models. This paper contributes to the literature concerning the significant event impacts on the crude oil prices and increasing prediction accuracy.

The most similar work in the existing literature compared to our study is by Yi et al.(2016). They established an EMD-VAR-SVR model to predict the executive amount of servicing outsourcing and added ARIMA models to analyze nonlinear time series. Nevertheless, the impacts of structural breaks are neglected. Besides, the white noise could be mitigated by EMD, which is not helpful for improving forecasting results. In our work, we consider the impact of structural and CEEMD in volatility forecasting and fill the existing research gap.

We propose an integrated framework to predict the volatilities of the crude oil prices, which consists of CEEMD, ARIMA, and Support Vector Machine (SVM). CEEMD aims to decompose price into different trends and mitigate the heterogeneous of crude oil price, which could increase prediction accuracy. The results show that the shock from extreme events makes the biggest difference in determining the prices of crude oil. ARIMA and SVM help forecast volatility. We also consider the extreme external events (structural breaks) into the models. In our work, we find that the Support Vector Machine (SVM) algorithm performs better for the prediction of the crude oil price series than the ARIMA model. We also compare different SVM-type models and ARIMA models, some of where CEEMD is integrated into the models. Based on the results of loss function, we prove that the CEEMD-ARIMA-SVM model's forecasting accuracy is the best. Besides, when we add structural breaks into the integrated framework, the performance is better compared to original basic models. We show that taking the structural breaks of the crude oil prices can effectively increase prediction accuracy.

This paper's contributions are as follows: (1) Using CEEMD to decompose a market fluctuation and mitigate the white noise of forecasting. Thus, the forecasting result improves; (2) It can help investigate the influences of significant events on crude oil prices to combine the CEEMD method with the ICSS test and the Chow test, which not only prepares for the setting of the dummy variables, but also analyzes the impacts that significant events have on crude oil price fluctuation; (3) This is the first work to combine CEEMD, ARIMA, SVM, and impacts of structural breaks to

forecast volatility. (4) The integrated framework can offer an insightful view to be applicable in tourism demand forecasting, commodity futures market forecasting and other industry.

The remainder of this paper is structured as follows: Section 2 is the literature review; Section 3 is an introduction of the basic principles and the models utilized in this paper; Section 4 is the description of the data sample; Section 5 is the seven forecast model construction and the empirical analysis, which mainly includes the crude oil price decomposition, the test of structural breaks, the forecast results and the robustness test and Section 6 concludes this paper.

2. Literature Review

Our research touches upon two streams of literature in the energy area: mathematical finance models and methods to mitigate heterogeneity. For the former one, we introduce the basic concepts of volatility and discuss the crucial volatility models. For the latter one, we review many recent applications to decompose and optimize volatility forecasting results.

2.1 Mathematical Finance Models on Volatility Forecasting

(Herrera et al. 2018) and Luo and Qin (2017) found that volatility was taken as a popular uncertainty proxy. To predict the volatility of financial market indices and product prices, Engle(1982), Bollerslev (1986), and Taylor (1986) successively established models like ARCH, GARCH, and SV. Then SV-type models and GARCH-type models were gradually developed. Later, a lot of experts and scholars improved GARCH models to make predictions (Bekiros and Diks 2008; Wang et al. 2019; Catania and Proietti 2020; Martin et al. 2020).

(Zhang et al. 2020), Jacquier et al. (2004) and Vo (2009) made breakthroughs on SV-type models. Geng et al.(2017) established the PCA-FRBF model to forecast crude oil and it performs well. Traditional econometric models simulate and predict crude oil price fluctuations under the assumption that the fluctuations are linear and there is no structural break in them. However, the assumption is contrary to the truth, resulting in unsatisfactory results. Therefore, this paper adopts artificial intelligence models for prediction, owing to data volume increase, computing power promotion and new machine learning algorithm emergence. Movagharnejad et al. (2011) and Shabri and Samsudin (2014) predicted crude oil prices based on ANN models.

Although ANN models can analyze data effectively and reduce operation difficulty, there exist some shortcomings, including a long debugging period and slow convergence, which reduce the ANN model efficiency most. Yu et al.(2017) and (Risse 2019) believed that SVM models predicted better than traditional models and other artificial intelligence models. Zhao et al. (2015) established the ARIMA-SVM model by taking into account crude oil prices, market elements and non-market elements to forecast the crude oil prices more accurately. Also, the SVM models and the hybrid models based on SVMs perform well in other areas (D. D. Wu et al. 2014). Kaytez et al. (2015) and Das and Padhy (2018) used the hybrid models based on SVMs for power consumption and commodity futures index forecast, achieving better results than the SVM models. From Wen et al. (2017), it is known that if inconsistency and non-synchronization of intrinsic fluctuations of the prices are neglected, the prediction error will be expanded further. Thus, this paper considers the SVM model and the hybrid models based on the multi-scale decomposition and the setting of the dummy variables to improve the forecasting results.

2.2 Methods to Mitigate Heterogeneity

Heterogeneity is considerably common in various financial markets, such as the commodity futures markets, the crude oil markets and the stock markets. Wen et al. (2014), Wen et al. (2014), Gong et al. (2017) and He et al. (2018) found that investment risks have asymmetric impacts (heterogeneous) on crude oil yield, indicating that crude oil prices are multidimensional. Meanwhile, scholars tried to solve this problem by establishing an EMD-based algorithm according to different frequencies of crude oil price fluctuations. Since Huang et al. (1998) invented the Empirical Mode Decomposition (EMD) method, many scholars have applied EMD to analyze the influential factors of price fluctuations. He et al. (2016) used multivariate EMD models to make a prediction of the crude oil prices and depict the diverse and heterogeneous data. Zhang et al. (2008) decomposed the price series at different frequencies into a market fluctuation, a shock from extreme external events and a long-run trend and analyzed extreme event impacts. Nevertheless, the EMD algorithm generates more components than needed to bring unnecessary low-frequency information and is likely to cause modal mixing, which does no good for further analysis. Wu and Huang (2009) optimized the EMD

algorithm by adding white noise multiple times and eliminating model mixing to create EEMD. Nevertheless, the residual white noise could not be eliminated by EEMD, and thus it would increase decomposition error. Yeh et al.(2010) created CEEMD by adding or subtracting white noise to eliminate unnecessary noise as much as possible. Based on the above analysis, this paper adopts CEEMD to decompose the crude oil price series aim to mitigate the influences of white noise. When it comes to hybrid models, Yu et al.(2008), Zhang and Zhou (2013), and Zhu et al. (2018) respectively established EMD-ANN models, EEMD-ICDSVM models, and EMD-LSSVM models to predict the crude oil price volatility more accurately. To determine the specific time of significant events and analyze its impact on crude oil yield, Mensi et al. (2014), Wen et al. (2016) and Wen et al. (2018) used the ICSS test designed by Inclan and Tiao (1994) to recognize structural breaks based on EMD. The scholars above explained the influences of significant events on the crude oil prices. Nevertheless, they did not consider how those events impact the crude oil price volatility in the long run and short run. We try to capture the impacts of structural breaks on the crude oil prices in different situations and forecast its volatility precisely in this work.

The most similar work in the existing literature compared to our study is by Yi et al. (2016). They established an EMD-VAR-SVR model to predict the executive amount of servicing outsourcing and added ARIMA models to analyze nonlinear time series. However, the effects of extreme events on price volatility are ignored by them. Also, EMD could not mitigate the impact of white noise, which causes the impurities existing during prediction. Besides, there is little literature on the characteristics of the short-run, mid-run and long-run crude oil price trends. We utilize CEEMD to decompose the prices to mitigate the markets' noises and identify the different terms of crude oil prices. Moreover, the events happening at the structural breaks of the crude oil prices are identified by the ICSS test, which helps to improve forecasting accuracy. This paper considers the above questions and utilizes the ARIMA model to emphasize the impacts that the interaction has on the crude oil price to supplement the existing literature.

3. Methodology of Analysis

The specific processes of establishing the CEEMD-ARIMA-SVM model with structural breaks and the other prediction models are presented in this section. First of all, the CEEMD algorithm, the ICSS test, the Chow test and the Support Vector Machine are briefly reviewed. Then the overall steps of establishing the CEEMD-based Support Vector Machine learning methodology are summarized.

3.1 Complementary ensemble empirical mode decomposition (CEEMD)

Tang et al. (2015) found that compared with traditional algorithms, EMD was an intuitive, empirical, and adaptive algorithm, which fitted for non-stationary and nonlinear data. Based on the assumption that original data has different coexisting fluctuation modes, EMD is designed to separate these coexisting modes to analyze fluctuation characteristics of original data. Among the EMD-type algorithms, Yeh et al. (2010) invented CEEMD based on EMD and EEMD. This method can avoid modal mixing and long computing period. Then Wen et al. (2017) proved that CEEMD performed better than EMD when analyzing price fluctuations. According to Imaouchen et al. (2017) , CEEMD was more effeicient in computation than EEMD by reducing unnecessary white noise. To analyze the multi-scale price trends, this paper adopts CEEMD to decompose crude oil prices into various IMFs and analyze the characteristics of each IMF' s fluctuation.

What's more, IMFs must satisfy the following two characteristics: (1) the number of extreme values (i.e., both maximum value and minimum value) should equal to the number of zero crossings, or at most differs from that by one; (2) the functions must be symmetric in terms of local zero mean. For a perspective of time series, according to the conditions that its IMFs satisfy, CEEMD is established as follows:

(1) Add the white noise $w_i(t)$ k times to the primary time series $x(t)$, and subtract the white noise $w_i(t)$ k times from the primary time series $x(t)$, thus getting the two complementary signals $x_{+i}(t)$ and $x_{-i}(t)$.

$$\begin{bmatrix} x_{+i}(t) \\ x_{-i}(t) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x(t) \\ kw_i(t) \end{bmatrix} \tag{1}$$

(2) Decompose the two complementary signals $x_{+i}(t)$ and $x_{-i}(t)$ by using EEMD

proposed by Wu et al. (Zhang et al. 2008) and get the two parts of each IMF.

$$\begin{cases} \bar{c}_{+i} = \frac{\sum_{j=1}^{N} c_{j,+i}}{N} \\ \bar{c}_{-i} = \frac{\sum_{j=1}^{N} c_{j,-i}}{N} \end{cases} \tag{2}$$

(3) Then the IMFs without white noise are calculated.

$$\bar{c}_i = \frac{c_{j,+i} + c_{j,-i}}{2} \tag{3}$$

(4) Thus, the primary time series $x(t)$ is decompounded into IMFs and a residual constituent $r_n(t)$.

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{4}$$

$r_n(t)$ is a non-oscillating trend item representing the intrinsic regularities of the primary time series $x(t)$.

3.2 The ICSS Test

The iterative cumulative sum of squares algorithm (ICSS) established by Inclan and Tiao (1994) has wide applications in examining structural breaks (Li et al. 2020). The series is assumed to show a steady variance in the beginning until the variance and the level of the previous fluctuation change. Then the variance of the series keeps at the new level before another structural break happens. A time series usually repeats the above process and obtains unknown times of variance change. Because the price mutation causes the variance of the price series to change during a certain period, the ICSS test fits for detecting structural breaks. This paper uses ICSS to examine structural breaks of each component and determine the specific time of each extreme event.

If there is a time series $Y_k = \mu + \varepsilon_k$, where μ means the mean of an unknown constant and σ^2 presents the unknown constant variance of the series ε_k. With its mean of 0 and its variance of σ_k^2, the iterative residual sequence $\{\varepsilon_k\}$ is assumed to be a time sequence. The variance of the time sequence in each segment is $\sigma_i^2 (i = 1,2,\cdots,N_T)$, where N_T means the number of variance breaks for T observations. Meanwhile,$1 < K_1 < K_2 < \cdots < K_{N_T} < T$ is the combination of breaks.

To investigate the breaks of variance and each change duration, the cumulative square sum C_k is used.

$$C_k = \sum_{i=1}^{k} \varepsilon_t^2 , \mathrm{k} = 1, \cdots, \mathrm{T}. \tag{5}$$

The equation is used to calculate the cumulative sums of squares of the sequence observations from the start point to the Kth point. Estimate the statistic D_k as follows.

$$D_k = \frac{C_k}{C_T} + \frac{k}{T}, \mathrm{k} = 1, \cdots, \mathrm{T}, D_0 = D_T = 0. \tag{6}$$

The statistic D_k moves around zero, if there is no break in the variance for the observations. But if there is one or more breaks, the value will be significantly different from zero. k^* is defined as the value of k at $\max_k |D_k|$. And if $\max_k \sqrt{(T/2)} |D_k|$ goes out of the range of the assumed confidence level, the k^* is seen as an assumed break and $T/2$ is seen as the standardised factor.

3.3 The Chow Test

The Chow test is a method for discriminating whether a significant variance mutation has happened at a time point set before. The time sequence is assumed to be decompounded into two parts, and then the F test is employed to examine if the structure undergoes a significant mutation according to if the parameters calculated from the previous part of the data are the same as the parameters calculated from the later part of the data. Ho and Huang (2015), and Kekolahti et al. (2016) used the Chow test to calculate the structural breaks of stock indexes, mobile phone sales and the gold price. The Chow test is utilized to determine if the significant events calculated by the ICS Krichene S test have significant impacts on the crude oil prices.

Concerning the preset time point for one specific structural break, the time sequence is decomposed into two parts with n_1 and n_2 values. And this paper utilizes the ARIMA model to establish two regression models with error terms being $\mu_i' \sim \mathrm{N}(0, \sigma^2)$, $\mu_i'' \sim \mathrm{N}(0, \sigma^2)$. And the errors are uncorrelated with each other. First, calculate the sum of squared residual errors SSE of the overall time sequence, and its degree of freedom is $n_1 + n_2 - m - 1$; then sum up the squared residual errors SSE_1 and SSE_2 of the above two parts respectively. And

their degrees of freedom are $n_1 - m - 1$ and $n_2 - m - 1$. Considering μ_i' and μ_i'' are not correlated with each other, the statistic F is defined as

$$F = \frac{\frac{SSE - SSE_1 - SSE_2}{m+1}}{\frac{SSE_1 + SSE_2}{n_1 + n_2 - 2m - 2}} \tag{7}$$

And it's obvious that $F \sim F(m+1, n_1 + n_2 - 2m - 2)$. Given a confidence level α, if the calculation result of the F test does not equal to the threshold value, it indicates that the statistical regression model is structurally unstable, which proves the preset time point represents a structural break.

In this paper, the structural break is preliminarily determined by the ICSS test, setting a time interval of 90 days before and after the assumed time point as the period where the extreme event happens, and the final structural break time is determined by the Chow test.

3.4 Support Vector Machine (SVM)

Vapnik (1995) established the SVM model for price forecast. The model is based on the Vapnik-Chervonenkis Dimension theory and the Structural Risk Minimization principle. The SVM model enjoys a lot of advantages, including simple construction, global optimization, strong generalization and fast learning. It can solve problems of high dimension, nonlinearity, small samples and local optimum. Yu et al. (2017) and Ren et al. (2019) believed that the SVM model had a better performance than the traditional models and other artificial intelligence models. Most artificial intelligence models have the shortcomings of a long debugging period and slow convergence, which decrease the efficiency of the ANN model most. Therefore, this paper prefers the SVM algorithm to predict crude oil prices. And this paper combines CEEMD, the ICSS test, the Chow test with SVM models for prediction.

The SVM mechanism is to find an optimal categorization of the hyperplane, making the blank margins maximized on both sides of the hyperplane to guarantee categorization accuracy. Take the two-type data classification as an example. Given the training set (x_i, y_i), $i = 1,2,\cdots,l$, $x \in R^n$, $y \in \{\pm 1\}$, and the hyperplane notated as $(\omega x) + b = 0$, the classification hyperplane must satisfy the following constraint: $y_i[(\omega x_i) + b \geq 1]$, $i = 1,2,\cdots,l$.

The categorization interval can be expressed as $2/_{\|\omega\|}$, so the problem of the optimum hyperplane construction is turned into that of solving the problem as follows under the constraint:

$$min\,\varphi(\omega) = \frac{1}{2}\|\omega\|^2 = \frac{1}{2}(\omega'\omega) \tag{8}$$

To solve this optimization problem subject to the constraint, the Lagrange function is employed:

$$L(\omega, b, a) = \frac{1}{2}\|\omega\| - a\big(y\big((\omega \cdot x) + b\big) - 1\big) \tag{9}$$

where $a_i > 0$ and a_i is the Lagrange multiplier. The opimization problem's solution is decided by the Lagrange function's saddle point.

To make a balance between the two goals of least sample misclassification and maximum classification margin, the object function is turned into:

$$min\,\varphi(\omega, \xi) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\xi_i \tag{10}$$

C is a constant larger than 0. The equation is called the penalty function, which decides the degree of punishment for sample misclassification.

As for the nonlinear problem, it can be turned into a dual problem by a high-dimensional space set by the SVM model. This paper uses the RBF kernel function, which is shown as follows.

$$k(x, x_i) = e^{(-\gamma\|x - x_i\|)} \tag{11}$$

3.5 Crude Oil Price Prediction Process

The prediction model for crude oil prices is constructed as follows, shown in Fig.1.

Step 1: Use CEEMD to decompose the crude oil prices into some IMFs and one residual component.

Step 2: Extract the three new sequences, i.e., the short-run market fluctuation, the shock from extreme external events and the long-run trend by the method proposed by Zhang et al.

(2008).

Step 3: In terms of the three sub-sequences, use the ICSS test and the Chow test to determine the specific time and impact of each external event.

Step 4: Set the corresponding dummy variables according to the effect of each event. Establish ARIMA models and SVM models separately to identify the future changes of the market fluctuation, the shock from extreme events and the long-run price trend.

Step 5: Add the results of the ARIMA models or those of the SVM models into the hybrid models based on SVMs.

Step 6: Establish the CEEMD-SVM models and CEEMD-ARIMA-SVM models with or without structural breaks, and compare their forecast results with those of the models, including ARIMA, cluster ananlysis, SVM, and CEEMD-ARIMA.

Step 7: According to the final forecast results, this paper chooses the best prediction model.

[Please insert Figure 1 about here]

4. Data and Summary Statistics

The validity of the proposed models is tested using the spot FOB price of WTI, which is acquired from the U.S. Energy Information Administration (EIA). In 2004, the WTI spot FOB price exceeded $40, fluctuating more and more violently, which was not in accord with the previous price trends, so many scholars used January 2004 as the starting point of the data set (Han et al. 2017). Therefore, after deleting days with excessively few transactions or shortened trading sessions, the daily data sample period spans from January 5, 2004 to December 31, 2019 and 4019 daily observations are obtained.

[Please insert Table 1 about here]

Table 1 reports the overall statistic characteristics of the WTI prices in this paper. It discovers that the maxima and the minima have a great difference. The value of the standard deviation shows the WTI price fluctuates dramatically. The skewness and kurtosis values tell us that the time series data are “right-skewed” and peaked and pricing high occurs more than

pricing relatively low. The original price series does not pass the stationarity test and it needs further decomposition.

Fig.2 presents the original series and volatility of the crude oil price. The crude oil price trend can be approximately separated into six stages. From January 2004 to July 2008, the international WTI market returned bullish with the WTI price jumping from $33.71 to $145.31 per barrel. From July 2008 to February 2009, the WTI price experienced a cliff-like decline, falling to $34.03 per barrel. From February 2009 to June 2014, the WTI price rose to $107.95 per barrel. From June 2014 to February 2016, the WTI price again fell to $26.19 per barrel. From February 2016 to June 2018, the WTI prices kept rising to $77.41 per barrel. From June 2018 to December 2019, the WTI prices kept decreasing.

From the right panel of Fig.2, the crude oil price fluctuated with a difference of 20 dollars before 2008. But the volatility of the WTI price has increased significantly in recent years, and it is difficult to grasp its characteristics simply through the statistical description of the original WTI price series. To better characterize the volatility of the WTI price, this paper utilizes CEEMD to figure out the characteristics of the WTI price volatilities.

[Please insert Figure 2 about here]

5. Empirical Results

5.1 In-sample analysis

5.1.1 Multi-scale analysis of crude oil price

This paper uses CEEMD to decompose the WTI crude oil price sequence. The IMFs and the residual component are presented in Fig.3. As for CEEMD, considering that white noise should not affect the distribution of extrema with high-frequency and it should reduce the interval distribution of low-frequency components, the integrated used here is 100, and the white noise variance is 0.2. From Table II, the crude oil price is a nonlinear and nonstationary time series, the relevant regularity defying characterization. As shown in Fig.3, the price series is divided into 9 IMFs at different frequencies and 1 residual element.

[Please insert Table 2 about here]

[Please insert Figure 3 about here]

To investigate the impacts of the short-term, the mid-term and the long-term trends on crude oil prices better, this paper first calculates the mean of each IMF and performs a t-statistic test to investigate if the mean of each component is zero. It is found that the means of IMF1 to IMF5 fluctuate around zero. This paper uses the method proposed by Zhang et al. (2008) and sums up IMF1 to IMF5 to obtain the crude oil price part at high frequency. It means the crude oil price market fluctuations in common cases, i.e., the influences of the short-run fluctuations on the WTI price. The means of IMF6 to IMF9 is not zero with statistical significance. And they consist of the crude oil price part at low frequency, which demonstrates the impacts that the extreme events have on the WTI price, i.e., the effects of the medium-term price fluctuations. Huang et al. (1998) pointed out the rest of the components, i.e., residuals, proved to be the long-term trend. The graphs of the combined IMFs and the residuals for the crude oil price series are shown in Fig.4.

[Please insert Figure 4 about here]

From Table 2 and Fig. 4, this paper makes the following summarization.

(1) The long-run trend can be observed from the rest of the components, which contributes only 31.1367 % of the variance and has a positive but lower correlation to the crude oil price (0.3147) than the shock from extreme events. Therefore, it does not make the greatest difference in the movements of the crude oil price. The shock from extreme events rather than the market fluctuation determines the crude oil price trend. The conclusion is different from Zhang et al. (2008), which can be explained from three perspectives. Firstly, Zhang's sample period is from 1946 to 2006, when economic globalization developed slower than from 2004 to 2019 and the events which occurred in non-oil producing areas hardly affected the oil producing areas. Secondly, as a kind of financial subject matter, crude oil can obtain the goal of speculation and investment value which cannot be achieved in the Zhang's period. At that time, crude oil prices appear lower volatility since the WTI crude oil futures began in 1983 and the Brent crude oil futures began in 1988, with their trading volumes expanding continuously after that. Thirdly,

the US Shale Oil Revolution has greatly reduced the crude oil prices by offering the great substitute and lessening the power of crude oil depletion theory. Therefore, the influences of the long-term crude oil price trend has also been greatly reduced. Besides, from Fig. 4, it can be observed that the long-term trend decreased gradually before Feburary 2007 but moved upwards steadily after that time.

(2) The mid-run trend can be known from the term representing the impacts of extreme events. Its cumulative variance contribution rate was 81.6868%. This illustrates that the term for extreme events is the most essential element determining the crude oil price, with significant events which indicate the global economic situation and have dramatic effects on the crude oil price, such as geopolitical events and the 2008 financial crisis. As exhibited in Table 2, IMF18 makes a big difference in affecting the mid-run price trend of the WTI spot crude oil price, reaching the variance contribution rate of 22.9225% and having a negative correlation to the crude oil price. That proves most geopolitical events have negative effects on the crude oil prices. During the economic boom, investors pay more attention to the financial derivatives of crude oil and invest in the industries related with crude oil more, thereby leading the crude oil prices to increase obviously. And investors do the opposite during the economic recession.

(3) From the market fluctuation term, the short-run price trend is not so much correlated to the crude oil prices, and its contribution rate of cumulative variance is relatively lower. This proves that the impacts that the market fluctuations on the crude oil prices are limited. The crude oil price is generally believed to fluctuate in the short run is affected by political or economic factors, but it can adjust itself in the short term and fluctuates moderately and temporarily under the influence of factors such as the governments' macro-economic policies. Also, investors do not react to these events as strongly as to the events of systematic importance.

5.1.2 Structural Breaks

To validate the impacts that extreme events have on the crude oil prices, this paper employs the ICSS test and the Chow test. They can help to calculate the structural breaks from the short-run trend, the mid-run trend and the long-run trend which are extracted by CEEMD. The confidence level is set to be higher than 99% for the null hypothesis that there is no structural

breaks in the series. The 2693 daily observations spanning from January 5, 2004 to September 17, 2014 is selected as the test set. This paper first combines the ARIMA model, the ICSS test with the Chow test and detects 34 breaks in the significant event term. The assumed time period for structural breaks is set to be 90 days before and after the breakpoints, and the criterion of the minimum F-statistic for the Chow test is applied to confirm the happening time for each breakpoint finally.

(1) Analysis of the external events based on the mid-run crude oil price trend.

[Please insert Table 3 about here]

This paper regards the market fluctuation term in Fig.4 as the middle-run price volatility which affects the crude oil price movement temporarily and is estimated from the happening time point of the minimum price to that of the maximum price. As is shown in Table 3, there are 34 structural breaks identified by the ICSS test and the Chow test in the market fluctuation.

5.2 Out-of-sample Forecast

To prove that taking multi-scale intrinsic characteristics between the crude oil price series, the trends at different frequencies and impacts of significant events into account can promote prediction accuracy, this paper uses the seven models in Section 5 to forecast the crude oil prices. All the ARIMA-type models and the linear regression with vce are established via the Stata 14.0 and the individual SVM model and the CEEMD-based support vector machine are implemented using the software package invented by Faruto and LIBSVM 3.22 provided by Chang and Lin(2001). The empirical study is run on Matlab R2017b. The CEEMD-SVM models apply SVM for combination and price prediction. The CEEMD-ARIMA-SVM models use the ARIMA model to predict market fluctuation, the shock from the extreme events and the long-term trend, and the use of SVM here is as same as in the CEEMD-SVM models. Comparing the performances of these models without and with structural breaks, this paper tests whether taking significant event impacts into account can promote prediction accuracy. And comparing the performances of the SVM model, the ARIMA model, the CEEMD-ARIMA type models, the CEEMD-SVM type models and the CEEMD-ARIMA-SVM type models, this

paper tests whether taking nonlinearity and irregularity of crude oil prices and external event impacts into account can promote prediction accuracy.

From January 2004, crude oil prices have been rising rapidly and fluctuating more violently than before. This paper selects 4019 daily observations, which spans from January 5, 2004 to September 17, 2014. The first part, the 2693 observation values is taken as the training data set, and the remaining part, 1326 observation values, is taken as the test data set. This paper adopts three popular loss functions, i.e., mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) (Wen et al. 2016; Gong et al. 2017). The MSE, MAE and MAPE index can be written as

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \widehat{x_i})^2 \tag{12}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|(x_i - \widehat{x_i})| \tag{13}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{I=1}^{N}\left|\frac{x(t)-\hat{x}(t)}{x(t)}\right| \tag{14}$$

where x_i denotes the real crude oil prices and $\widehat{x_i}$ are the predicted values.

Besides, the statistic valuable D_{stat} can measure the capability of forecasting movement direction or changing point. It is proposed by Yao and Tan (2000). The Direction Change Statistics can be written as

$$D_{stat} = \frac{1}{N}\sum_{t=1}^{N} a_t \times 100\% \tag{15}$$

Because the short-run fluctuation term, the significant event term and the long-run crude oil price trend identify its own tendency, it is greatly essential to promote forecasting accuracy and predict the crude oil prices as accurately as possible. In this paper, the LIBSVM algorithm with the RBF kernel is chosen to predict its volatility of different scales (C.-C. Chang & Lin, 2001.).

[Please insert Table 4 about here]

From Table 4, the ARIMA model has the biggest prediction error and the SVM-type models have a better performance than the ARIMA-type models in terms of the values of MSE

and MAE. That indicates it can promote forecast accuracy to take the multi-scale volatilities and the nonlinearity of the crude oil prices into consideration. And from the predictions of SVMs, this paper shows that the extreme events impact on the crude oil price greatly and are essential factors when predicting the time series.

[Please insert Table 5 about here]

From Table 5, the CEEMD-ARIMA-SVM model with structural breaks has less prediction error than that without structural breaks, which also indicates the importance of adding the impacts of extreme events to the prediction model.

From Table 4 and Table 5, the CEEMD-ARIMA-SVM model with structural breaks has the least prediction error from the value of MSE and MAPE. This proves that when forecasting the crude oil price, it's of great importance and necessity to take into account the nonlinearity of the price series and the impacts of significant events. And the CEEMD-ARIMA-SVM models can have a more accurate prediction of the crude oil prices than the models in Table VI. It means the multi-scale volatilities are nonlinear and complex and the previous price trends have influences on the later ones.

5.3 Robustness Check

Furthermore, this paper examines whether the CEEMD-ARIMA-SVM model keeps good prediction accuracy in other periods by changing the test set. Since the crude oil prices have the features of complexity, irregularity and nonlinearity and the CEEMD-SVM type models have larger prediction errors. Utilizing different window regions could examine the robustness of the model (Rossi and Inoue 2012). This paper mainly examines the robustness of the CEEMD-ARIMA-SVM type models by decreasing 400, 300, 200 and 100 market days respectively based on the time interval of the out-of-sample prediction , which is following the works of Gong (2017). The results are as follows.

[Please insert Table 6 about here]

[Please insert Table 7 about here]

From Table 6 and Table 7, it is found that the CEEMD-ARIMA-SVM model with

structural breaks has smaller values which are calculated from the loss functions except the index D_{stat}. The CEEMD-ARIMA-SVM model with structural breaks has a better performance than the CEEMD-ARIMA-SVM model without structural breaks during different test periods which is consistent with our original results, indicating the good robustness of the CEEMD-ARIMA-SVM model with structural breaks.

6. Conclusions

We consider the impacts of extreme events and establish an integrated model named CEEMD-ARIMA-SVM model to make a forecast of the crude oil price volatility. CEEMD decomposes the crude oil price into the price constituents at various frequencies to extract a market fluctuation, a shock from extreme events and a long-run trend. Shocks from extreme events are significantly vital to impact on volatility. We used the above price trends respectively and utilized the SVM and ARIMA models to make a forecast of the price volatility. Compared with the ARIMA models and the linear regression with vce, the SVM models perform better in predicting the crude oil prices, which means the crude oil prices are nonlinear and complex. Thus, investors should work out speculation and investment plans according to their financial needs, capital capacity and risk preference when they make crude oil investments.

Moreover, without considering the impacts of extreme events, CEEMD-ARIMA-SVM has the highest accuracy and the lowest calculation results of loss function, which means it performs best, compared to the other models, such as SVM, ARIMA, the linear regression with vce, CEEMD-SVM, CEEMD-ARIMA and so on. Considering the impact of extreme events, we examine the points of structural breaks in different periods by the ICSS test and the Chow test. After adding the structural break factors into the model, the forecasting results improved based on CEEMD-ARIMA-SVM. Besides, we examine the robustness of the model.

The framework of our study can be utilized in different fields such as tourism demand forecasting, silver future price forecasting, carbon emission forecasting and so on. It can help different industries' decision-makers to make optimal decisions and support related industry policies.

Compliance with Ethical Standards

Ethical approval: This article does not contain any studies with human participants or animals.

Disclosure statement: No potential conflict of interest was reported by the author(s).

Data availability statement: Data will be made available on reasonable request

Author contributions: All authors contributed to the study conception and design. The conceptualization and the methodology were formulated by Yuxiang Cheng. Datacuration was completed by Jiayu Yi. And the formal analysis was finished by Xiaoguang Yang, Kin Keung Lai and Luis Seco. All authors have read and agreed to the published version of the manuscript.

References

Bekiros SD, Diks CGH (2008) The relationship between crude oil spot and futures prices: Cointegration, linear and nonlinear causality. Energy Econ 30:2673–2685

Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. J Econom 31:307–327

Catania L, Proietti T (2020) Forecasting volatility with time-varying leverage and volatility of volatility effects. Int J Forecast 36:. https://doi.org/10.1016/j.ijforecast.2020.01.003

Chang C-C, Lin C-J LIBSVM: A library for support vector machines

Chang CL, Mcaleer M, Tansuchat R (2013) Conditional Correlations and Volatility Spillovers between Crude Oil and Stock Index Returns. North Am J Econ Finance 25:116–138

Coleman L (2012) Explaining crude oil prices using fundamental measures. Energy Policy 40:318–324

D. D. Wu, L. Zheng, D. L. Olson (2014) A Decision Support Approach for Online Stock Forum Sentiment Analysis. IEEE Trans Syst Man Cybern Syst 44:1077–1087. https://doi.org/10.1109/TSMC.2013.2295353

Das SP, Padhy S (2018) A novel hybrid model using teaching-learning-based optimization and a support vector machine for commodity futures index forecasting. Int J Mach Learn Cybern 9:97–111

Engle RF (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the variance of United Kingdom Inflation. Econometrica 50:987–1007

Gong X, Wen F, Xia X, et al (2017) Investigating the risk-return trade-off for crude oil futures using high-frequency data. Appl Energy 196:152–161

Gupta R, Yoon SM (2018) OPEC news and predictability of oil futures returns and volatility: Evidence from a nonparametric causality-in-quantiles approach. North Am J Econ Finance 45:206–214

Han L, Lv Q, Yin L (2017) Can investor attention predict oil prices? Energy Econ 66:547–558. https://doi.org/10.1016/j.eneco.2017.04.018

He K, Zha R, Wu J, Lai KK (2016) Multivariate EMD-Based Modeling and Forecasting of Crude Oil Price. Sustainability 8:387

He Z, He L, Wen F (2018) Risk Compensation and Market Returns: The Role of Investor Sentiment in the Stock Market. Emerg Mark Finance Trade 55:704–718

Herrera A, Hu L, Pastor D (2018) Forecasting crude oil price volatility. Int J Forecast 34:622–635. https://doi.org/10.1016/j.ijforecast.2018.04.007

Ho LC, Huang CH (2015) The nonlinear relationships between stock indexes and exchange rates. Jpn World Econ 33:20–27

Hu C, Liu X, Pan B (2018) Asymmetric Impact of Oil Price Shock on Stock Market in China: A Combination Analysis Based on SVAR Model and NARDL Model. Emerg Mark Finance Trade 54:1693–1705

Huang NE, Shen Z, Long SR, et al (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc Math Phys Eng Sci 454:903–995

Imaouchen Y, Kedadouche M, Alkama R, Thomas M (2017) A Frequency-Weighted Energy Operator and complementary ensemble empirical mode decomposition for bearing fault detection. Mech Syst Signal Process 82:103–116

Inclan C, Tiao GC (1994) Use of cumulative sums of squares for retrospective detection of changes of variance. J Am Stat Assoc 89:913–923

Jacquier E, Polson NG, Rossi PE (2004) Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. J Econom 122:185–212

Ji Q, Guo J (2015) Oil price volatility and oil-related events: An Internet concern study perspective. Appl Energy 137:256–264

Kaytez F, Taplamacioglu MC, Cam E, Hardalac F (2015) Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. Int J Electr Power Energy Syst 67:431–438

Kekolahti P, Kilkki K, Hammainen H, Riikonen A (2016) Features as predictors of phone popularity:

An analysis of trends and structural breaks. Telemat Inform 33:973–989

Krichene N (2006) Recent Dynamics of Crude Oil Prices. IMF Work Pap 6:

Li W, Cheng Y, Fang Q (2020) Forecast on silver futures linked with structural breaks and day-of-the-week effect. North Am J Econ Finance 101192. https://doi.org/10.1016/j.najef.2020.101192

Luo X, Qin S (2017) Oil price uncertainty and Chinese stock returns: New evidence from the oil volatility index. Finance Res Lett 20:29–34

Martin V, Tang C, Yao W (2020) Forecasting the volatility of asset returns: The informational gains from option prices. Int J Forecast 37:. https://doi.org/10.1016/j.ijforecast.2020.09.012

Martina E, Rodriguez E, Escarela-Perez R, Alvarez-Ramirez J (2011) Multiscale entropy analysis of crude oil price dynamics. Energy Econ 33:936–947

Mensi W, Beljid M, Managi S (2014) Structural breaks and the time-varying levels of weak-form efficiency in crude oil markets: Evidence from the Hurst exponent and Shannon entropy methods. Int Econ 140:89–106

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

Oladosu G (2009) Identifying the oil price–macroeconomy relationship: An empirical mode decomposition analysis of US data. Energy Policy 37:5417–5426

Rafiq S, Salim R, Bloch H (2009) Impact of crude oil price volatility on economic activities: An empirical investigation in the Thai economy. Resour Policy 34:121–132

Ren R, Wu DD, Liu T (2019) Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine. IEEE Syst J 13:760–770. https://doi.org/10.1109/JSYST.2018.2794462

Risse M (2019) Combining wavelet decomposition with machine learning to forecast gold returns. Int J Forecast 35:601–615. https://doi.org/10.1016/j.ijforecast.2018.11.008

Rossi B, Inoue A (2012) Out-of-Sample Forecast Tests Robust to the Choice of Window Size. J Bus Econ Stat 30:432–453. https://doi.org/10.1080/07350015.2012.693850

Ruelke JC, Pierdzioch C, Stadtmann G (2011) On the internal consistency of short-term, medium-term and long-term oil price forecasts. Appl Econ 44:2757–2765

Shabri A, Samsudin R (2014) Daily Crude Oil Price Forecasting Using Hybridizing Wavelet and Artificial Neural Network Model. Math Probl Eng 1:1–10

Tang L, Dai W, Yu L, Wang S (2015) A novel CEEMD-based EELM ensemble learningparadigm

for crude oil price forecasting. Int J Inf Technol Decis Mak 14:141–169

Taylor SJ (1986) Modeling Financial Time Series. Econ J 97:

Vapnik VN (1995) The Nature of Statistical Learning Theory. Springer, New York, NY, New York

Vo MT (2009) Regime-switching stochastic volatility: Evidence from the crude oil market. Energy Econ 31:779–788

Wang lu, Ma F, Liu J, Yang L (2019) Forecasting stock price volatility: New evidence from the GARCH-MIDAS model. Int J Forecast 36:. https://doi.org/10.1016/j.ijforecast.2019.08.005

Wei Y (2019) Oil price shocks, economic policy uncertainty and China's trade: A quantitative structural analysis. North Am J Econ Finance 48:20–31

Wen F, Gong X, Cai S (2016) Forecasting the volatility of crude oil futures using HAR-type models with structural breaks. Energy Econ 59:400–413

Wen F, He Z, Dai Z, Yang X (2014a) Characteristics of Investors' Risk Preference for Stock Markets. Econ Comput Econ Cybern Stud Res 48:235–254

Wen F, He Z, Gong X, Liu A (2014b) Investors' risk preference characteristics based on different reference point. Discrete Dyn Nat Soc 2014:1–9

Wen F, Xiao J, Huang C, Xia X (2018) Interaction between oil and US dollar exchange rate: nonlinear causality, time-varying influence and structural breaks in volatility. Appl Econ 50:1–16

Wen F, Yang X, Gong X, Lai KK (2017) Multi-Scale Volatility Feature Analysis and Prediction of Gold Price. Int J Inf Technol Decis Mak 16:205–223

Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: A noise-assisted data analysis method. Adv Adapt Data Anal 1:1–41

Xiao J, Zhou M, Wen F, Wen F (2018) Asymmetric impacts of oil price uncertainty on Chinese stock returns under different market conditions: Evidence from oil volatility index. Energy Econ 74:777–786

Yao J, Tan CL (2000) A case study on using neural networks to perform technical forecasting of forex. Neurocomputing 34:79–98

Yeh JR, Shieh JS, Huang NE (2010) Complementary Ensemble Empirical Mode Decomposition: a Novel Noise Enhanced Data Analysis Method. Adv Adapt Data Anal 2:135–156

Yi S, Guo K, Chen Z (2016) Forecasting China's Service Outsourcing Development with an EMD-VAR-SVR Ensemble Method. Procedia Comput Sci 91:392–401

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

Yu L, Zhang X, Wang S (2017) Assessing Potentiality of Support Vector Machine Method in Crude Oil Price Forecasting. EURASIA J Math Sci Technol Educ 13:7893–7904

Z. Geng, J. Chen, Y. Han (2017) Energy Efficiency Prediction Based on PCA-FRBF Model: A Case Study of Ethylene Industries. IEEE Trans Syst Man Cybern Syst 47:1763–1773. https://doi.org/10.1109/TSMC.2016.2523936

Zhang B, Chan J, Cross J (2020) Stochastic volatility models with ARMA innovations: An application to G7 inflation forecasts. Int J Forecast 36:. https://doi.org/10.1016/j.ijforecast.2020.01.004

Zhang X, Lai KK, Wang S (2008) A new approach for crude oil price analysis based on empirical mode decomposition. Energy Econ 30:905–918

Zhang X, Yu L, Wang S, Lai KK (2009) Estimating the impact of extreme events on crude oil price: An EMD-based event analysis method. Energy Econ 31:768–778

Zhang X, Zhou J (2013) Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines. Mech Syst Signal Process 41:127–140

Zhao L, Cheng L, Wan Y, et al (2015) A VAR-SVM model for crude oil price forecasting. Energy Econ 38:126

Zhu B, Ye S, Wang P, et al (2018) A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting. Energy Econ 70:143–157

Figures

Figure 1

Crude oil price prediction process.

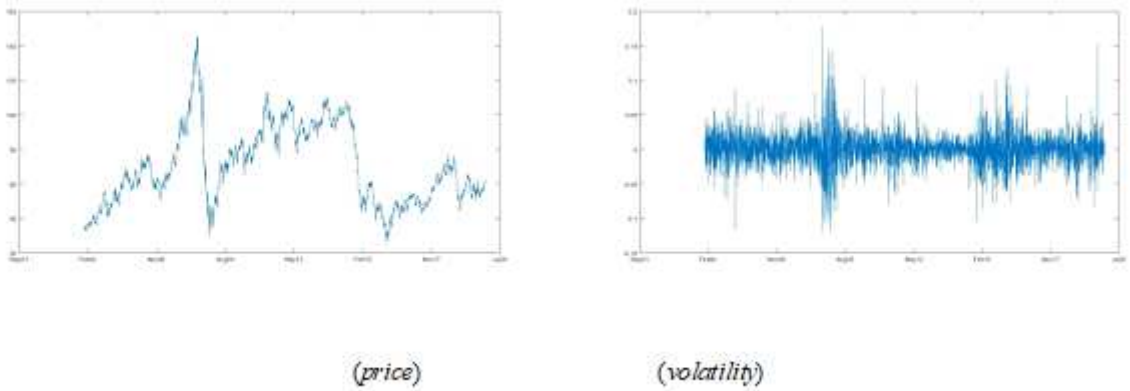

Figure 2

The crude oil price and its volatility.

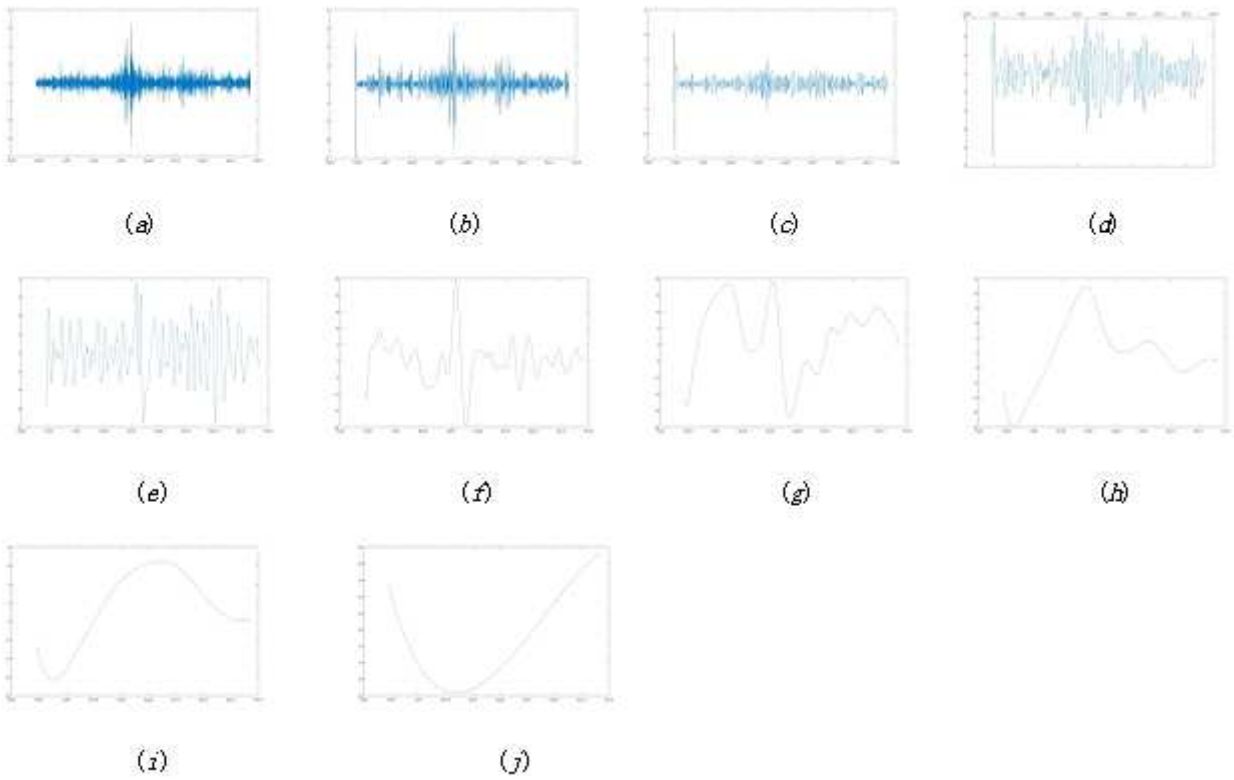

Figure 3

IMFs of the WTI price and the residual components.

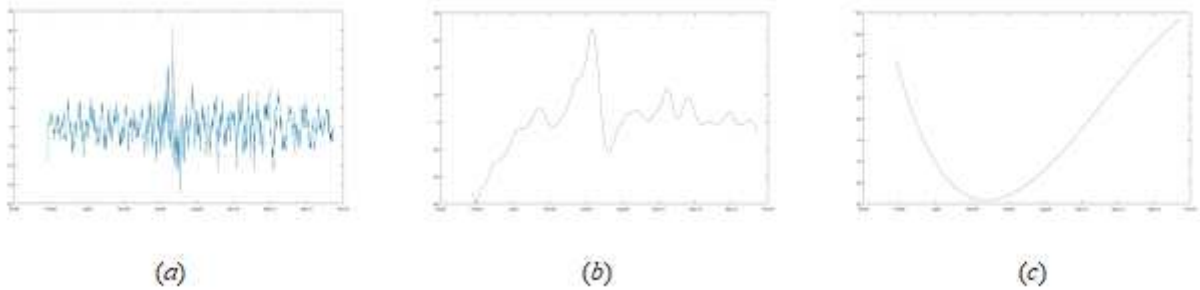

Figure 4

The three combined IMFs for crude oil price series.

Note: himf is the market fluctuation, limf is the shock from extreme events and res is the residual component.