

Research Article

Multiscale Combined Model Based on Run-Length-Judgment Method and Its Application in Oil Price Forecasting

Wang Shu-ping, Hu Ai-mei, Wu Zhen-xin, Liu Ya-qing, and Bai Xiao-wei

School of Economics and Management, North China University of Technology, Beijing 100144, China

Correspondence should be addressed to Wang Shu-ping; dwangshuping@163.com

Received 30 December 2013; Revised 15 April 2014; Accepted 14 May 2014; Published 12 June 2014

Academic Editor: Jianping Li

Copyright © 2014 Wang Shu-ping et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Forecasting of oil price is an important area of energy market research. Based on the idea of decomposition-reconstruction-integration, this paper built a new multiscale combined forecasting model with the methods of empirical mode decomposition (EMD), artificial neural network (ANN), support vector machine (SVM), and time series methods. While building the model, we proposed a new idea to use run length judgment method to reconstruct the component sequences. Then this model was applied to analyze the fluctuation and trend of international oil price. Oil price series was decomposed and reconstructed into high frequency, medium frequency, low frequency, and trend sequences. Different features of fluctuation can be explained by irregular factors, season factors, major events, and long-term trend. Empirical analysis showed that the multiscale combined model obtained the best forecasting result compared with single models including ARIMA, Elman, SVM, and GARCH and combined models including ARIMA-SVM model and EMD-SVM-SVM method.

1. Introduction

International commodity price fluctuates sharply and rises in the long run. Commodity prices' fluctuation creates large impact on global economy, for example, surging the cost of imports, triggering inflation, slowing economic growth, and diminishing the effectiveness of fiscal and monetary policy. Therefore analyzing the features of international commodity price fluctuations so as to predict the price and trend has a very important significance for global economy.

From the view of nations and governments, commodities are indispensable for economic development and have major strategic impact on national economic security. If the prices of commodities can be predicted more accurately, commodities can be imported when prices are low so that the imported inflation pressures will be greatly eased. Due to lower import costs, government can reduce the subsidies for enterprises, therefore, increasing the flexibility of fiscal policy. What is more, due to the reduction in foreign exchange expenditure, a country's foreign exchange reserves can be used to adjust the stability of the exchange rate, therefore, increasing the flexibility of monetary policy.

From the view of producers, commodities are raw materials for industries such as aviation, transportation, food processing. They are also products for industries such as oil exploration and nonferrous metals enterprises. Commodity price fluctuations have impact on business costs and profits. If the price fluctuation of commodities can be predicted more accurately, producers may make production plans more precisely, thus reducing costs and obtaining higher profits.

As for the trading companies, sharp fluctuations in commodity prices often result in great loss. However, due to lack of research team, operations team, and corresponding decision-making mechanism, the risk of price fluctuations is underdetermined. Especially when dealing commodity futures, the predicted price deviates greatly with actual price. Through building commodity price forecasting model and price fluctuation warning system, trading companies are able to avoid the risk of price fluctuations and reduce trade losses.

In general, in order to provide rational support for government and business decision makers, it is necessary and urgent to research on commodity prices forecast.

This paper is organized as follows. Section 2 is a literature review, mainly discusses single models and combined models

of forecasting oil prices, as well as the shortcomings of the existing multiscale combined model. Section 3 gives the detailed process of establishing a new multiscale combined model and gives a new idea, that is, the run-length judgment to reconstruct subsequences. In Section 4 the empirical results of forecasting oil prices are reported and commented. Section 5 concludes the paper.

2. Literature Review

Researchers have developed quite a lot of methods to analyze and forecast oil price, which can be roughly divided into single models and combined models. Single models include qualitative methods, causality regression methods, time series methods, and mathematical methods. Combined models are formed by combining single models according to certain rules.

Abramson and Finizza [1] predicted the trend of oil price based on OPEC member countries' output volume, transport capacity policy, and expert judgment. This qualitative method is flexible, easy, and quick to use. But it is of constraint to the knowledge, experience, and capacity of people who use it and it lacks the quantitative description. Ye et al. [2] established a regression forecasting model based on oil reserves, oil output, and oil import volumes. As a primitive constant coefficient regression model, it has some limitations. On one hand, the factors affecting oil prices are numerous and complex. It is impossible to include all the influencing factors. On the other hand, this model only considers the linear impact on oil prices. However, many studies showed that the fluctuation of commodity prices is nonlinear. Gori et al. [3] examined the evolution of oil price and consumption of oil in the last decades to construct a relationship between them and then predicted the trend of oil price. Lanza et al. [4] used the ECM specification to predict crude oil prices. In addition, some researcher applied mathematical methods into oil price forecasting. Fan et al. [5] proposed futures weighted oil price multistep prediction method based on PMRS. Ghaffari and Zare [6] applied a data filtering algorithm to predict daily oil price variation. Meade [7] established an oil price forecasting model based on Gaussian process. Mathematical methods are outstanding in fitting complex nonlinear functions and make improvement in oil price forecasting.

Based on the above single models, Krogh and Vedelsby [8] proved such a thought: when the single models are accurate and diversified enough, establishing a combined model with them can obtain better forecasting result. Combined models have become important for oil price forecasting after that. Wang et al. [9, 10] integrated text mining techniques, econometric models, and artificial neural networks to establish TEI@I, a new combined model, and used it to predict oil price. Nguyen and Nabney [11] built a combined model by putting together wavelet decomposition, adaptive machine learning methods, and adaptive GARCH and improved the prediction accuracy. de Souza e Silva et al. [12] used wavelets and hidden Markov models to predict oil price trends. Ahmad Kazem et al. [13] built a combined model

based on chaotic map, firefly algorithm, and support vector regression. Their empirical results showed that the combined model performed better than artificial neural network and support vector regression model based on genetic algorithm. Azadeh et al. [14] presented a flexible algorithm based on artificial neural network (ANN) and fuzzy regression (FR) to cope with optimum long-term oil price forecasting.

As oil price sequence showed characteristics of nonlinear, nonstationary, and multiscale, researchers began to use multiscale methods such as wavelet analysis, empirical mode decomposition (EMD) to analyze the fluctuations of oil price. These methods have good time resolution and frequency resolution and can enhance the variation regularity. Xie et al. [15] demonstrated that EMD has remarkable effect in time series decomposition and provided a powerful tool for adaptive multiscale analysis of nonstationary signals. At the same time, they proposed a bandwidth criterion for EMD and used bandwidth EMD to decompose electricity consumption data into cycles and trend which help us recognize the structure of the electricity consumption series. Zhang et al. [16, 17] used EMD to analyze the underlying characteristics of international crude oil price movements. They gave economic explanations to the reconstruction components such as normal supply-demand disequilibrium, the shock of significant events, and long-term trend. Yu et al. [18] decomposed oil price series with EMD, predicted all components with FNN, and then integrated them with ALNN. Li et al. [19] proposed a decomposition hybrid approach (DHA) to predict country risk for oil exporters based on the principle of "decomposition and ensemble" and the strategy of "divide and conquer." Experimental results, with ten major oil exporters as study samples, demonstrated that DHA with decomposition process was statistically proved to be much stronger and more robust than other prediction models. A hybrid ensemble learning paradigm integrating EEMD and LSSVR was proposed in Tang et al. [20]. The hybrid ensemble method was helpful to predict time series with high volatility. Mingming and Jinliang [21] proposed a multiwavelet recurrent neural network model.

Currently, the multiscale combined model is still in its early development stage, but its application prospect can be wide. There are still some problems. Firstly, the decomposed sequences are relatively numerous and the forecasting workload is large. Many papers did not reconstruct the subsequences and some papers reconstructed the subsequences subjectively rather than objectively. This paper puts forward run-length-judgment method to reconstruct all subsequences based on the feature that it has unified criterion and can reconstruct objectively. Secondly, the reconstructed sequences have little economic implications. This paper explores relevant economic implication which will help us understand oil price fluctuations. In addition, the selection of prediction methods will affect the accuracy of prediction. This paper chooses different prediction methods to predict different reconstructed sequences according to their characteristics.

3. The Establishing and Analysis of Multiscale Combined Model

3.1. The Basic Idea of Multiscale Combined Model. Generally speaking, the fluctuations of commodities price sequences have showed features of nonlinear, nonstationary, and multiscale. Multiscale refers to multiple frequencies. These features bring difficulties to the forecast of oil price. Some general prediction methods failed to grasp the complex characteristics and the laws of the data fluctuation therefore the prediction accuracy was reduced. In order to excavate the rules behind the commodity prices fluctuations, it is a good idea to decompose the original data sequence and analyze the features of each decomposed sequence. Among many decomposition methods, the multiscale method can mine inherent laws and essential features hidden in the data fluctuations of different scales and can determine the effect size of influencing factors. It also has advantages such as simple and low computation complexity. Therefore, the multiscale method is used in this paper to decompose the data sequences. It is efficient to analyze the influencing factors of price sequences and to better predict the trend of price series.

The basic idea of multiscale combined model is as follows: firstly to decompose the sequence in different scales with multiscale decomposition method; secondly to reconstruct the subsequences; then to predict reconstructed sequences with different prediction methods; and finally to integrate the predictions and obtain the final prediction value.

The process is shown in Figure 1.

3.2. The Process of Multiscale Combined Model. In detail, the basic process of model building is as follows.

Firstly, EMD is applied to decompose commodity price sequence. EMD is superior to wavelet analysis and Fourier decomposition method, in that it is direct and self-adaptive. EMD has better time resolution and frequency resolution, highlighting the local characteristics of data and enhancing the fluctuation rules of subsequence. It is also helpful in grasping the characteristic information of the original data in further analysis. With EMD decomposition, the original data sequence will be decomposed to n IMF components whose frequency is from high to low and one residual component.

Secondly, a new method, that is, run-length-judgment method, is proposed in this paper to reconstruct these subsequences. Generally, price sequence can be reconstructed into high frequency, medium frequency, low frequency, and trend part by run-length-judgment method. But some commodity price sequences may only be reconstructed into high frequency, low frequency, and trend part. Here we consider the general case.

Thirdly, ANN, SVM, and time series methods are used to predict the four parts. The machine learning methods such as ANN and SVM have advantages for high vibration frequency sequences. They have good learning ability and capture the nonlinear features of the data. The time series method such as ARIMA is a linear model and suitable for trend part forecasting.

Finally, SVM is selected to integrate.

The concrete steps of establishing multiscale model are as follows.

(1) *The First Step: Multiscale Decomposition by EMD.* Assume the original price series are $y(t)$ ($t = 1, 2, \dots, N$), and it can be decomposed into several IMFs by EMD; these IMFs are obtained by “sieving.”

(a) Find out all maxima of $y(t)$ and use cubic spline function to interpolate them into the upper envelope of the original data. Find out all minima of $y(t)$ and interpolate them into the lower envelope of the original data with cubic spline function. The mean of upper and lower envelope of original data is the average envelope $m_1(t)$. Subtracting the average envelope from the original data $y(t)$ will obtain a new data sequence $h_1(t)$:

$$y(t) - m_1(t) = h_1(t). \quad (1)$$

If $h_1(t)$ is an IMF, it will be the first subsequence. IMF must satisfy two conditions. Firstly, the number of its extreme points must be the same as or a different one at most with the number of zero crossing points. Secondly, the upper and lower envelope must be locally symmetric about the axis.

(b) If $h_1(t)$ is not an IMF, put $h_1(t)$ into the original data then repeat process (a) k times until the average envelope value tends to be 0. Then the first IMF component $C_1(t)$ is obtained which represents the high frequency sequence.

(c) Subtracting the first IMF $C_1(t)$ from the original data $y(t)$ will obtain a new sequence $R_1(t)$. Put $R_1(t)$ into the original data and repeat the processes (a) and (b), we will get the second IMF component $C_2(t)$. Repeat down until the last $R_n(t)$ is a monotonic function. Finally, the original data can be expressed by n IMF components and a residual component.

Consider

$$y(t) = \sum_{j=1}^n C_j(t) + R_n(t). \quad (2)$$

In (2), $R_n(t)$ represents the trend of the original price series; $y(t)$ and $C_j(t)$ ($j = 1, \dots, n$) represent frequency components from high frequency to low frequency, respectively.

(2) *The Second Step: Reconstruction by Run-Length-Judgment Method.* According to EMD decomposition theory, $R_n(t)$ represents the trend of signal, so we classify it as the trend term at first. Then, how to reconstruct these IMF sequences? This paper proposed a new idea as follows.

For n IMF sequences $C_1(t), C_2(t), \dots, C_n(t)$, use run-length-judgment method to calculate the run number of each IMF component. Assume that the mean of $\{C_i(t)\}$ is \overline{C}_i ; if the observed data is smaller than \overline{C}_i , we mark “-”; if the observed data is greater than \overline{C}_i , we mark “+”. Then we obtain a symbol sequence. Among the symbol sequence, each continuous identical symbol sequence is a run; thus, we can calculate the run number of each IMF. The run number reflects the volatility of data sequence. The larger the number is, the higher the volatility is. It is obvious that the run number of any IMF component will not exceed N , the total number of sample data.

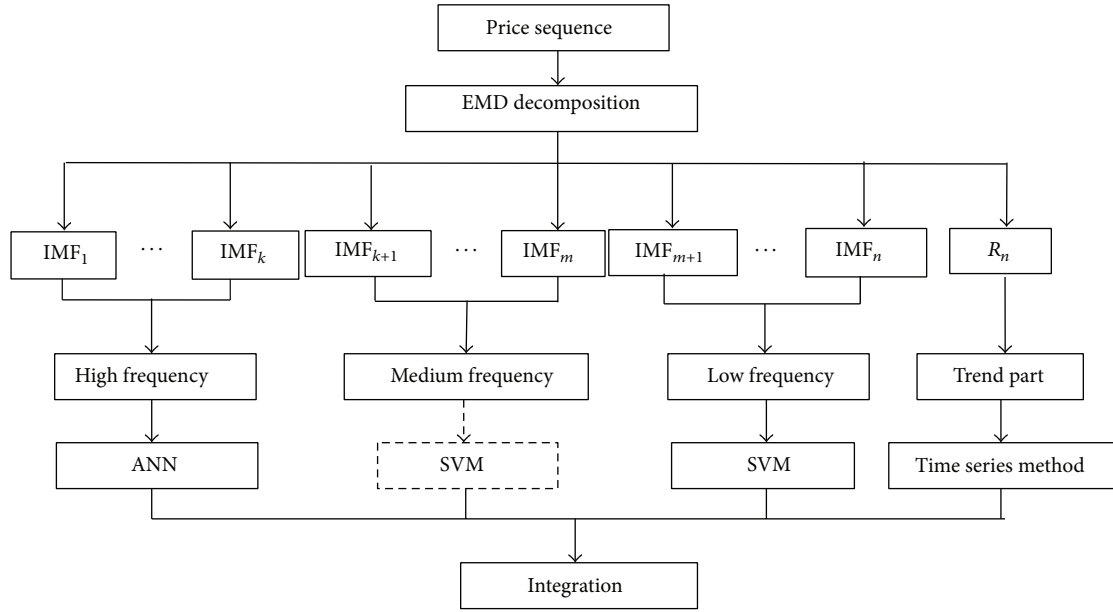


FIGURE 1: The process of establishing multiscale combined model.

After that, since the maximum possible run number is equal to the total number of samples, we divided the run-length number $[1, 2, \dots, N - 1, N]$ equally into n intervals. Then, the IMFs whose run-length number falls in the same interval will be reconstructed into one item. Those falling in the interval of larger run number are high frequency items, followed by medium frequency items and low frequency items. Thus, we can use run-length-judgment method to reconstruct IMFs into high frequency $y_1(t)$, medium frequency $y_2(t)$, and low frequency $y_3(t)$. The residual component $R_n(t)$ is the long trend item $y_4(t)$.

It can be seen that with the run-length-judgment method, the run interval number is determined by the number of IMF components and the interval length is determined by the number of samples and the number of IMF components. Therefore, the reconstruction method is completely dependent on characteristics of data fluctuation and data length and it is objective.

(3) *The Third Step: Prediction of Different Subsequences with Different Methods.* According to different scale features of reconstructed items, different methods are selected to predict.

(a) It can be learned from previous literatures [9, 10, 18] and our research work, Elman neural network is a suitable forecasting approach for high frequency data. So here Elman method is selected to predict high frequency sequences $y_1(t)$ and its predicted value is assumed as $\hat{y}_1(t)$.

(b) SVM is selected to predict medium frequency $y_2(t)$ and low frequency $y_3(t)$. SVM model maps the input variable $y(t)$ to the high feature space through a nonlinear mapping ϕ , then runs linear regression in the feature space, and constructs the optimal learning machine. According to the principle of minimizing structural risk, the Duality theory

and Saddle Point condition, we will obtain the following output variables:

$$\hat{y}(t) = \sum_{i=1}^n (\alpha_i^* - \alpha_i) k(y(t)) + b. \quad (3)$$

Therefore, we will get the predicted values of medium frequency $y_2(t)$ and low frequency $y_3(t)$, which we assumed as $\hat{y}_2(t)$ and $\hat{y}_3(t)$.

(c) ARIMA is selected to predict trend item $y_4(t)$. We firstly determine the order of the model and estimate the unknown parameters. Then we establish ARIMA (p, q) model $\hat{y}(t) = c + \varphi_1 y(t-1) + \varphi_2 y(t-2) + \dots + \varphi_p y(t-p) + \varepsilon(t) + \theta_1 \varepsilon(t-1) + \dots + \theta_p \varepsilon(t-p)$. Finally we will obtain the predicted value of trend item $y_4(t)$, which is assumed as $\hat{y}_4(t)$.

(4) *The Fourth Step: Integration of Prediction Results of All Subsequences with SVM Model.* The method of SVM integration is to set all predicted results as input and actual price as output. After learning certain amount of samples, it will build a mapping function between component predicted results and actual price. For a trained model, when the inputs are predicted results of four components, the output is the final predicted value $\hat{y}(t)$.

We can get the final predictive value $\hat{y}(t)$ by integrating the predicted values of the four components with SVM. The equation of this multiscale combined model is

$$\hat{y}(t) = \text{svm}(\hat{y}_1(t), \hat{y}_2(t), \hat{y}_3(t), \hat{y}_4(t)). \quad (4)$$

TABLE 1: The run number of each IMF.

IMF	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7
Run number	186	77	32	13	8	5	3

TABLE 2: The cycle and variance contribution rate of each sequence.

	High frequency	Medium frequency	Low frequency	Trend
Cycle	1.6	4	13	312
Variance contribution rate (%)	0.4	1.8	14.5	82.2

4. International Oil Price Forecasting

In this paper, we use the above multiscale combined model to analyze the characteristics of oil price fluctuation and forecast the oil price trend.

4.1. Data Selection and Evaluation Criteria. This paper selects monthly spot price of the US West Texas international (WTI) crude oil from January 1986 to November 2013, 335 data in total. The unit of price is US dollar/barrel. Data from January 1986 to December 2011 are used as the training set, and data from January 2012 to November 2013 are used as test set. All oil price data are from the US Energy Information Administration (<http://www.eia.doe.gov>).

NMSE and MAPE are used to measure the prediction accuracy and DS are used to measure the trend prediction ability. These three evaluation index are calculated as follows:

$$\begin{aligned}
 \text{NMSE} &= \frac{1}{n\delta^2} \sum_{t=1}^n (y(t) - \hat{y}(t))^2, \\
 \delta^2 &= \frac{1}{n-1} \sum_{t=1}^n (y(t) - \overline{y})^2 \\
 \text{MAPE} &= \frac{1}{n} \sum_{t=1}^n \left| \frac{y(t) - \hat{y}(t)}{y(t)} \right| \\
 \text{DS} &= \frac{1}{n} \sum_{i=1}^n d_i, \\
 d_i &= \begin{cases} 1, & (y(t) - y(t-1))(\hat{y}(t) - \hat{y}(t-1)) \geq 0 \\ 0, & (y(t) - y(t-1))(\hat{y}(t) - \hat{y}(t-1)) \leq 0. \end{cases}
 \end{aligned} \tag{5}$$

The smaller NMSE and MAPE are, the larger DS is, and the prediction method is better. MATLAB and EViews are used in this paper to process the data. The core part of SVM modeling was finished based on LIBSVM designed by Lin. EMD and Elman were developed in MATLAB. ARIMA modeling was done in EViews.

4.2. Oil Price Sequence Decomposition and Reconstruction. In this part of the training set data, 312 in total, from January 1986 to December 2011 are decomposed and reconstructed.

Firstly, EMD is applied to decompose WTI oil spot price series into seven IMF components and a residual component. The result is shown in Figure 2.

According to the EMD theory, we firstly classify R as the trend component. Then we calculate the run number of the seven IMF components. The result is shown in Table 1.

The run-length-judgment method is used to reconstruct the seven IMF components. Firstly, we divide the run-length number $[1, 2, \dots, 311, 312]$ into 7 equal intervals, that is, $[1, 44]$, $[45, 88]$, $[89, 132]$, $[133, 176]$, $[177, 220]$, $[221, 264]$, and $[264, 312]$. It is obvious that IMF1 falls in the fifth interval $[177, 220]$, IMF2 falls in $[45, 88]$, and IMF3, IMF4, IMF5, IMF6, and IMF7 fall in $[1, 44]$. Therefore, according to the run-length-judgment method, IMF1 is classified as high frequency item, IMF2 as medium frequency, and IMF3 to IMF7 as low frequency after summation. The WTI oil prices and the four reconstructed sequences are shown in Figure 3.

4.3. Fluctuation Features of Reconstructed Sequences. To observe the fluctuation features of the four sequences, the cycle and variance contribution rate are calculated. The cycle is determined by the data number of each sequence divided by the number of extreme point. The variance contribution rate is the ratio of the variance of each sequence divided by the overall variance of the original data; that is, $A_i = \xi_i/\xi$, ξ_i is the variance of the i sequence and ξ is the variance of the original oil prices series. Statistical results are shown in Table 2.

It is shown from the table that the four reconstructed sequences have different features.

For trend component, its variance contribution rate reaches 82.2% and its cycle is 312. Trend component is the most important component of oil price and plays a decisive role for long-term fluctuation. The rising trend component is synchronous with the world economy growth, which indicates that the long-term trend of oil price is fundamentally determined by world economy. Although oil price may fluctuate heavily by the influence of some major events such as wars, it will return to trend price in the long run. In a word, trend component represents the long-term trend of oil price without the influence of other factors.

For low frequency component, its variance contribution rate reaches 14.5% and its cycle is 13 months. It is shown in Figure 3 that the shape of the low frequency sequence is consistent with oil price sequence. Specially, every volatility point is correspondent with major event influencing the oil price. For example, oil price rise from 1990 to 1991 (at 58 point in Figure 3) resulted from the Gulf War. High volatility in 2008 (at 270 point in Figure 3) was caused by the US subprime crisis. These illustrate that low frequency sequences reflect oil price volatility affected by major events. Therefore

TABLE 3: Oil price forecasting in short-term: one year forecasting (Jan. 2012 to Dec. 2012).

	Multiscale combined model	ARIMA	Elman	SVM	GARCH	ARIMA-SVM	EMD-SVM-SVM
NMSE	0.5594	1.7079	0.8756	2.1201	2.0306	1.733	0.7999
MAPE	0.0506	0.0927	0.0728	0.1017	0.0999	0.0963	0.0579
DS (%)	75	66.67	66.67	66.67	66.67	50	75

TABLE 4: Oil price forecasting in long-term: 23 months forecasting (Jan. 2012 to Nov. 2013).

	Multiscale combined model	ARIMA	Elman	SVM	GARCH	ARIMA-SVM	EMD-SVM-SVM
NMSE	1.0925	1.844	1.3895	1.6791	2.2636	2.1074	1.1434
MAPE	0.0548	0.0844	0.0697	0.0787	0.0946	0.0938	0.0653
DS (%)	65.22	60.9	65.22	65.22	60.87	52.17	58.33

it is believed that low frequency component represents the significant impact of major events. It is important to separate low frequency to predict oil price.

For medium frequency, its variance contribution rate is 1.8% and its cycle is about 4 months. In Figure 3, medium frequency sequence shapes as sine wave or cosine wave and its cycle is about one season. In general, economic time series is vulnerable to seasonal factors. It is believed that medium frequency mainly reflects seasonal factors.

For high frequency component, its variance contribution rate is 0.4% and its cycle is 1-2 months. Although high frequency component has little effect on oil price, its cumulative effect is not neglectable. With the financialization of international crude oil market, the impact of speculation on short-term fluctuation is on the rise. High frequency component is highly volatile and its regularity is not obvious. This paper argues that high frequency term represents the impact of psychological factors, speculation, and other irregular factors on oil price.

4.4. Oil Price Forecasting and Comparative Analysis. We select oil price data from January 1986 to December 2011 as a training set to train the multiscale combined model.

ANN is used to predict high frequency component, SVM to predict medium frequency and low frequency component, and ARIMA to predict trend component. The prediction and integration methods are as follows.

- (1) Predict high frequency component $y_1(t)$ with Elman Neural Network. Input layer node is 3, output layer node is 1, and hidden layer node is 10.
- (2) Predict medium frequency component with SVM. Use cross-validation method to find the optimal parameters: $c = 6$, $\gamma = 0.5$, and $\varepsilon = 0.05$.
- (3) Predict low frequency component with SVM. Use cross-validation method to find the optimal parameters: $c = 3$, $\gamma = 5$, and $\varepsilon = 0.05$.
- (4) Predict trend component with ARIMA model. After repeating attempts we chose ARIMA (5, 1, 3).

- (5) Integrate the prediction results of the four components with SVM. Use cross-validation method to find the optimal parameters: $c = 4$, $\gamma = 2$, and $\varepsilon = 0.05$.

To further prove the validity of the proposed multiscale combined model, we select oil price data from January 2012 to December 2013 as a test set. Then we make comparative analysis with single models including ARIMA, Elman, SVM, and GARCH, as well as combined models including ARIMA-SVM model, EMD-SVM-SVM method which is a decomposition ensemble model without “run-length-judgment method” (i.e., using a uniform tool of SVM to forecast all of IMFs) and the EMD-SVM-SVM method advocated in the literatures [16, 17].

The short-term and long-term forecasting effects are shown in Tables 3 and 4.

It can be seen from Tables 3 and 4 that the multiscale combined model improves the prediction accuracy and direction accuracy of oil prices forecasting. No matter in short-term or in long-term, according to the three evaluation indicators including NMSE, MAPE, and DS, the multiscale combined model established in this paper performs better than the single models such as ARIMA, Elman, SVM, and GARCH, as well as the combined model such as ARIMA-SVM model and EMD-SVM-SVM method. For examples, in short-term forecasting, the most precise model among previous models is the EMD-SVM-SVM method. Its NMSE and MAPE are 0.7999 and 5.79%, respectively, and the DS is 75%. However, for the multiscale combined model, NMSE and MAPE are 0.5594 and 5.06%, respectively, and the DS is 75%. It is even better than EMD-SVM-SVM method. In long-term forecasting, the NMSE and MAPE of the multiscale combined model are smaller than those of other methods, meanwhile the DS of the multiscale combined model is not smaller than that of any other methods.

Overall, the multiscale combined model is a good forecasting method and suitable to forecast international oil prices.

4.5. A Description of the Model Application. Comparing with other existing prediction model including single model and combined model, the multiscale combined model established in this paper has several characteristics as follows.

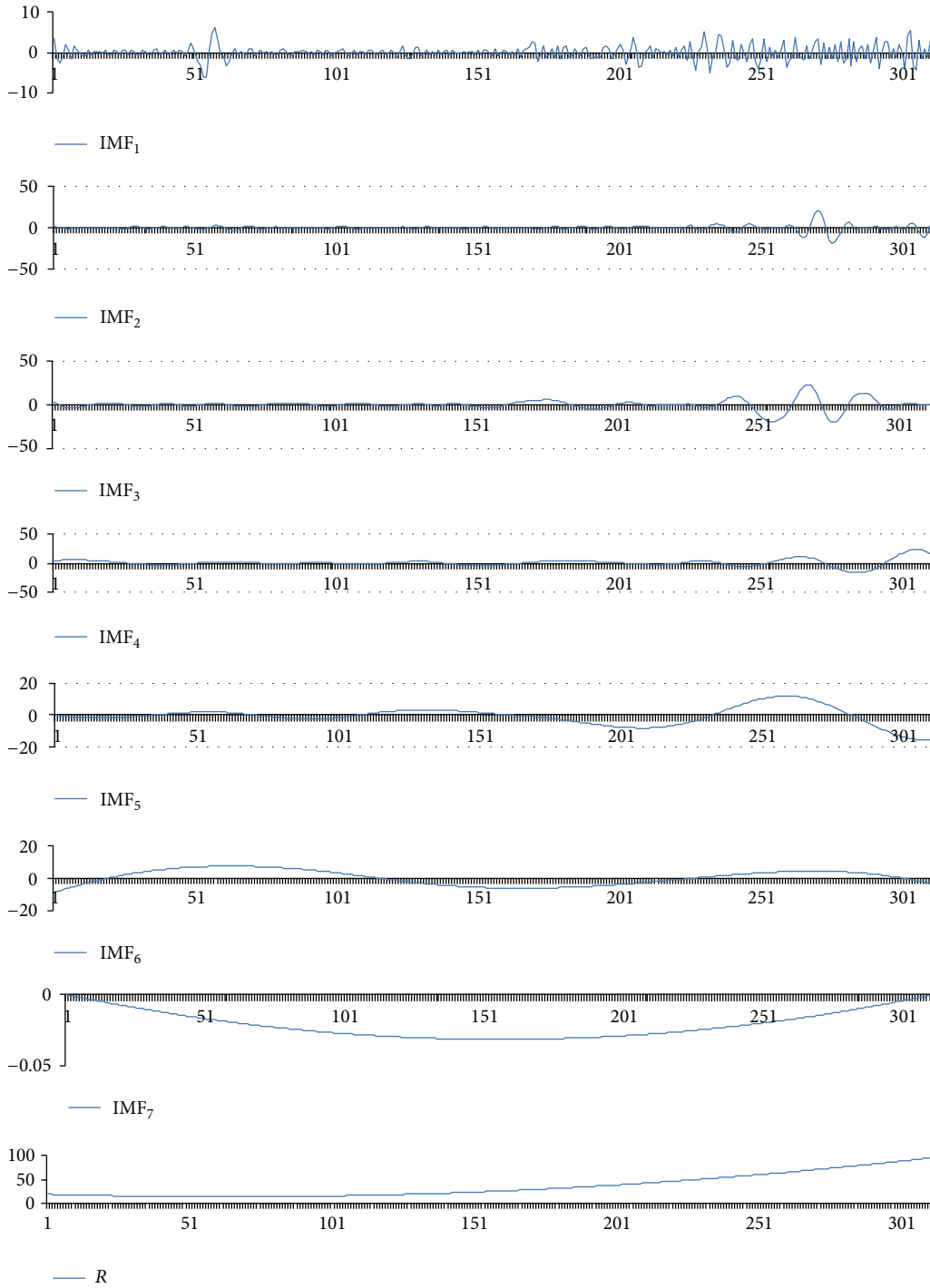


FIGURE 2: EMD decomposition of WTI oil spot price.

- (1) According to the different scale characteristics of data, the model decomposes the data into seven subsequences and a trend component. These subsequences are independent with each other. So, this model can grasp the data's features from different scales and reveal the underlying rules.
- (2) Instead of analyzing the components sequences obtained by EMD decomposition, this paper applies

the run-length-judgment method to reconstruct the components sequences. The reconstruction has three advantages. Firstly, the main feature of data fluctuation is revealed and the movement pattern of reconstructed sequences is found out. Secondly, the run-length-judgment method is only dependent on the data fluctuation features and the data length, so the reconstruction of component sequences is objective.

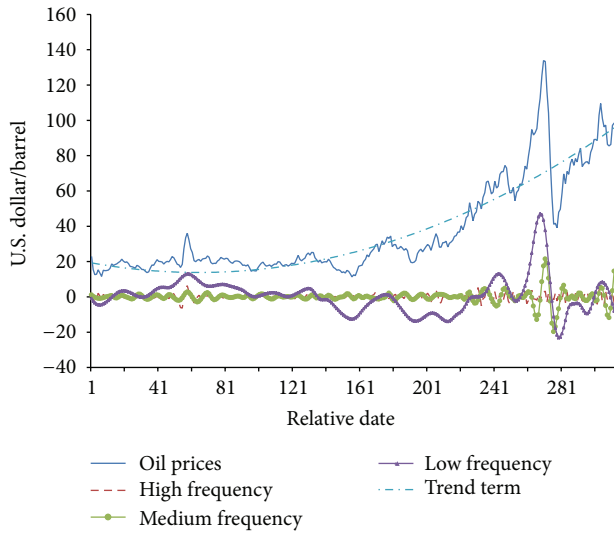


FIGURE 3: The fluctuation graph of oil prices and the four reconstructed sequences.

Thirdly, it is easy to find out influencing factors according to the fluctuation features of reconstructed sequences and give economic explanations. If there are too many sequences, it is impossible to analyze and explain.

- (3) This model applies ANN, SVM, and time series method to predict the four reconstructed sequences, respectively, which take the advantages of each prediction method. It is different from other existing combined forecasting model such as the literatures [16–18], which use the same method to predict the reconstructed sequences.
- (4) Comparing with the simple linear integration, this model prevails for using SVM to integrate and capture the nonlinear relationship between variables.

Based on the above characteristics, the multiscale combined model established in this paper has obvious advantage in forecasting the oil price and improves the prediction accuracy.

When using this model to predict commodity prices, attention needs to be paid to several aspects. One is that this model is suitable for the economic time series which have characteristics of nonlinear, nonstationary, multiscale. But it is not suitable for the high frequency financial time series (for example, there is a data in one minute or several seconds). This model is established through a complex process of “decomposition-reconstruction-prediction-integration.” It is necessary to give economic explanation to reconstructed sequence. However, for the high frequency financial time series data, it is hard to give the economic meaning to the reconstruction sequences. The second point is that the sample data need to be above certain length, which is required by the multiple methods. For the small sample (no more than 30 data), this model is not appropriate.

5. Conclusions

Forecasting of oil price is a challenging problem. This paper establishes a new multiscale combined model to analyze oil price fluctuations and predict its trend. EMD is used to decompose nonstationary oil price sequence into several IMFs and a residual component. For the IMFs, we propose a new run-length-judgment method to reconstruct them. Empirical test shows that oil price sequences can be reconstructed into high frequency sequence, medium frequency sequence, low frequency sequence, and trend sequence. By analyzing the fluctuation features of these four components and referring to existing researches, we point out that high, medium, and low frequency and trend sequences, respectively, reflect the irregular factors, seasonal factors, major events, and the world economy’s impact on oil price. Then Elman neural network method, support vector machine (SVM), and time series method are applied to forecast the four components. Finally SVM is used to integrate the four predicted results and obtained the final prediction value. Empirical analysis shows that both in short-term and long-term forecasting, the multiscale combined model is better than single models including ARIMA, Elman, and SVM, as well as combined models including ARIMA-SVM model and EMD-SVM-SVM method.

This multiscale combined model not only enhances the prediction accuracy, but also confers certain economic implications to reconstructed sequences. The proposed multiscale combined model is a combination of “data-driven modeling” and “theory-driven modeling” and is suitable for commodity price volatility analysis and prediction.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research is supported by the Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality (PHR20110869), the Special Fund of Subject and Graduate Education of Beijing Municipal Education Commission (PXM2013_014212_000005), and the National Natural Science Foundation Project of China (71301002), as well as Funding for training talents in Beijing (401053711403).

References

- [1] B. Abramson and A. Finizza, “Probabilistic forecasts from probabilistic models: a case study in the oil market,” *International Journal of Forecasting*, vol. 11, no. 1, pp. 63–72, 1995.
- [2] M. Ye, J. Zyren, and J. Shore, “A monthly crude oil spot price forecasting model using relative inventories,” *International Journal of Forecasting*, vol. 21, no. 3, pp. 491–501, 2005.
- [3] F. Gori, D. Ludovisi, and P. F. Cerritelli, “Forecast of oil price and consumption in the short term under three scenarios: parabolic,

- linear and chaotic behaviour,” *Energy*, vol. 32, no. 7, pp. 1291–1296, 2007.
- [4] A. Lanza, M. Manera, and M. Giovannini, “Modeling and forecasting cointegrated relationships among heavy oil and product prices,” *Energy Economics*, vol. 27, no. 6, pp. 831–848, 2005.
- [5] Y. Fan, Q. Liang, and Y.-M. Wei, “A generalized pattern matching approach for multi-step prediction of crude oil price,” *Energy Economics*, vol. 30, no. 3, pp. 889–904, 2008.
- [6] A. Ghaffari and S. Zare, “A novel algorithm for prediction of crude oil price variation based on soft computing,” *Energy Economics*, vol. 31, no. 4, pp. 531–536, 2009.
- [7] N. Meade, “Oil prices—Brownian motion or mean reversion? A study using a one year ahead density forecast criterion,” *Energy Economics*, vol. 32, no. 6, pp. 1485–1498, 2010.
- [8] A. Krogh and J. Vedelsby, “Neural network ensembles, cross validation and active learning,” *Neural Computing and Applications*, vol. 25, pp. 231–238, 1995.
- [9] S. Y. Wang, L. A. Yu, and K. K. Lai, “A novel hybrid AI system framework for crude oil price forecasting,” in *Data Mining and Knowledge Management*, Y. Shi, W. Xu, and Z. Chen, Eds., vol. 3327 of *Lecture Notes in Computer Science*, pp. 233–242, 2004.
- [10] S. Y. Wang, L. A. Yu, and K. K. Lai, “Crude oil price forecasting with TEI@I methodology,” *International Journal of Systems Science and Complexity*, vol. 18, pp. 145–166, 2005.
- [11] H. T. Nguyen and I. T. Nabney, “Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models,” *Energy*, vol. 35, no. 9, pp. 3674–3685, 2010.
- [12] E. G. de Souza e Silva, L. F. L. Legey, and E. A. de Souza e Silva, “Forecasting oil price trends using wavelets and hidden Markov models,” *Energy Economics*, vol. 32, no. 6, pp. 1507–1519, 2010.
- [13] A. Kazem, E. Sharifi, F. K. Hussain, M. Saberi, and O. K. Hussain, “Support vector regression with chaos-based firefly algorithm for stock market price forecasting,” *Applied Soft Computing Journal*, vol. 13, no. 2, pp. 947–958, 2013.
- [14] A. Azadeh, M. Moghaddam, M. Khakzad, and V. Ebrahimipour, “A flexible neural network-fuzzy mathematical programming algorithm for improvement of oil price estimation and forecasting,” *Computers and Industrial Engineering*, vol. 62, no. 2, pp. 421–430, 2012.
- [15] Q. Xie, B. Xuan, S. Peng, J. Li, W. Xu, and H. Han, “Bandwidth empirical mode decomposition and its application,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 6, no. 6, pp. 777–798, 2008.
- [16] X. Zhang, K. K. Lai, and S.-Y. Wang, “A new approach for crude oil price analysis based on empirical mode decomposition,” *Energy Economics*, vol. 30, no. 3, pp. 905–918, 2008.
- [17] X. Zhang, L. Yu, S. Wang, and K. K. Lai, “Estimating the impact of extreme events on crude oil price: an EMD-based event analysis method,” *Energy Economics*, vol. 31, no. 5, pp. 768–778, 2009.
- [18] L. Yu, S. Wang, and K. K. Lai, “Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm,” *Energy Economics*, vol. 30, no. 5, pp. 2623–2635, 2008.
- [19] J. P. Li, L. Tang, X. L. Sun, L. A. Yu, W. He, and Y. Y. Yang, “Country risk forecasting for major oil exporting countries: a decomposition hybrid approach,” *Computers and Industrial Engineering*, vol. 63, no. 3, pp. 641–651, 2012.
- [20] L. Tang, L. Yu, S. Wang, J. Li, and S. Wang, “A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting,” *Applied Energy*, vol. 93, pp. 432–443, 2012.
- [21] T. Mingming and Z. Jinliang, “A multiple adaptive wavelet recurrent neural network model to analyze crude oil prices,” *Journal of Economics and Business*, vol. 64, no. 4, pp. 275–286, 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

