A New Prediction Method of Gold Price: EMD-PSO-SVM

Jian-hui Yang
School of Business Administration, South China University of Technology, Guangzhou , China
Email: bmjhyang@scut.edu.cn

Wei Dou
School of Business Administration, South China University of Technology, Guangzhou , China
Email: dauwile@qq.com

***Abstract*—The current gold market shows a high degree of nonlinearity and uncertainty. In order to predict the gold price, Empirical Mode Decomposition (EMD) was introduced into Support vector machine (SVM). Firstly, we used the EMD method to decompose the original gold price series into a finite number of independent intrinsic mode functions (IMFs), and then grouped the IMFs according to different frequencies. Secondly, SVM was used to predict each IMF in which particle swarm optimization (PSO) was applied to optimize the parameters of SVM. Finally, the sum of each IMF's forecasting result will be the final prediction. In order to validate the accuracy of the proposed combination model, the London spot gold price and the Shanghai Futures gold price series were employed. Empirical studies indicated that the EMD-PSO-SVM model outperformed the WT-PSO-SVM model, and was feasible and effective in gold price prediction. We can promote the EMD-PSO-SVM to other related financial areas.**

***Index Terms*— empirical mode decomposition, independent intrinsic mode functions, support vector regression, wavelet transform, PSO, gold price**

I. INTRODUCTION

Gold is a symbol of wealth since from the ancient times. Since the disintegration of the Bretton Woods system in 1973, the gold is non-monetary which makes it become an important tool of financial market. The current gold market has high-yield and high risks. After long-term practice, a gold price theory gradually formed. However, due to the price of gold market is affected by many factors, there are a variety of uncertainties. Many scholars have done researches on it. Qian chose a fuzzy time series model to determine the initial parameters of the fuzzy system, and used Type -2 Fuzzy Systems and Type-1 fuzzy system to train and forecast the price of gold [1]. According to the basic principles of the gray prediction model and the Markov chain, Qin and Chen constructed gray Markov model to predict the price of gold, in which two methods can complement each other, making the predictions more reasonable and reliable[2]. Zhang, Yu and Li predicted the gold price on model based on wavelet neural network [3]. Summarize previous studies, we find that the gold price shows a high degree of nonlinearity and uncertainty, and currently, a variety of model results is not satisfactory.

Support vector machine (SVM), a novel learning machine based on statistical learning theory, was developed by Vapnik etc. in 1995 [4]. SVM can be used to solve problems in pattern recognition and makes out decision-making rules on generalization performance. SVM is attar-ctive and has widely been applied to various different fields [5,6]. Meanwhile, many scholars have studied the combination of innovative methods of SVM. Abdulhamit proposed that hybridized the particle swarm optimization and SVM method to improve the EMG signal classification accuracy [7]. Jazebi combined SVM and wavelet transforms, and used on a power transformer in PSCAD/EMTDC soft-ware [8]. Wei etc. studied wavelet decomposition, EMD and R / S valuation process, and finally got a fast, high-precision method of valuation of the Hurst exponent [9].

In this paper, empirical mode decomposition (EMD) and particle swarm optimization (PSO) are introduced into the SVM model to establish the gold price forecasting model. We used the London spot gold price and the Shang-hai Futures gold prices to validate the innovative method. The EMD-PSO-SVM model is expected to be more accu-rate and feasible in gold price prediction.

II. METHODOLOGY FORMULATION

Support vector machine (SVM) is a new and promising technique for data classification, regression and forecasting. In this section we give a brief description of SVM. Assume $\{(x_1,y_1),\ldots,(x_l,y_l)\}$ is the given training data sets, where each $x_i \subset R_n$ shows the input date of the sample and has a corresponding target value $y_i \subset R$ for i=1, . . . ,l. where l represents the size of the training data. The support vector machine solves an optimization problem:

$$\min \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i$$

China National Natural Science Foundation (71073056). Authors: YANG Jian-hui(1960-),male, professor, postdoctoral, the main research direction : the investment decision-making and risk management; DOU Wei(1989-),male, graduate student, the main research direction: finan-ceial decision-making ,corresponding author' E-mail adderss :dauwile @foxmail.com;

doi:10.4304/jsw.9.1.195-202

$$\text{Subjected to}\begin{cases} y_i - < w, x_i > -b \leq \epsilon_i + \xi \\ < w, x_i > +b - y_i \leq \epsilon_i + \xi^* \\ \xi_i, \xi_i^* \geq \quad i = \ \ldots l \end{cases}$$

Where x_i is mapped to a higher dimensional space by the function Φ,ξ_i is the upper training error(ξ_i^* is the lower) subject to the ε insensitive tube $y_i - < w, x_i > -b \leq \varepsilon$. The parameters which control the regression quality are the cost of error C, the width of the tube E, and the mapping function Φ. The constraints imply that we would like to put most data x_i in the tube $y_i - < w, x_i > -b \leq \varepsilon$. If x_i is not in the range, there is an error ξ_i or ξ_i^* which we would like to minimize in the objective function. SVM avoids under fitting and over fitting the training data by minimizing the training error $\sum_{i=0}^{l}(\xi \text{ or } \xi^*)$, as well as the regularization term $\frac{1}{2}\|w\|^2$. By contrast, traditional least square regression ε is always 0, and data are not mapped into a higher dimensional spaces. Therefore, SVM is a more general and flexible model on regression problems.

Many works in forecasting have demonstrated the favorable performance of SVM before. Therefore, SVM is adopted in this paper. The selection of the three parameters γ, ε and C of SVM will influence the accuracy of the forecasting result. However, there is no standard method of selection of these parameters. In the paper, particle swarm optimization technique is used in the proposed model to optimize parameter selection.

A. WT-PSO-SVM Forecasting Model

Let $\Psi(t) \in L^2(R)$, its Fourier transform as $\Psi(\omega)$, to meet permit conditions $C_\Phi = \int_R^d |\Psi(\omega)|^2 / |\omega| d\omega < \infty$. $\Psi(\omega)$ is basic wavelet, in continuous case,

$$\Psi_{ab}(t) = a^{(-1/2)}\ \Psi(\frac{(1-b)}{a}),$$

Where a is the dilation factor, b is the translation factor. Given any function $f(x) \in L^2(R)$, the continuous wavelet transform and its reconstruction formula is:

$$W_f(a,b) = (f, \Psi_{ab}) = |a|^{-1/2} \int_R^d f(t)\Psi(\frac{t-b}{a})dt$$

$$f(t) = \frac{1}{C_\Psi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{1}{a^2} W_f(a,b)\Psi\left(\frac{t-b}{a}\right) da db$$

Decomposition contains a signal of lower frequency component, while details contain higher frequency components. After wavelet decomposition and reconstruction, we can get a different frequency component. The original signal Y will be decomposed into $Y=D_1+D_2+\ldots+D_N+A_N$, Where $D_1,D_2,\ldots,D_N$, respectively, for the first layer, second layer to the N-tier decomposition of high-frequency signal (i.e., the detail signal). Then plus D_1 to D_N, and using PSO to optimize SVM's parameters to get the prediction result of the signals. Using PSO method to train the parameter of SVM again, and get the prediction result of A_N. Plusing the two prediction results, we can get the final prediction [9] [10].

Nowadays, the WT is a mature model to decompose the signal, and will get a good performance. WT method decomposed the signal into approximation signal and detail signals, then we use a combination of PSO and SVM model to predict the signals that WT decompose the original signal, finally can get a better result. Its prediction accuracy is higher than the traditional SVM method. EMD is a new type of signal decomposition tool, so we combine EMD, PSO and SVM together, and expect to have a higher accuracy and reduce errors. Therefore, comparing the WT-PSO-SVM method, we can test whether EMD-PSO-SVM is affective or not. In our next section, we will introduce EMD-PSO-SVM method.

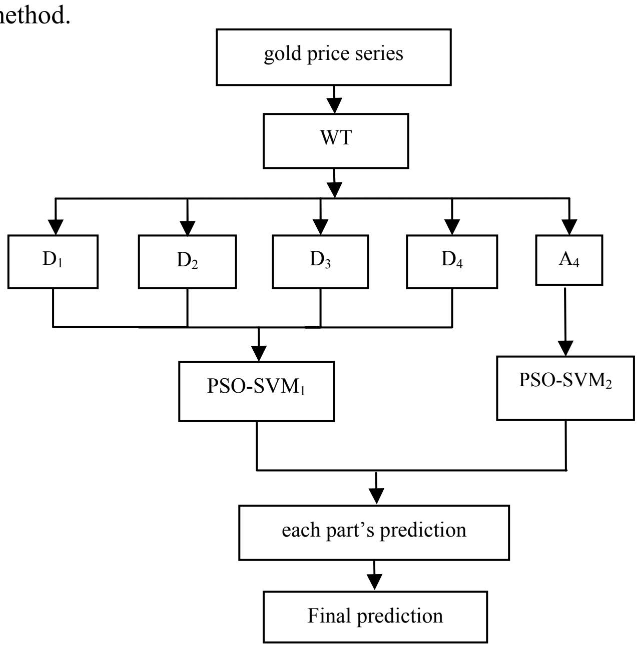


Figure 1. The construction of Gold price forecasting model by WT-PSO-SVM

B. EMD-PSO-SVM Prediction Model

EMD is a generally nonlinear, non-stationary data processing method developed by Huang et al. (1998) [11],[12]. It assumes that the data, depending on its complexity, may have many different coexisting modes of oscillations at the same time. EMD can extract these intrinsic modes from the original time series, based on the local characteristic scale of data itself, and represent each intrinsic mode as an intrinsic mode function (IMF), which meets the following two conditions:

1) The functions have the same numbers of extreme and zero-crossings or differ at the most by one;

2) The functions are symmetric with respect to local zero mean.

The two conditions ensure that an IMF is a nearly periodic function and the mean is set to zero.IMF is a harmonic-like function, but with variable amplitude and frequency at different times. In practice, the IMFs are extracted through a sifting process.

The EMD algorithm is described as follows:

1) Identify all the maxima and minima of time series x(t);

2) Generate its upper and lower envelopes, $e_{min}(t)$ and $e_{max}(t)$, with cubic spline interpolation.

3) Calculate the point-by-point mean (m(t)) from upper and lower envelopes:

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

4) Extract the mean from the time series and define the difference of x (t) and m (t) as d (t): $d(t) = x(t) - m(t)$

5) Check the properties of d (t):

If it is an IMF, denote d (t) as the ith IMF and replace x (t) with the residual $r(t) = x(t) - d(t)$.

The ith IMF is often denoted as ci(t) and the i is called its index;

If it is not, replace x (t) with d (t);

6) Repeat steps 1)-5) until the residual satisfies some stopping criterion.

The original time series can be expressed as the sum of some IMFs and a residue:

$$x(t) = \sum_{j=1}^{N} c_j(t) + r(t)$$

Similar with the wavelet decomposition, the price series is decomposed into a number of different frequencies IMFs. Based on fine-to-coarse reconstruction rule, the IMFs are composed into high-frequency sequence, low-frequency sequence and trend series. Then we choose the right kernel functions to build different SVM to predict each IMF. Finally, we plus the prediction of the high-frequency sequence, low-frequency sequence and trend series to get the final result[13].

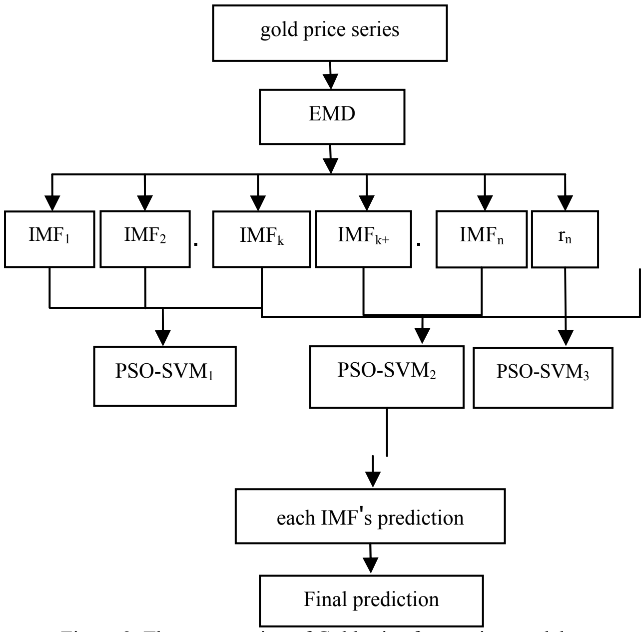


Figure 2. The construction of Gold price forecasting model by EMD-PSO-SVM

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. *Research Data*

The experimental data set comes from the London gold market's afternoon fixing price from January 4, 2005 to November, 28, 2011 which consists of a total of 1860 data. All data from http://www.lbma.org.uk, the London Bullion Market Association. For the missing data, we use the interpolation method to fill. In the London gold price series, the data from January 4, 2005 to February 21, 2011 are used as the training data, while February 22, 2011 to November 28, 2011 as the testing data.

At the same time, we chose the Shanghai gold futures closing price as another experimental data to test whether the combination method has good adaptability and stability. The time span is from January 8, 2008 to December.28 2011. The data from January 8, 2008 to September 15, 2011 are used as training samples, while the data from September 16, 2011 to December 28, 2011 as the testing data.

B. *Wavelet Decomposition and Prediction*

First of all, using wavelet method, we decomposed the gold price series in the London market to a detail signal and an approximation signal. Then continue to decompose the approaching signal, and get the next level of approximation and details signals. Repeat the above steps until we get four detail signals Di and an approximation signal A4, as illustrated in Fig. 3.

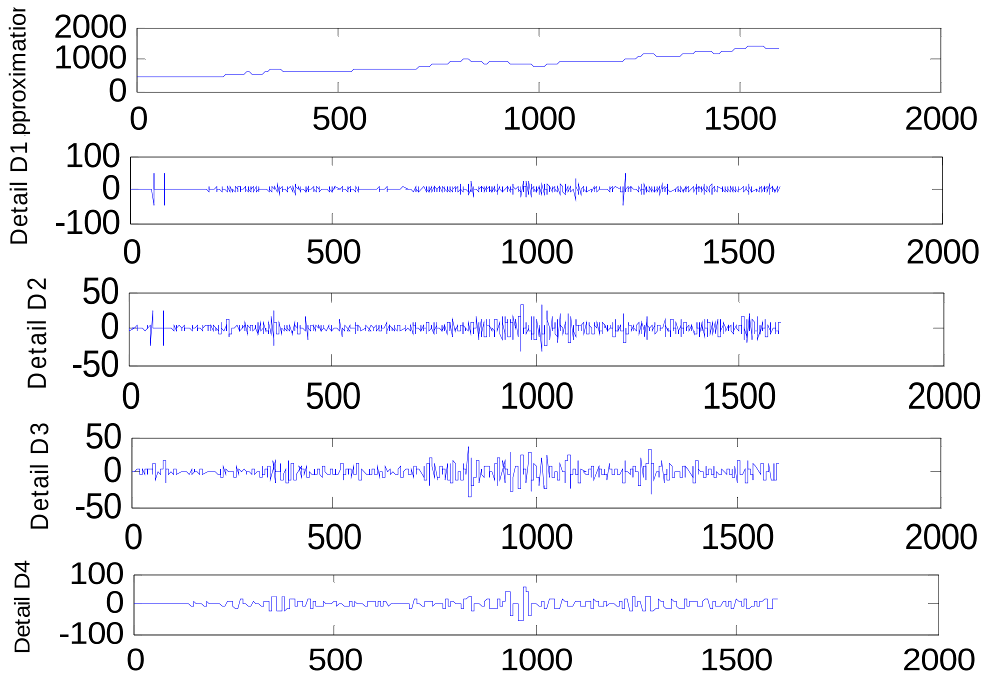
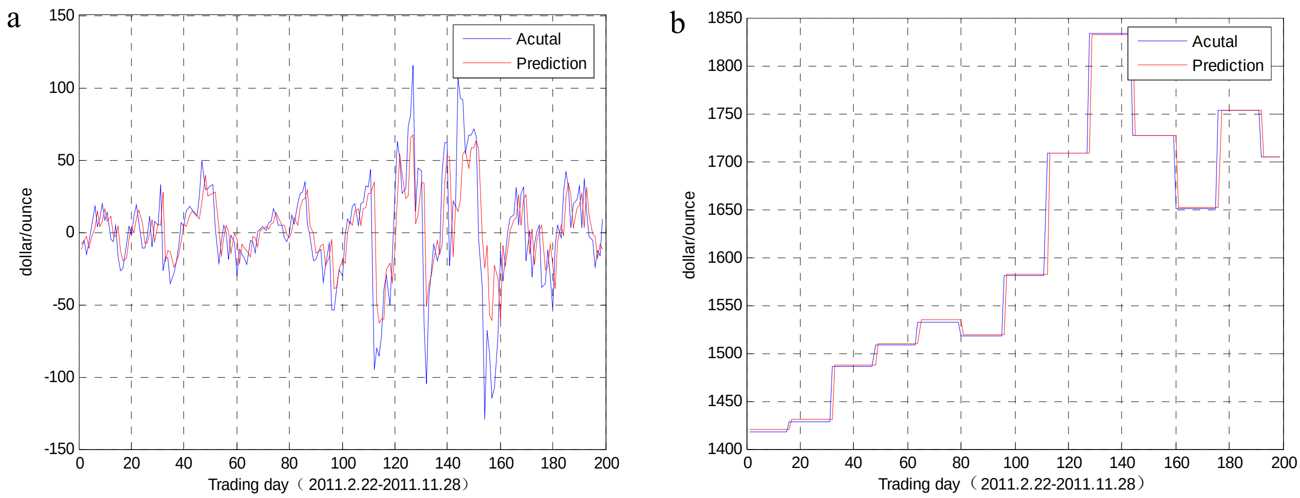


Figure 3. The decomposed using WT in London market

Adding the four details signal Di together, we can get the detail part of the entire sequence, representing the fluctuations in the market. While approaching signal indicates the direction of the market trend. We use the PSO-SVM method to predict the approximation signal and detail signal. The detail signal is in a short period of time, and has high frequency fluctuations, but less volatile. The detail signal basically fluctuates to 0 and prediction deviation is a little large. While the approaching signal's change frequency is low, but significant, showing a trend of change. The prediction deviation of the approaching signal is small. The predicted data of the details signal are plotted in Fig. 4(a) while the predicted approaching signal is shown in Fig. 4(b). The final prediction of WT-PSO-SVM is given in Fig. 4(c).



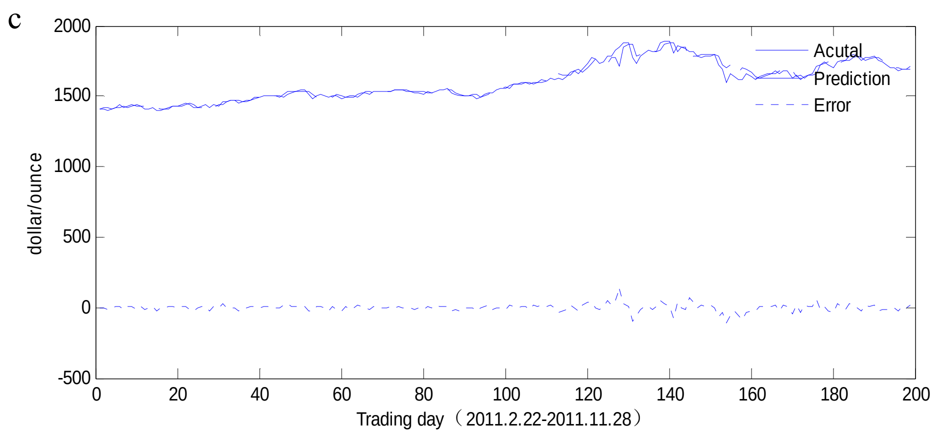


Figure 4. Forecasting results by WT-PSO-SVM in London market

As shown in the previous method, we have chosen the Shanghai gold futures market data to analyze. And we can get the final forecasting result that is substantially the same with the London market prediction, as illustrated in Fig 5.

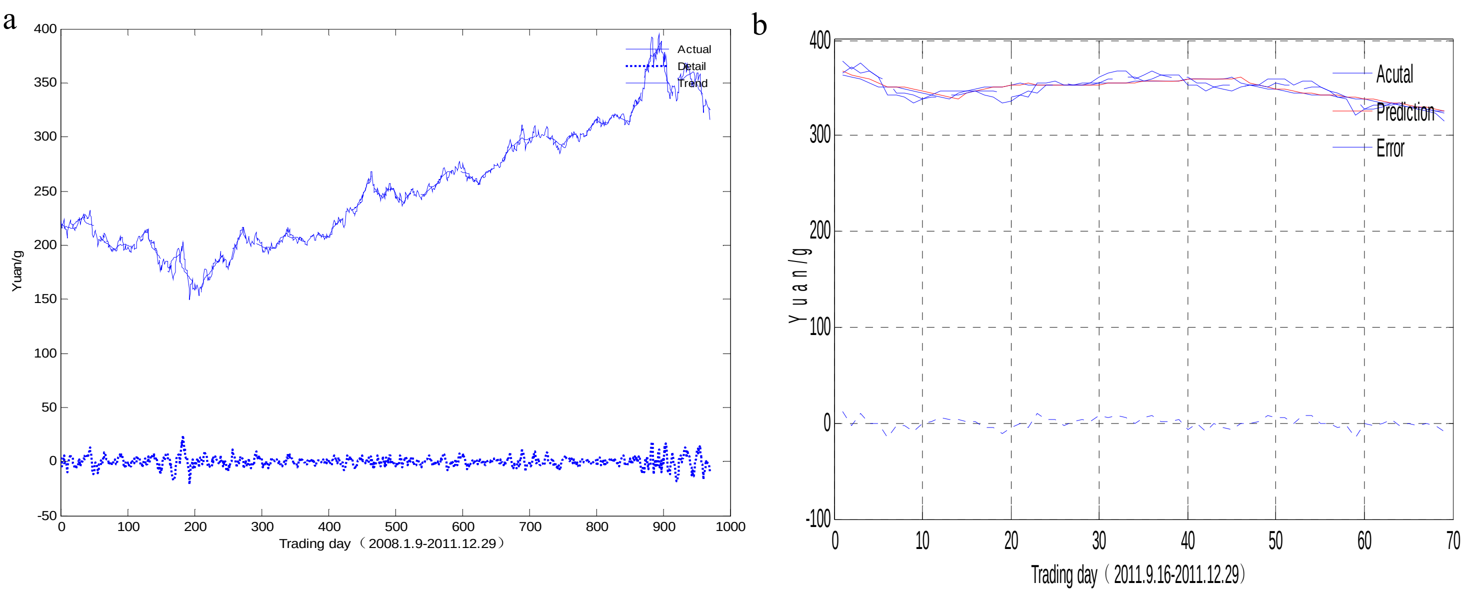
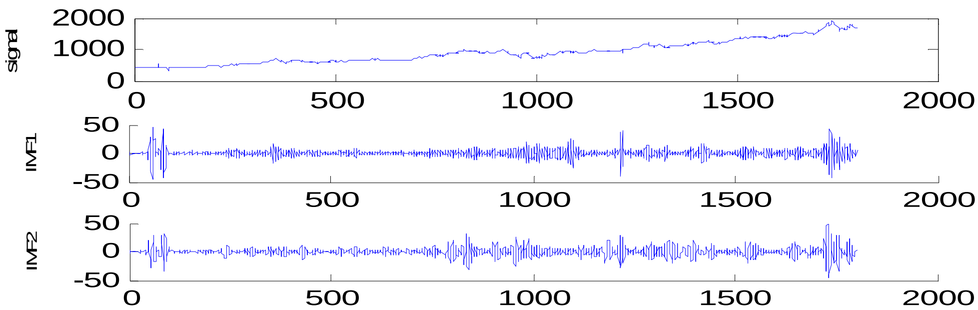


Figure 5. Forecasting results by WT-PSO-SVM in shanghai market

C. EMD Decomposition and Prediction

Firstly, using EMD method, we decomposed the gold price series in the London market. Unlike the wavelet composition, EMD method directly divided the gold price series into many IMFs and a residual component R. It can be seen from Fig. 6 that we finally get 7 IMFs and a residual component R.



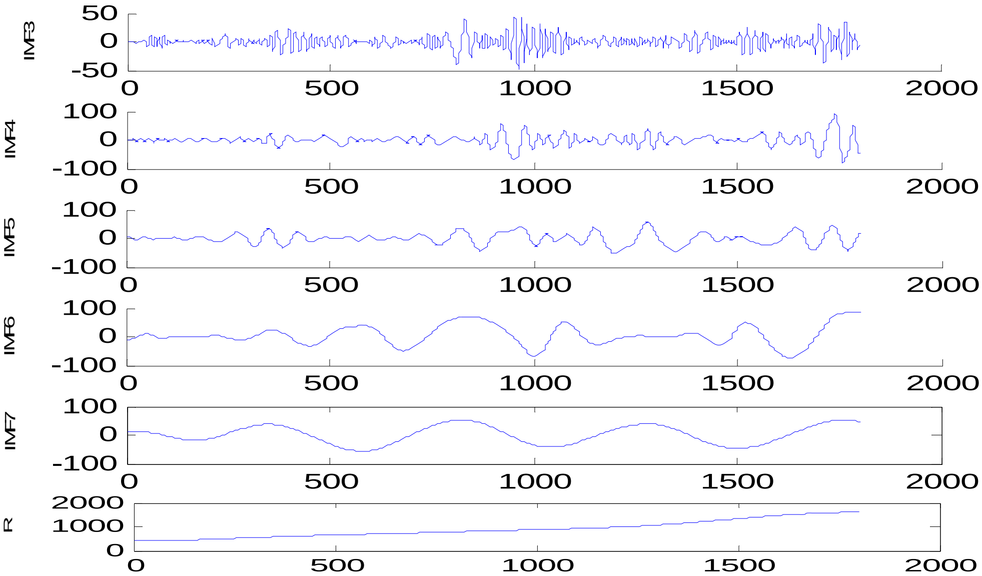
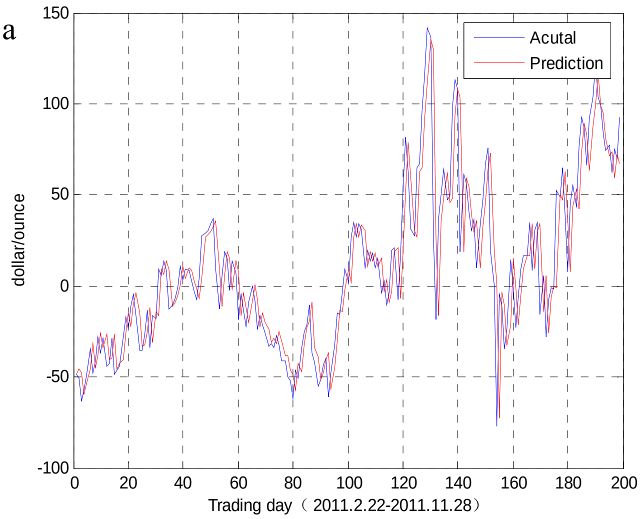


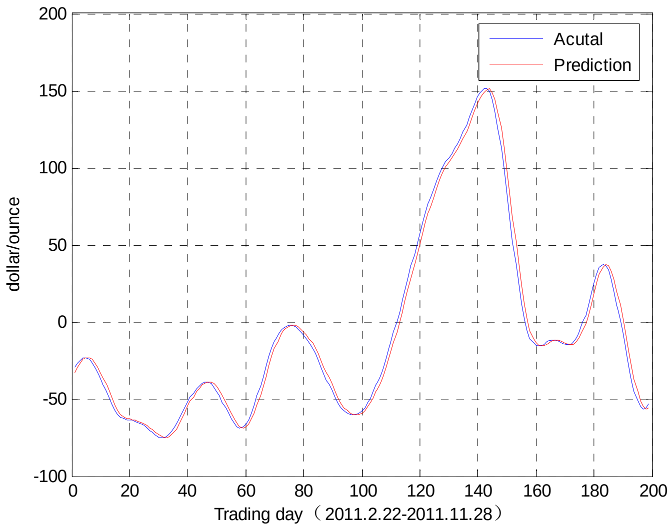
Figure 6. The decomposed using EMD decomposition

From Fig. 6, we can find that the average value of IMF1-IMF4 is approximately equal to zero and presents a very dramatic fluctuation. Group IMF1-IMF4 into a high-frequency sequence part, representing minor changes in the market, for example, some fabricated rumors. At the same time, the average value of IMF5-IMF7 is significantly greater than zero and presents a smaller frequency fluctuations, but long life cycle and impact. Group IMF5-IMF7 into a low-frequency sequence,representing big changes in the market which have long-term and stable affection to gold price, for example, raising interest rates by the central bank or enhancing the deposit reserve ratio. Finally, R represents the trend in the market, considered the impact of the global nature of the entire financial industry. For example, central banks around the world increase e-currency launch, causing the devaluation of the local currency. So gold as a hard currency, presents a continuous uplink irreversible trend. Then we use PSO-SVM method to predict different frequency sequences. The sum of each forecasting value will be the final prediction. The actual data and predicted data of different frequency sequences are shown in Fig. 7(a) to Fig. 7(c). And the final prediction of EMD-PSO-SVM method is given in Fig. 7(d).

a



b



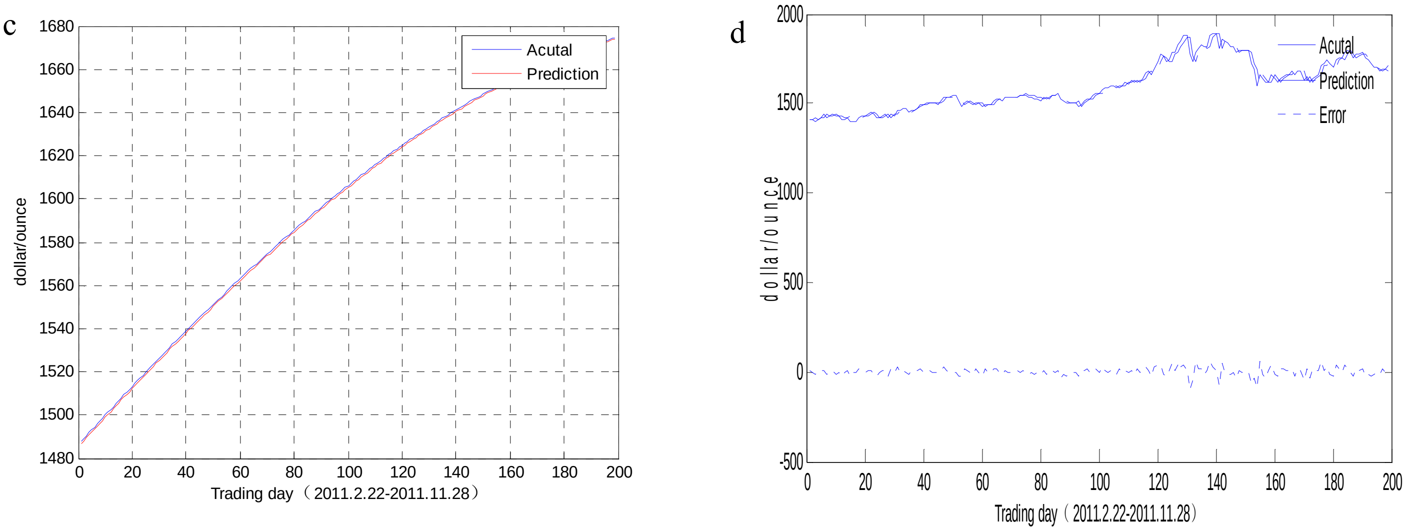


Figure 7. Forecasting results by EMD-PSO-SVM and details enlargement

As shown in the previous method, we have chosen the Shanghai gold futures market data to analyze. We can find that the gold price movements and error are substantially the same with the London market prediction, as illustrated in Fig 8. This shows that EMD-PSO-SVM model has stability and strong adaptability, and can be applied in different markets.

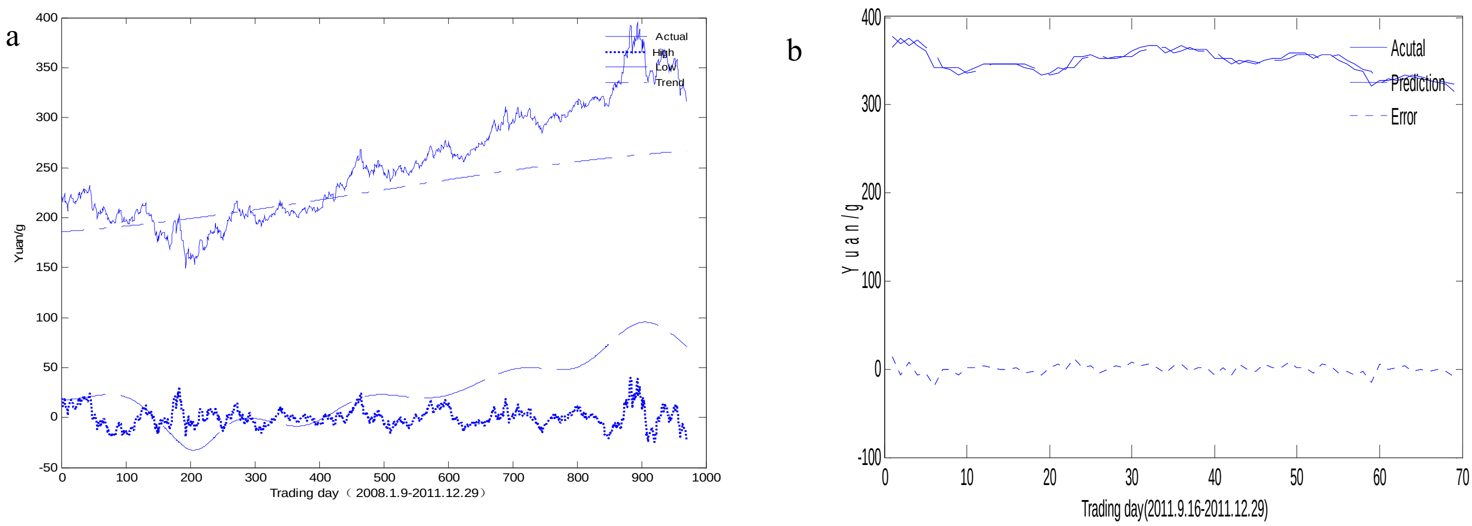


Figure 8. Forecasting results by EMD-PSO-SVM in shanghai market

D. The Comparison of the Two Methods' Results

The training errors of the two methods are all small. But we can still see that, the average error of WT-PSO-SVM method is 1.09% in London market. While in shanghai market, the error is 1.13%; the average error of EMD-PSO-SVM method is 0.46% in London market. While in shanghai market, the error is 0.49%. As a result, the EMD-PSO-SVM model outperformed the WT-PSO-SVM model, and was feasible and effective in gold price prediction.

E. Discussion

The results in the above experimental studies prove that comparing to the WT-PSO-SVM method, EMD-PSO-SVM has achieved a significant improvement in prediction accuracy. The accurate rate is at a high level. The reason is the following:

(1)The EMD is a better signal decomposition tool than the WT model, but it was generally used in industrial and oceanography field before. We introduce EMD into the financial sector, and the practice has proved that it is also effective. In particular, EMD portrays the characteristics of the series more accurately. Through analyzing the IMFs, we can find economic factors which affect the gold market movements and provide decision-making for investment.

(2) In the paper, SVM was used to predict each IMF in which PSO method was applied to optimize the

parameters of SVM. The most suitable parameters directly decide the prediction accuracy of SVM.

(3) Since gold price is non-stationary, using EMD method decompose the series, then regroup the sequence in accordance with the different frequency which avoid the errors of accumulation in the respective segments in the prediction process.

(4) Comparing to the London gold market, the prediction error is larger in Shanghai gold market. It mainly manifests in the short-term high frequency part. Because when influenced by some market fluctuations, the gold price in Shanghai market will appear a larger fluctuation, showing that the Shanghai gold market is not as mature as the London gold market, and it has poorer resistance to shock. Nowadays, the investors are also sensitive to market noise impact. Long-term value investment idea has not deeply rooted.

IV. Concluding Remarks

In this article, we study the London and Shanghai gold market price data, and successfully establish the prediction model. Firstly, we use the EMD method and wavelet method to decompose each set of data, and then reconstruct them. Add the predicted data together respectively and obtain our final results. In the forecasting part, we combine the PSO and the SVM together. In order to improve the prediction accuracy of the SVM, the PSO is used to optimize the parameters. Empirical studies showed that: comparing the EMD method to the traditional wavelet method, the prediction accuracy is significantly improved. Especially when processing nonlinear and non-stationary data, EMD has its own superiority. EMD method will be as much as possible to retain the basic features of the original data. The PSO combined with SVM method can effectively improve the prediction accuracy. The combination of EMD, PSO and SVM can decompose the data into several layers which we can analyze the characteristics of each time series data to get better explanatory of the prediction results. Also, through the analysis of each sequence prediction results, we can know short and long term trend of fluctuations of the gold price, and find the influencing factors which can guide our investment. The EMD-PSO-SVM method has a wide range of practical value, and can be promoted to other related financial areas.

References

[1] Qian Bingbing, "The Application of Type-2 Fuzzy System to Forecast the Price of Gold," Journal of Jiamusi University, vol. 25, pp. 397-399, 2007.

[2] Qin Sheng and Chen Yang, "Prediction Gold Price base on Grey Markov Model," China Economist, vol.10, pp. 29-31, 2010.

[3] Zhang Kun, Yu Yong and Li Tong, "Application of wavelet neural network in prediction of gold price," Computer Engineering and Applications, vol. 46(27), pp. 224-226,2010

[4] V.N. Vapnik, "The Nature of Statistical Learning Theory, " Springer, New York, 1995.

[5] Dongxiao Niua,Yongli Wanga,Desheng Dash Wu, "Power load forecasting using support vector machine and ant colony optimization," Expert Systems with Applications,2010, 37(3):2531-2539.

[6] Koknar-Tezel, LJ Latecki, "Improving SVM classification on imbalanced time series data sets with ghost points," Knowledge and information systems2011,28:1-23.

[7] Abdulhamit Subasi, "Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders," Computers in Biology and Medicine,2013,In Press.

[8] S. Jazebi ,B. Vahidi, M. Jannati, "A novel application of wavelet based SVM to transient phenomena identification of power transformers," Energy Conversion and Management,2011,52(2): 1354-1363.

[9] Wei Bin, Wu Chongqing and Shen Ping, "Fast Hurst Value Computation Base on Wavelet Transform and EMD," Computer Engineering, vol. 34, pp.1-3,2008

[10] Chen Wei, Wu Jiejun, Duan Weijun, "Model of Urban Air Pollution Concentration Forecast Based on Wavelet Decomposition and Support Vector Machine," Modern Electronics Technique, vol.34,pp145-148,2011

[11] HUANG N E,SHEN Z and LONG S R, "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis, " The Royal Society A: Mathematical, Physical & Engineering Sciences, pp. 903-995, 1998

[12] HUANG N E,SHEN Z and LONG S R, "A New View of Nonlinear Water Waves: The Hilbert Spectrum, " Annual Review of Fluid Mechanics, vol. 31, pp. 417-457, 1999.

[13] Wang Wei, Zhao Hong, Liang Zhaohui and Ma Tao, "Hybrid intelligent prediction method based on EMD and SVM and its application," Computer Engineering and Applications, vol. 48,pp225-227,2012