Vol. 48 (2017) ACTA PHYSICA POLONICA B No 11

A NEW APPROACH TO GOLD PRICE ANALYSIS BASED ON VARIATIONAL MODE DECOMPOSITION AND INDEPENDENT COMPONENT ANALYSIS

E Jianwei, Li Shenggang

College of Mathematics and Information Science, Shaanxi Normal University
Xi'an 710119, China

Ye Jimin

School of Mathematics and Statistics, Xidian University, Xi'an 710071, China

(Received April 24, 2017; accepted November 11, 2017)

Gold plays a significant role in the investment market. The analysis of influence factors of the gold price is a research hotspot for the portfolio managers, investors and regulatory authorities. There are a lot of articles which investigate the gold price using different methods. However, there is a scarce literatures on the research of the gold price from the perspective of internal structure of the temporal sequence. In this paper, we generate a novel approach which combines variational mode decomposition (VMD), principal component analysis (PCA) and independent component analysis (ICA) (VMD-PCA-ICA). Specifically, three steps are contained: firstly, the gold price is decomposed into the sum of intrinsic mode functions (IMFs) via VMD method and the recombination of IMFs (RIMFs) is obtained in light of the contribution coefficients between IMF and gold price data. Secondly, PCA is applied to RIMFs and the principal components (PCs) are acquired. Finally, ICA method is utilized to separate the gold price and independent components (ICs) are obtained. The proposed approach demonstrates that the gold price sequence is a linear combination of ICs. Meanwhile, the influence factors of the gold price are analysed from the viewpoint of the correspondence between ICs and economic factors. Compared with the EMD-ICA, EEMD-ICA and VMD-ICA, the approach in this paper has the fastest convergence speed.

DOI:10.5506/APhysPolB.48.2093

1. Introduction

Gold, as a special commodity in the present international finance markets, includes the properties of merchandise, money and the financial nature. Gold price is tightly related to macroeconomic variable, international politics, speculations *etc.* Thus, the fluctuation of the gold price becomes complex. In recent years, academic and industry circles were engaged in researching the influence factors of the gold price. Mandelbrot [1] is the earliest researcher to study the time sequence by the non-linear physics research. Wen *et al.* [2] extracted three influence factors of the gold price via complementary ensemble empirical mode decomposition (CEEMD). Dai *et al.* [3] studied multifractal properties of the gold price and further found that the gold price is heterogeneous. Malliaris [4] employed a decision-tree analysis method to research the gold price. Mali and Mukhhopadhyay [5], and Bolgorian and Gharli [6] analysed the long-range correlation and fluctuation by multifractual detrended fluctuation analysis approach (MF-DFA) respectively. Similarly, Mills [7] used the detrended fluctuation analysis (DFA) to analyse daily gold price series. Yurdakul and Sefa [8] identified four main variables of the gold price in the Istanbul gold exchange by exponential generalized autoregressive conditional heteroscedasticity (EGARCH). Ntim *et al.* [9] revealed that macroeconomic variables can explain the behaviour of the return sequence of the gold price. Meanwhile, Li [10] combined wavelet neural network (WNN) with artificial bee colony (ABC) to predict the gold price. Yu [11] researched the gold price data by visibility graph network (VGN), and pointed out that the return sequence of log-gold series is a multifractal Gaussian noise sequence. However, few references analysed gold price based on the internal structure of the gold price data. At the same time, the existing research techniques lack theoretical evidence and the combination of multivariant methods. In this paper, we present a novel approach to analyse the gold price based on VMD, PCA and ICA.

VMD method was first proposed by Dragomiretskiy and Zosso [12]. It is an adaptive and non-recursive signal decomposition approach. The primary purpose of VMD is to decompose observed signals into the sum of IMFs (here, the number K of IMFs is pre-determined). Different from EMD [13], VMD is sustained by an adequate theory and can generate better decomposition results. At the same time, it can tremendously avoid modal aliasing and boundary effect. VMD method has been applied to solve many practical problems. Liu *et al.* [14] put forward signal denoising technique based on VMD and detrended fluctuation analysis (DFA-VMD), which demonstrates that VMD-based method is superior than EMD-based approach. Abdoos [15] and Zhang *et al.* [16] introduced VMD to predict the short-term wind power respectively. Additionally, VMD can also be applied to forecast financial temporal series. Lahmiri [17] combined VMD with back-propagation neural network (BPNN) to forecast stock price. Wang *et al.* [18] employed VMD to detect rub-impact faults of the rotor system.

PCA, first introduced by Pearson [19], is a dimension reduction technique, which transforms multiple variables into fewer variables. The main aim of PCA is to project data from a high-dimension space to lower-dimension space in which each new variable is called principal component (PC). Zhang *et al.* [20] combined PCA with artificial neural networks (ANN) to solve the selection problem of quantification in overlapped capillary electrophoresis peaks. Subsequently, Polat *et al.* [21] utilized PCA in medicine.

ICA was proposed by Comon [19]. It is a statistical technique which can be used to reveal the underlying components of random variables, observed data or signals. In recent two decades, ICA was extensively applied to the fields of medicine, image signal processing neural network *etc.* Naganawa *et al.* [22] applied ICA in medicine to extract the plasma time-activity curve from the dynamic brain PET images. Fuzzy support vector machine (FSVM)-ICA was presented by Liu *et al.* [23] to recognize the face feature. Lu [24] used ICA to predict stock price based on denoising scheme with neural network.

The residue of the paper is shaped as follows. Section 2 contains some basic theories such as VMD, PCA and ICA methods. Section 3 proposes a novel approach which is called VMD-PCA-ICA. Section 4 is the description of experiments, including the description of data, the decomposition of the gold price, the extraction of PCs, separation results by ICA, trend analysis, robust regression analysis and performance comparison. The conclusion of the paper is given in Section 5.

2. Theoretical framework

2.1. Variational mode decomposition

Empirical mode decomposition (EMD) is an adaptive technique of signal decomposition which has been proposed by Huang *et al.* [13]. Shi *et al.* [25] integrated EMD with cascaded multi-stable stochastic resonance system (CMSRS) to extract signal features. However, there exists the phenomenon of modal aliasing and boundary effect in EMD. To overcome the issues, Wu and Huang [26] presented Ensemble EMD (EEMD), which adds a Gaussian white noise series to the original signal in the process of using EMD. Zhang *et al.* [27] applied EEMD to decompose the crude oil price. Although the EMD has been improved, the problem of modal aliasing still exists. Dragomiretskiy and Zosso [12] proposed a new method — VMD, which solves the drawback of EMD and EEMD. VMD has been successfully introduced to the fields of analysis and forecasting of finance by Lahmiri [28]. In this paper, VMD method is employed to decompose the gold price data.

VMD is a novel signal processing technique for adaptive and non-recursive decomposition. The purpose of VMD is to get the discrete IMF u_k from the original signal $x(t)$, where u_k has limited bandwidth and compacted around a center pulsation ω_k, determined along with the decomposition. The constrained variational issue is a minimum problem as follows [12]:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] * e^{-j\omega_k t} \right\|_2^2 \right\}, \text{s.t.} \sum_{k=1}^{K} u_k(t) = x(t)\,, \tag{1}$$

where $\delta(t)$ is impulse function, $*$ represents convolution and K is the number of IMF; $\{u_k\} := \{u_1, u_2, \cdots, u_k\}$ and $\{\omega_k\} := \{\omega_1, \omega_2, \cdots, \omega_k\}$; $j = \sqrt{-1}$.

The minimum problem can be shifted into unconstraint problem by the method of the augmented Lagrangian multipliers. In this connection, quadratic penalty factor α is introduced as the balance parameter of original signal fidelity constraint. At the same time, Lagrangian multiplier λ is needed. Then, the augmented Lagrangian function L can be expressed as [12]

$$\begin{aligned} L\left(\{u_k\}, \{\omega_k\}, \lambda\right) \;=\; & \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| x(t) - \sum_k u_k(t) \right\|_2^2 + \left[\lambda(t), x(t) - \sum_k u_k(t) \right]. \end{aligned} \tag{2}$$

Here, $\|x(t) - \sum_k u_k(t)\|_2^2$ is quadratic penalty term for accelerating the velocity of convergence.

In order to solve equation (2), alternate direction method of multipliers (ADMM) is introduced to find the final solutions [29]. The optimal solution of equation (2) can be acquired by updating iterate the $\hat{u}_k^{n+1}(\omega)$, ω_k^{n+1}, and $\hat{\lambda}^{n+1}(\omega)$. Here, n represents the number of iterations.

Dragomiretskiy and Zosso [12] has given the algorithm of VMD which can be seen in Algorithm 1.

2.2. Principal component analysis

PCA is a signal analysis method based on linear algebra, it is widely applied in statistical data analysis, feature extraction and data compression. The basic assumption of PCA is that the raw data is centralized. PCA can linearly transform the p-dimension vector X into q-dimension vector Y ($q \leq p$). The process of using PCA is to maximum project variance of X in the new coordinate axis, such that the first axis corresponds to the maximum

Algorithm 1 VMD optimization

set $\{\hat{u}_k^1\}, \{\omega_k^1\}, \hat{\lambda}_1, n \leftarrow 0$
repeat
  $n \leftarrow n+1$
  for each $k = 1 : K$ **do**
    Update $\hat{u}_k$, $\omega \geq 0$:

$$\hat{u}_k^{n+1}(\omega) \leftarrow \frac{\hat{x}(\omega) - \sum_{i<k} \hat{u}_i^{n+1}(\omega) - \sum_{i>k} \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \tag{3}$$

    Update ω_k:

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 \mathrm{d}\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 \mathrm{d}\omega} \tag{4}$$

  end for

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau(\hat{x}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)) \tag{5}$$

until convergent condition: $\sum_k \frac{\|u_k^{n+1} - u_k^n\|_2^2}{\|u_k^n\|_2^2} < \epsilon$

variance and the second axis orthogonalizing to the first axis corresponds to the maximum variance.

Therefore, PCA is a dimension reduction technique. The main ideology is to find a rotational orthogonal coordinate system, and to ensure that new vector Y reaches the maximum projection variance on the new coordinate axis. In detail, Statheropoulos and Dong *et al.* gave the algorithm of PCA in references [30, 31].

2.3. Independent component analysis

ICA is a data processing technique for separating the source signals from the mixed signal [19]. The main purpose of ICA is to recover the potential ICs. Let $X = (x_1, x_2, \cdots, x_m)^T$ be mixed signals, and the size of each component x_i be $n \times 1$. Suppose X is the linear combination of $s_i, i = 1, 2, \cdots, m$, and the size of each source signal s_i is $n \times 1$. Hyvärinen *et al.* [19] has defined the basic ICA model as follows:

$$X = AS\,, \tag{6}$$

where $S = (s_1, s_2, \cdots, s_m)^T$, A is mixed matrix and we have no knowledge about A and S.

ICA method can be identified based on some hypotheses and ambiguity factors [32]. Firstly for each source signal s_i exists at most one Gaussian distribution. Secondly, the s_i are mutually-independent. Finally, for convenience, an unknown mixed matrix is assumed as a square matrix. In ICA method, there exist two ambiguity factors. On the one hand, ICs' variance (or energy) cannot be confirmed. On the other hand, we have no ability to determine the order of ICs. Hyvärinen *et al.* [19] pointed out that ICA model is aimed to find a separation matrix B such that

$$Y = BX = BAS\,. \tag{7}$$

Here, $Y = (y_1, y_2, \cdots, y_m)^T$, y_k and y_l ($k, l = 1, 2, \cdots, m$, and $k \neq l$) are statistically-independent. Y is the estimation of S and $\hat{s}_i = y_i$ iff $B = A^{-1}$.

The solution of equation (7) can be acquired through maximizing non-Gaussianity cost function. Hyvärinen and Oja [19, 32] introduced the approximation of negentropy and kurtosis as in equations (8) and (9) respectively

$$J(y) \;\propto\; [E\{G(y)\} - E\{G(\nu)\}]^2\,, \tag{8}$$

$$\mathrm{kurt}(y) \;=\; E\left\{y^4\right\} - 3\left(E\left\{y^2\right\}\right)^2\,, \tag{9}$$

where y is a random variable, $E\{y\} = 0$ and $\mathrm{Var}\{y\} = 1$, $\nu \sim N(0,1)$. Normally, non-linear function $G(y)$ can be chosen as follows:

$$G(y) = y \exp\left(-\frac{y^2}{2}\right)\,. \tag{10}$$

3. The hybrid methodology

Hyvärinen *et al.* [19] pointed out that the temporal series was constructed by a variety of reciprocally ICs. In this paper, the gold price is considered as temporal series. What is more, there are a lot of analytical methods for time series. Rounaghi and Zadeh [33] analysed S&P 500 and London stock exchange by autoregressive moving average (ARMA) method. Zhang *et al.* [27] researched the crude oil price based on EMD and EEMD respectively. Lahmiri [28] used VMD approach to analyse and forecast financial time series. However, with the advent of the era of big data, the composition of time series is becoming more complicated. The single analysis method could not function well any more. Simultaneously, it cannot adequately mine the internal structure of temporal series. Based on this, academic circle tends to research the temporal series with a combination technique. EEMD-ICA method was introduced by Mijović *et al.* [34]. Subsequently, Xian *et al.* [35] introduced EEMD-ICA method to investigate international gold price.

Meanwhile, Wang *et al.* [36] combined the ARMA with generalized autoregressive conditional heteroskedasticity (GARCH) to examine stock indices. Zhang *et al.* [16] integrated VMD and machine learning method to predict short-term wind power generation.

In this paper, a novel approach is presented to research the international gold price. The proposed method is a hybrid model of VMD, PAC and ICA which is called VMD-PCA-ICA. It directly projects the single channel mixed signal to the high-dimension space, and then recovers the source signals by the ICA method. The framework of the VMD-PCA-ICA consists of three modules which can be seen in Fig. 1

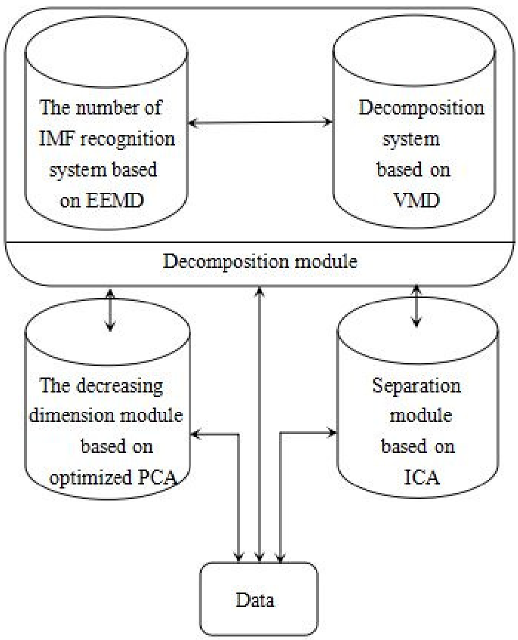

Fig. 1. The framework of VMD-PCA-ICA.

The first one is decomposition module. In this module, the number N of IMFs must be fixed previously. EEMD is applied to intelligently identify N. Then, we can get the picture of the internal structure of the gold price from $u_i(t), i = 1, 2, \cdots, N$. Due to the different influence of the IMFs on the original signal, we can regroup the IMFs by contribution coefficient of the N IMFs. In this paper, we choose relative hamming distance to calculate the contribution coefficient of the i^{th} IMF, and the formula is defined as [35]

$$\mathrm{CC}_i = 1 - \frac{1}{T-1}\sum_{t=1}^{T-1} r(t)\,. \tag{11}$$

Here,

$$r(t) = \begin{cases} 1, & (x(t+1) - x(t))(\hat{x}(t+1) - x(t)) \\ 0, & \text{else} \end{cases} \tag{12}$$

and $\hat{x}(t) = \sum_{j=1, i \neq j}^{N} u_j(t)$. In order to regroup IMFs $\{u_i(t), i = 1, 2, \cdots, N\}$ according to CC_i, we can give a hard threshold β (normally, β is 0.2 or 0.3). Here, $\beta = 0.3$ is applied. If the CC_i of IMFs is less than β, we sum up these IMFs as a new mode function. Furthermore, we can get the M RIMFs $\{u_j(t), j = 1, 2, \cdots, M, M < N\}$.

The second one is decreasing dimension module. Applying PCA to RIMFs, we can obtain the K PCs $\{y_k(t), k = 1, 2, \cdots, K, K < M\}$.

The final module is to separate out ICs by using ICA method. On the other hand, the novel approach VMD-PCA-ICA comprises of the following five steps:

Step 1. EEMD is applied to intelligently identify the number N of IMFs.

Step 2. Apply VMD method to the original data $x(t)$ and obtain IMFs $\{u_i(t), i = 1, 2, \cdots, N\}$, such that

$$x(t) = \sum_{i=1}^{N} u_i(t)\,. \tag{13}$$

Step 3. According to contribution coefficient CC_i between $u_i(t)$ and $x(t)$, we reconstruct a set of RIMFs $\{u_j(t), j = 1, 2, \cdots, M, M < N\}$. The original signal $x(t)$ satisfies the following equation:

$$x(t) = \sum_{j=1}^{M} u_j(t)\,. \tag{14}$$

Step 4. PCA technique is used to decrease the dimension of the RIMFs. The PCs $\{y_k(t), k = 1, 2, \cdots, K, K < M\}$ can explain RIMFs $\{u_j(t), j = 1, 2, \cdots, M\}$ at least in 90 percent.

Step 5. ICs $\{s_k(t), k = 1, 2, \cdots, K\}$ are separated out by the ICA method. And the original data $x(t)$ can be estimated by the linear combination of ICs as follows:

$$\hat{x}(t) = \sum_{k=1}^{K} b_k\, s_k(t)\,, \tag{15}$$

where b_k is the combination coefficient of $s_k(t)$, and it is calculated by the sum of k^{th} column of B^{-1}.

4. Empirical results

This section contains the following three parts: Firstly, we analyse the results which are obtained from the method of VMD-PCA-ICA. Secondly, robust regression analysis verifies that our analysis is of statistical significance. Finally, we compare our proposed approach with EMD-ICA, EEMD-ICA and VMD-ICA methods.

4.1. Data description

In this subsection, the analysed gold price data comes from the World Gold Council (WGC, `http://www.gold.org/statistics`). Throughout the trend of the gold price, it has relatively large fluctuations at the beginning of 1980s. For example, the gold price reached $675.30 per ounce in January 1980. However, the gold price decreased to the annual average of $317.26 per ounce in 1985. Put it another way, the gold price has gone down by 48.2 percent over the five years. In July 1999, the gold price had reached the lowest point $256.08 per ounce in the history. Nevertheless, over the next twelve years, it had a steady increasing trend. In September 2011, the gold price was traded at highest in the history point of $1771.88 per ounce. In the following three years, it exhibited a declining trend.

Given the facts above, the monthly gold price data from January 1980 to December 2016 have been collected to consider as the original signal and the gold price time series is described in Fig. 2.

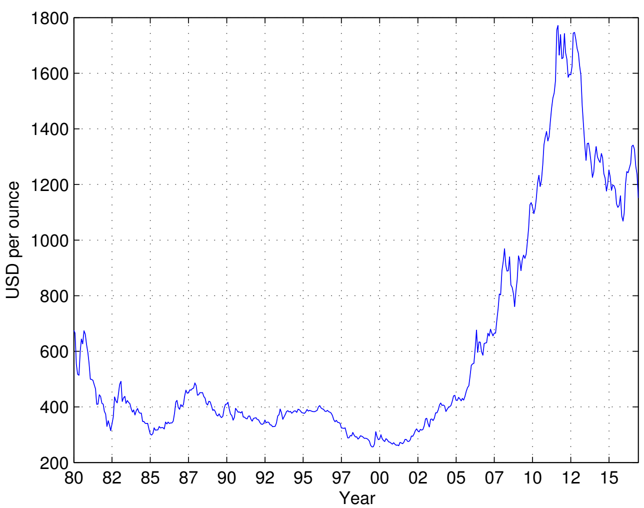

Fig. 2. The monthly gold price data from January 1980 to December 2016.

4.2. Decomposed results

We apply EEMD method to the gold price $x(t)$ data from January 1980 to December 2016. The decomposed results are illustrated in Fig. 3. We obtain IMF1 to IMF7 and a residual function (RF) by EEMD method. Therefore, the number N of IMFs is fixed as 8. Hence, the IMFs series which are obtained by VMD method are plotted in Fig. 4. We conclude that the gold price $x(t)$ can be decomposed into 8 IMFs $\{u_i(t), i = 1, 2, \cdots, 8\}$.

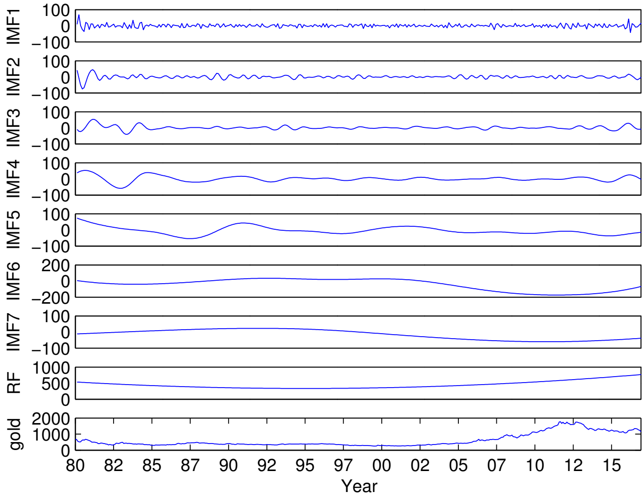

Fig. 3. The decomposed results by EEMD for monthly gold price data from January 1980 to December 2016.

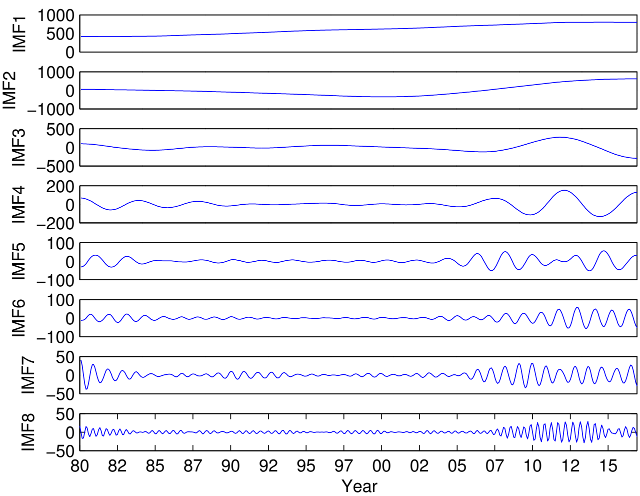

Fig. 4. The decomposed results by VMD for monthly gold price data from January 1980 to December 2016.

By reference [12], each $u_i(t)$ is a portion of the original data $x(t)$. According to equation (11), the contribution coefficients CC_i between $u_i(t)$ and $x(t)$ are listed in Table I.

TABLE I

Contribution coefficients between IMF and gold price.

MFs	$u_1(t)$	$u_2(t)$	$u_3(t)$	$u_4(t)$	$u_5(t)$	$u_6(t)$	$u_7(t)$	$u_8(t)$
CC_i	0.4853	0.4831	0.3950	0.3905	0.2844	0.2370	0.2257	0.2053

From Table I, $u_1(t)$ shows the strongest contribution coefficient with the gold price as 0.4853. $u_2(t)$ shows the relatively weak contribution compared with $u_1(t)$. The last four IMFs present weak contribution coefficient with the gold price to be lower than 0.3. According to the realignment regulation of IMFs, we sum them up as a new IMF. The RIMFs are illustrated in Fig. 5.

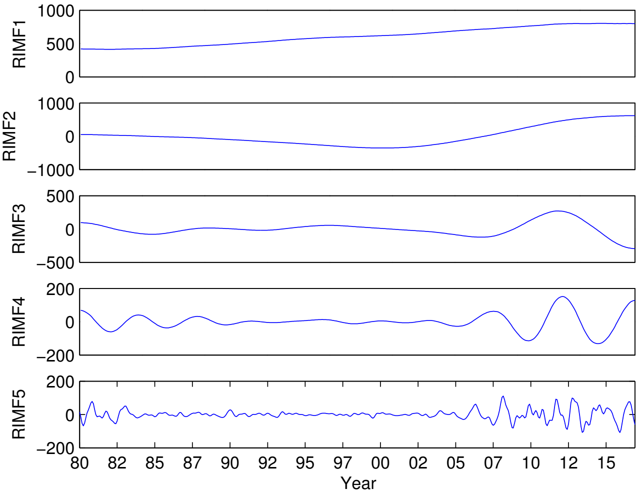

Fig. 5. RIMFs for monthly gold price data from January 1980 to December 2016.

There are a lot of factors which can affect the gold price. However, what attract us are the dominant factors. Consequently, the PCA method is used to reduce the number of RIMFs. Based on Subsection 2.2, we calculate the eigenvalues, eigenvectors, contribution ratios and accumulative contribution ratios in Table II.

From Table II, the cumulative contribution ratio of the first four $\mathrm{AC}R_4$ reaches 90.5735 percent, and it belongs to the interval $[80\%, 100\%]$. Based on this fact, we reduce the dimension of the IMFs $u_j(t)$ from 5-dimension to 4-dimension. The PCs $\{y_k(t), k = 1, 2, 3, 4\}$ can be obtained and depicted in Fig. 6.

TABLE II

The eigenvalue of new mode functions.

t_i	t_1	t_2	t_3	t_4	t_5
$v_1(t)$	0.8630	−0.1158	0.0484	−0.0614	−0.4854
$v_2(t)$	0.8628	−0.1217	0.0457	−0.0554	0.4854
$v_3(t)$	0.2261	0.5808	−0.4137	0.6636	−0.0011
$v_4(t)$	0.0831	0.7237	−0.1811	−0.6607	0.0038
$v_5(t)$	0.0310	0.4337	0.8819	0.1819	0.0018
λ_i	1.5482	1.0775	0.9862	0.9168	0.4713
$CR_i(\%)$	30.9640	21.5493	19.7246	18.3356	9.4265
$ACR_i(\%)$	30.9640	52.5133	72.2379	90.5735	100.0000

Note: t_i is eigenvector. $CR_i(\%)$ and $ACR_i(\%)$ represents the contribution ratios and accumulative contribution ratios respectively.

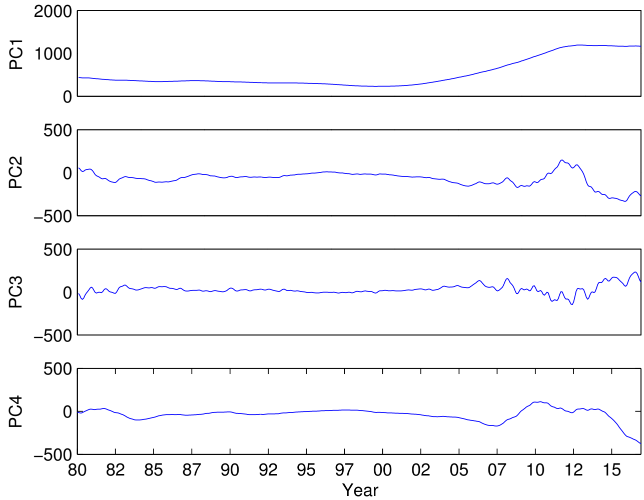

Fig. 6. PCs from the RIMFs.

4.3. Separated results

In this subsection, ICs $\{s_k(t), k = 1, 2, 3, 4\}$ are separated from the PCs. According to the ICA algorithm, we have output the separating matrix B in equation (16) and the gold price series can be estimated in equation (17). The separated ICs are shown in Fig. 7.

$$B = \begin{pmatrix} -0.0029 & -0.0095 & 0.0049 & 0.0027 \\ 0.0005 & 0.0150 & 0.0280 & 0.0020 \\ 0.0013 & -0.0068 & 0.0096 & 0.0127 \\ 0.0011 & -0.0011 & -0.0030 & -0.0123 \end{pmatrix}, \tag{16}$$

$$\hat{x}(t) = -225.6332 s_1(t) + 15.8043 s_2(t) + 194.1609 s_3(t) + 72.6075 s_4(t)\,. \tag{17}$$

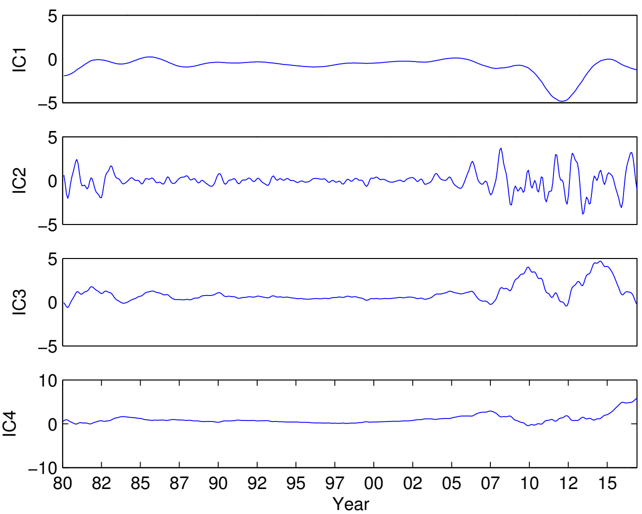

Fig. 7. ICs for monthly gold price data from January 1980 to December 2016.

For accurately illustrating the underlying meaning of ICs, it is necessary to analyse data characteristics. Therefore, we choose the ICs' basic statistics, including the mean, skewness, kurtosis, adjust Jarque–Bera statistics, Pearson correlation coefficients, Hurst exponent and Phillips–Perron statistics. These statistics are calculated and listed in Table III.

The mean is an indicator that reflects the central tendency of ICs, and it can be used to simply express the differences between groups.

The normality of temporal series can be analysed through the skewness, kurtosis and adjusted Jarque–Bera statistics. The skewness is an indicator of the distribution of symmetry. When the skewness is greater than 0, the heavy tail is on the right. When the skewness is lesser than 0, the heavy tail is on the left. The kurtosis is used to describe the index of top flat or sharp. When the value of kurtosis is greater than 3, the distribution curve has positive kurtosis (peakedness). When the value of kurtosis is less than 3, the distribution curve has negative kurtosis (flatness). The adjusted

TABLE III

The basic statistics of ICs.

ICs	S	K	AJB	ρ_i	b_i	H	PP(h)
IC1	1.5705	6.0500	264.5720	−0.7535	−225.6332	0.9455	−1.0136(0)
IC2	−2.5812	9.5675	177.8004	0.0306	15.8043	0.5165	−5.0354(1)
IC3	1.8915	6.2552	472.7922	0.5373	194.1609	0.8767	−0.8767(0)
IC4	−0.1650	6.0173	1.32e+03	0.3760	72.6075	0.9757	3.6779(0)

Note: S and K represent the skewness and kurtosis respectively. AJB is adjusted Jarque–Bera statistics. ρ_i represents Pearson correlation coefficient, b_i represents combination coefficient, H Hurst exponent, absolute value method is used. PP(h) is Phillips–Perron statistics which is used to test the unit root, where h demonstrates whether the temporal series is stable. $h = 0$ shows that we fail to reject to null hypothesis (non-stationarity), while $h = 1$ shows that we reject to null hypothesis (stationarity). Here, null hypothesis exists unit root.

Jarque–Bera statistics has faster convergence speed than the Jarque–Bera statistics, which is used to test whether ICs can be viewed from normal population. Generally, when the value of the adjusted Jarque–Bera statistics tends to 0, ICs can be viewed from normal population. Conversely, when it is greater than 11, ICs can be viewed from non-normal population. In light of Table III, IC1 and IC3 have heavy tails on the right, while IC2 and IC4 have heavy tails on the left, and all ICs are positive kurtosis with sharp peaks. Furthermore, the values of adjusted Jarque–Bera statistics of all ICs are greater than 11. This is statistical evidence to show that ICs come from the non-Gaussian distribution. Hence, ICA is a reasonable alternative in our proposed approach since it satisfies the basic hypothesis.

Pearson correlation coefficients are used to measure the linear correlation between ICs and gold price sequence. The value of correlation coefficient belongs to $[-1, 1]$, when it is greater than 0, it is a positive correlation, when it is less than 0, it is a negative correlation. Table III shows that IC2 exhibits a weak linear correlation with gold price, while others ICs have certain correlation with gold price. Because of the two ambiguity factors in ICA method, the symbol of correlation coefficient and the variance of ICs are neglected.

The Hurst exponent reflects the memory of temporal series. Normally, the value of Hurst exponent of random time series tends to 0.5. When Hurst exponent is close to 1, it means that the temporal sequence has long memory. From Table III, IC1, IC3 and IC4 have long memory, while IC2 has relatively weak memory.

The stabilization of temporal series is tested by the Phillips–Perron statistics which is one of the unit root test. The results of test for the unit root in Table III show that the IC2's Phillips–Perron statistics value is -5.0354. According to Table III, $h = 1$, so we reject to null hypothesis (exists unit root). Therefore, IC2 is the unique stationarity series.

In Fig. 8, the dotted line is the estimation of the gold price data, because the PCA process selects 4 as the number of PCs, and the variance explained percentage is 90.5735. Thus, the estimated results shown in equation (17) reflect 90% information at least. The clear result of $\hat{x}(t)$ can be acquired by improved approach VMD–PAC–ICA, and it is meaningful and convenient for scholars and researchers to analyse the gold price. In practice, ICs $s_k(t)$ in equation (17) attached underlying implications in terms of specific circumstance, which will be interpreted in the next subsection.

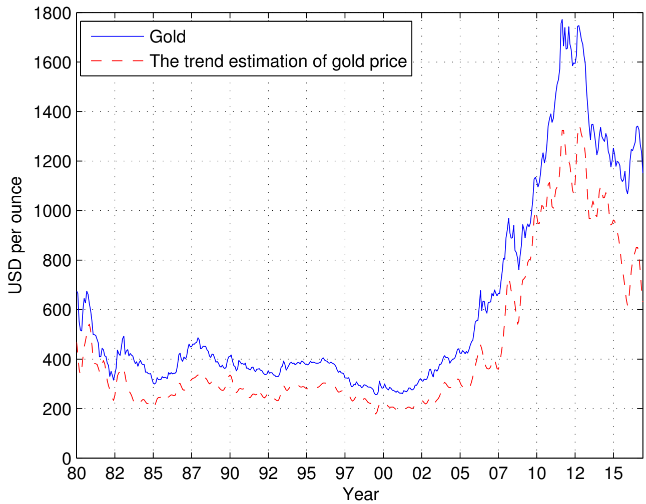

Fig. 8. The estimation of monthly gold price data from January 1980 to December 2016 by VMD-PCA-ICA.

4.4. Trend analysis of ICs

Since the huge fluctuation of the gold price sequence in the past decades, it has attracted the attention of academic, portfolio managers, regulatory authorities *etc.* There exists many factors affecting the volatility of the gold price. Lin *et al.* [37] studied the relationship between the US dollar (USD) and the gold price. Chen and Lin [38] pointed out that the stock would affect the gold price. Xian *et al.* [35] separated three types of influence factors of the gold price series via EEMD-ICA method, including long-term trend influence factors, cyclical effects and stochastic market factors. Nistor and

Ciupacalici [39] concluded that the crude oil price and consumer price index (CPI) would influence the gold price. In this subsection, we use VMD-PCA-ICA method to investigate the influence factors of the gold price.

4.4.1. IC1

From Table III, IC1 has the largest correlation with the gold price, and it has long memory feature in light of Hurst exponent with the value being 0.9455. By analysing the index of global GDP from The World Bank (WB, `http://data.worldbank.org.cn`), the comparison of IC1 and the growth rate of GDP is depicted in Fig. 9. Since the long-term trend is being analysed, we take the year as the time point. Based on this, we observed the growth rate of GDP exhibited the negative tendency in 1982, 1997, 1998, 2001, 2009, 2015, wherein the growth rate of GDP reached the relatively low level of −5.21% and −5.69% in 2009 and 2015 respectively. It demonstrates that the global economy was in recession. Correspondingly, IC1 exhibits the fluctuation tendency of decreasing in these years. Especially, in 1982, 2009 and 2015, IC1 presented a relatively decreasing trend. Furthermore, in the next subsection, we use regression analysis to verity this comparison.

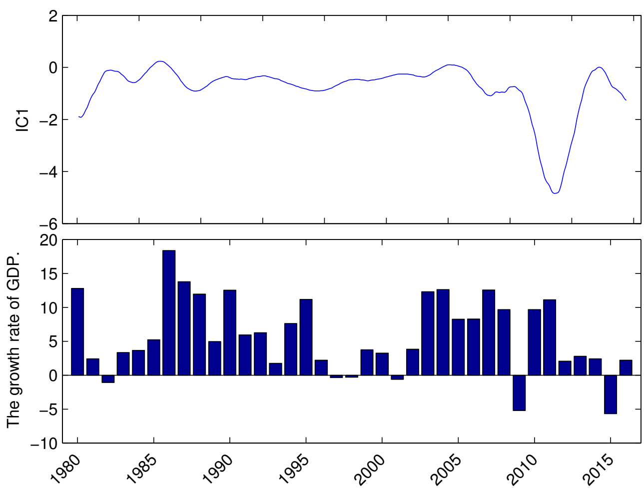

Fig. 9. The annually growth rate of GDP data from 1980 to 2016.

4.4.2. IC2

The trend of IC2 exhibits objectively statistical regularity of randomness. In terms of the gold price series, the large volatility appeared from 1980 to 1985 and from 2005 to 2016. Correspondingly, IC2 rises and falls considerably. These changes are the same with IC2. From Table III, the value of

Hurst exponent is 0.5165 which shows that IC2 is a random time series. The correlation coefficient between IC2 and gold price is 0.0306 which demonstrates that the linear relationship between IC2 and gold price is not obvious. This result suggests that the emergencies and international events have an impact on the gold price. For instance, international wars, financial crisis, hideous attacks *etc.* They contribute to the diverse tendency of increments or declines. The comparative plot is shown in Fig. 10.

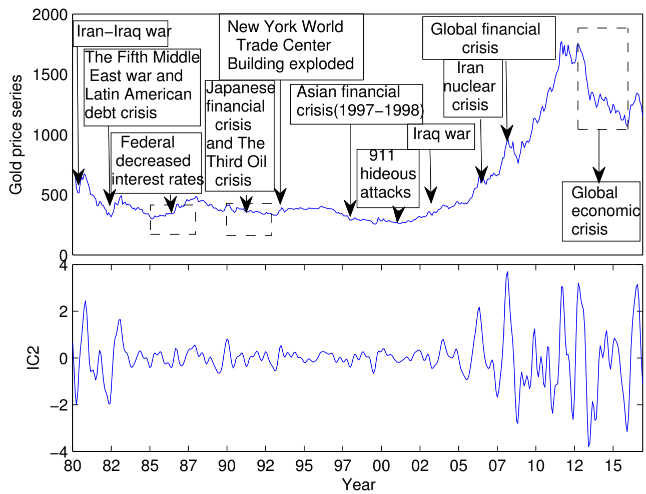

Fig. 10. The trend of IC2 is compared with the gold price series from January 1980 to December 2016.

Generally, the international wars, hideous attacks and financial crisis made the gold price increase. They play a significant roles in the fluctuation of the gold price. When the wars or hideous attacks happened, the gold price increased. For example, during the war between Iraq and Iran in 1980, gold price was up to \$675.30 per ounce. During the Fifth Middle East war, the gold price reached \$491.96 per ounce. After the 911 hideous attacks (Afghanistan war) in 2001, the gold price increased from \$265.49 per ounce in January 2001 to \$281.65 per ounce in January 2002, and the growth of rate reached 6.09 percent. In 2003, the Second Gulf war (Iraq war) spurred gold price to increase.

The crisis events result in the decline of the gold price. For example, Latin American debt crisis had made the gold price decline five years successively, and the gold price tailed to below \$300 per ounce in February 1985. Until the Federal Reserve System decreasing interest rates from 1985 to 1987, the price of gold was picked up. Japanese financial crisis in 1990, the Third Oil crisis in 1991 and a global financial crisis made the gold price descend at different degree. From the beginning of 2012, the gold price was affected by the European debt crisis and improvement in the global economy. Gold's storage function dropped and the gold price presented the tendency of decreasing.

4.4.3. IC3

IC3 is an unstable temporal series in light of the Phillips–Perron test for the unit root. The combination coefficient of IC3 reaches -194.4175, the Hurst exponent is 0.8767, which implies long memory of IC3. We have analysed annually supply and demand data from 1995 to 2016, and the data is derived from WGC (`http://www.gold.org.com/`). The relationship between supply and demand is reflected by supply minus demand in Fig. 11. The compared result illustrates that the volatility of IC3 is accordant with the relationship between supply and demand. It is an evidence that IC3 reflects the fluctuation of supply and demand trend in gold. Furthermore, in the next subsection, we use regression analysis to verity these comparison.

As a matter of fact, the gold price will decrease under the circumstance of the supply exceeding the demand. For example, before 2003, the value of supply minus demand is positive, which shows the supply of gold is sufficient and the gold price is relatively low. Correspondingly, the trend of IC3 is weak. However, at the beginning of 2004, with the increasing demand for gold, the production of gold cannot keep up with the pace of demand, which leads to the increment of the gold price. In conclusion, the declining trend of the relationship between supply and demand corresponds to the growth of the gold price through contrastive analysis.

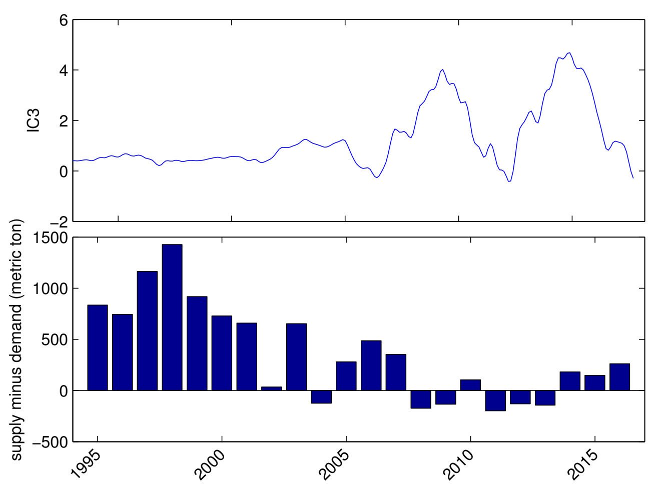

Fig. 11. The annually supply and demand data from 1995 to 2016.

4.4.4. IC4

From Table III, the value of Hurst exponent is 0.9757, which implies long memory of IC4. In equation (17), the combination coefficient of IC4 is 72.6075. IC4 comes from the non-Gaussian population in light of skewness, kurtosis and adjust Jarque–Bera statistics. The experiment suggests that the trend of IC4 is evidently in line with the DJS data from the Dow Jones Company (`http://www.dowjones.com/`). The comparison of IC4 and DJS is shown in Fig. 12. Hence, the potential formation mechanism of the gold price includes the DJS.

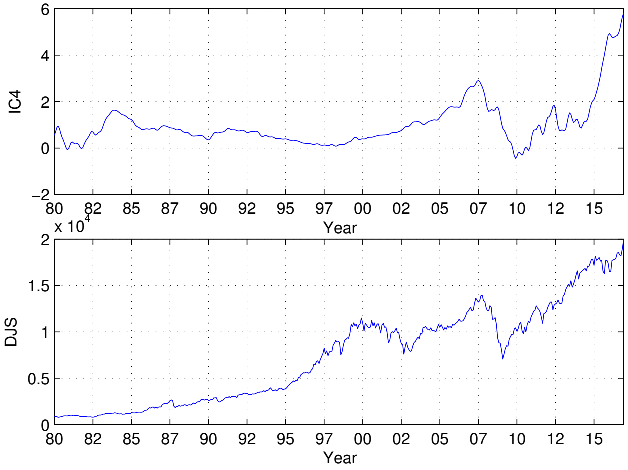

Fig. 12. IC4 compares with DJS time series from January 1980 to December 2016.

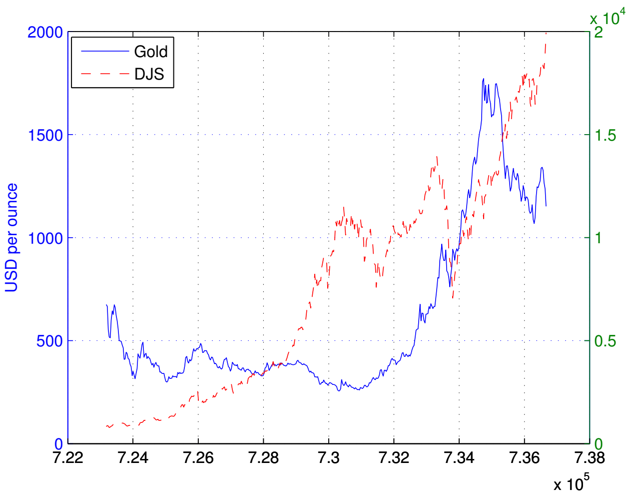

Fig. 13. The comparison of DJS with gold from January 1980 to December 2016.

In fact, DJS is at present extensively utilized in stock index. The stock and gold are most important in the portfolio market. The relationship between DJS and gold price is complicated. Due to the competitive nature, the market of stock and gold takes on negative correlation in Fig. 13. When the stock market is in the flourishing epoch, portfolio managers tend to invest fund in the stock with high returns. Furthermore, cash flows to stock market, which leads to the decline of the gold price. Whereas, when the stock market is in doom with depressing macroeconomic, the hedging property of gold drive cash backflowing to the gold market, which results in the rise of the gold price. In the following, we use regression analysis to verity these comparisons.

4.5. Robust regression analysis

In this subsection, the correlation between ICs and economic interpretation is invested and proved. The robust regression is less influenced by the outlier than the ordinary least-squares regression (OLS) [40]. We choose the robust regression analysis in our paper. Regressing each ICs on the correlative economic factors (except IC2, owing to accidental events and high randomness in it), including the growth rate of GDP, supply minus demand of gold and DJS index. They represent economic development, gold supply and demand, and stock market, respectively [33, 35, 41]. The robust regression analysis results are shown in Table IV.

TABLE IV

The robust regression of the ICs on the economic factors.

ICs	$C1$	$C2$	R^2	Adj-R^2	Sig.	F
IC1-GDP	-0.4609	-0.0055	0.4770	0.3795	$2.7313e-18$	31.8843**
IC3-SMD	0.8392	-0.0004	0.5543	0.4009	0.0015	4.3057**
IC4-DJS	0.5229	$3.8925e-05$	0.5819	0.5790	$2.2270e-80$	27.1962**

Note: $C1$ and $C2$ represent intercept and coefficient in the robust regression model; Adj-R^2 is adjusted R^2; Sig. is confidence level; F is F-statistics; GDP, SMD and DJS represent the growth rate of GDP, supply minus demand of gold and DJS index respectively; ** $p < 0.01$.

From Table IV we can see that the ICs on relevant economic factors exhibits statistical significance based on the F-statistics. At 1% confidence level, the coefficients of robust regression analysis model exhibits statistical significance. The above results provide the statistical evidence for the linear relationship between ICs and relevant economical factors, which implies the results of our analysis are correct.

4.6. Performance comparison

The separation effect of the observed signals largely depends on the selection of the decomposition method. EEMD remedies the deficiency of the EMD method, and it can efficiently solve the boundary effect. However, there exist some problems in the application of the EEMD-ICA, such as high iteration times and low convergence speed. Dragomiretskiy and Zosso [12] proposed the VMD method, which can effectively handle the above problems. In our proposed approach, we substituted original EMD and EEMD method with VMD, and PCA preprocessing technique is introduced to analyse the gold price. Then, hidden ICs are separated from the gold price by ICA method. The comparison results of EMD-ICA, EEMD-ICA, VMD-ICA and VMD-PCA-ICA are shown in Fig. 14.

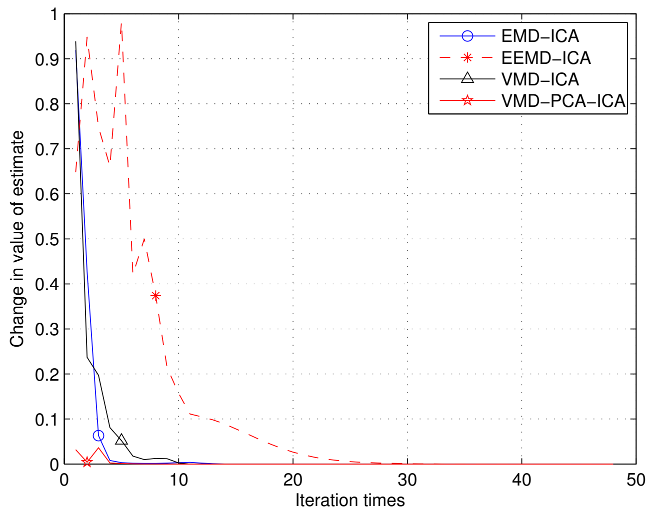

Fig. 14. The comparison results of iteration times of methods.

From Fig. 14, EMD-ICA method reaches convergence after 15 steps. Although EEMD modified the shortage of boundary effect, the phenomenon of mode mixing has not been conquered yet. EEMD-ICA costs 37 steps to converge. In fact, EEMD method adds white noise sequence to the observed signal series during the process of iteration, and it costs many computations to achieve convergence. VMD method sorts out the modal aliasing and boundary effect, and VMD-ICA method reached convergence after 12 steps. Since the dimension reduction technique PCA was introduced in our proposed method, it improves the degree of adaption and makes the integrated approach more intelligent. Meanwhile, it reduces the complexity of ICA process. Whereas, our proposed approach only needs 7 steps to reach the convergence.

5. Conclusion

In this paper, the potential ICs of the gold price are separated out via VMD-PCA-ICA and the underlying economic meanings are interpreted in light of ICs' data feature. More specifically, we identify four influence factors: GDP, sudden great historic events, the relationship between supply and demand, and DJS. Meanwhile, robust regression analysis verifies that the economic meaning of ICs exhibits statistical significance.

In detail, this paper can be summarized as follows: Firstly, we decompose the gold price into IMFs by VMD method and the RIMFs are obtained in light of the contribution coefficient between IMF and gold price data. In order to reduce the dimension of RIMFs, PCA is applied to RIMFs, and the experiment result suggests that RIMFs are explained by PCs at 90.5735%. The gold price series is estimated by the simple linear combination of ICs which are separated from the PCs, which is a fantastic result. Secondly, robust regression analysis has verified that our comparative analysis shows a statistical significance. Finally, compared with EMD-ICA, EEMD-ICA and VMD-ICA, our approach uses the minimum number of iterations to converge, which implies VMD-PCA-ICA outperforms other methods.

The authors would like to thank the editor and the unknown responsible reader for their ideas in improving manuscript sincerely. This work is supported by the National Natural Science Foundation of China — grants No. 11771263 and No. 61573014.

REFERENCES

[1] B.B. Mandelbrot, *Fractals and Scaling in Finance*, Springer, 1997.

[2] F. Wen, X. Yang, X. Gong, K.K. Lai, *Int. J. Info. Tech. Dec. Mak.* **16**, 205 (2017).

[3] M. Dai *et al.*, *Fractals* **24**, 1650046 (2016).

[4] A.G. Malliaris, M. Malliaris, *Finance Res. Lett.* **13**, 45 (2015).

[5] P. Mali, A. Mukhopadhyay, *Phys. Scr.* **90**, 035209 (2015).

[6] M. Bolgorian, Z. Gharli, *Acta Phys. Pol. B* **42**, 159 (2011).

[7] T.C. Mills, *Physica A* **338**, 559 (2004).

[8] F. Yurdakul, M. Sefa, *Procedia Economics and Finance* **23**, 77 (2015).

[9] C.G. Ntim, J. English, J. Nwachukwu, Y. Wang, *Int. Rev. Financ. Anal.* **41**, 218 (2015).

[10] B. Li, *Comput. Intel. Neurosc.* **2014**, 270658 (2014).

[11] Y. Long, *Physica A* **392**, 3374 (2013).

[12] K. Dragomiretskiy, D. Zosso, *IEEE Trans. Signal Process.* **62**, 531 (2014).

[13] N.E. Huang *et al.*, *Proc.: Math. Phys. Eng. Sci.* **454**, 903 (1998).
[14] Y. Liu, G. Yang, M. Li, H. Yin, *Signal Process.* **125(C)**, 349 (2016).
[15] A.A. Abdoos, *Neurocomputing* **203**, 111 (2016).
[16] Y. Zhang, K. Liu, L. Qin, X. An, *Energy Convers. Manage.* **112**, 208 (2016).
[17] S. Lahmiri, *J. Comput. Sci.* **12**, 23 (2015).
[18] Y. Wang, R. Markert, J. Xiang, W. Zheng, *Mech. Syst. Signal Process.* **60**, 243 (2015).
[19] A. Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.
[20] Y. Zhang, H. Li, A. Hou, J. Havel, *Chem. Intell. Lab. Syst.* **82**, 165 (2006).
[21] K. Polat, S. Güneş, *Digital Signal Process.* **17**, 702 (2007).
[22] M. Naganawa, Y. Kimura, K. Ishii, K. Oda, *IEEE Trans. Biomed. Eng.* **52**, 201 (2005).
[23] Y. Liu, G. Chen, J. Lu, W. Chen, *Face Recognition Based on Independent Component Analysis and Fuzzy Support Vector Machine*, in: Proc. of the 6$^{\rm th}$ World Congress on Intelligent Control and Automation, June 21–23, 2006, Dalian, China, DOI:10.1109/WCICA.2006.1713929.
[24] C.J. Lu, *Expert Syst. Appl.* **37**, 7056 (2010).
[25] P. Shi, S. An, P. Li, D. Han, *Measurement* **90**, 318 (2016).
[26] Z. Wu, N.E. Huang, *Adv. Adapt. Data Anal.* **1**, 1 (2009).
[27] X. Zhang, K.K. Lai, S.Y. Wang, *Energ. Econ.* **30**, 905 (2008).
[28] S. Lahmiri, *Expert Syst. Appl.* **55**, 268 (2016).
[29] E. Ghadimi, A. Teixeira, I. Shames, M. Johansson, *IEEE Trans. Autom. Control* **60**, 644 (2015).
[30] M. Statheropoulos, A. Pappa, P. Karamertzanis, H.L.C. Meuzelaar, *Anal. Chim. Acta* **401**, 35 (1999).
[31] D. Dong, T.J. McAvoy, *Comput. Chem. Eng.* **20**, 65 (1996).
[32] A. Hyvrinen, E. Oja, *Neural Networks* **13**, 411 (2000).
[33] M.M. Rounaghi, F.N. Zadeh, *Physica A* **456**, 10 (2016).
[34] B. Mijovi *et al.*, *IEEE Trans. Biomed. Eng.* **57**, 2188 (2010).
[35] L. Xian, K. He, K.K. Lai, *Physica A* **454**, 11 (2016).
[36] W. Wang, Y. Guo, Z. Niu, Y. Cao, in: IEEE International Conference on Industrial Engineering and Engineering Management, 2009, pp. 2143–2147.
[37] F.L. Lin, Y.F. Chen, S.Y. Yang, *Int. Rev. Econ. Finance* **43**, 59 (2016).
[38] A.S. Chen, J.W. Lin, *Appl. Econ. Lett.* **21**, 158 (2014).
[39] I.A. Nistor, M.L. Ciupaculici, *Metal. Int.* **18**, 256 (2013).
[40] A. Preminger, R. Franck, *Int. J. Forecasting* **23**, 71 (2007).
[41] M.B. Devereux, E.G. Mendoza, *J. Int. Econ.* **69**, 1 (2006).