



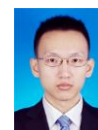
## STUDY ON PRICE FLUCTUATION OF INDUSTRY INDEX IN CHINA'S STOCK MARKET BASED ON EMPIRICAL MODE DECOMPOSITION



Fu Qiao

School of Economics and Management, University of Chinese Academy of Sciences, Beijing, China.

Email: [qiaofu17@mailsucas.edu.cn](mailto:qiaofu17@mailsucas.edu.cn) Tel: +86 13051361751



### ABSTRACT

#### Article History

Received: 17 February 2020

Revised: 27 March 2020

Accepted: 1 May 2020

Published: 28 May 2020

#### Keywords

China's stock market

Stock price

Industry index

Empirical mode decomposition (EMD)

Ensemble empirical mode

Decomposition (EEMD)

Rotation

Principal component analysis (PCA).

#### JEL Classification:

E30; G10.

Rotation is a classic method to describe how securities performance varies with specific factors, such as industry. However, in a short time interval, it is difficult to prove or to disprove the existence of rotation. This paper uses empirical mode decomposition (EMD) and principal component analysis (PCA) in combination to build a framework for analyzing the price fluctuation. Based on this framework, we analyzed the price fluctuations of the China Shenwan (SWS) Industry Index from 2007 to 2019. In the overall fluctuation of the 28 SWS industry index prices, the component that makes all index prices fluctuate in the same direction accounts for more than 75%. The one that makes the fluctuations in index prices varying with the industries accounts for 7%. The former component mainly features the periodic fluctuation, the later one mainly features the aperiodic, trending fluctuations. It means that in a time interval on a decade basis, the inter-industries difference of securities performance should be viewed as the difference of long-term trend rather than the periodic fluctuation (which is usually termed rotation). Therefore, investors who are interested in China's capital market should pay more attention to analyzing the long-term development trend of China's industrial economy, especially the major industrial policy issued by the government.

**Contribution/ Originality:** The paper's primary contribution is finding that in a decade interval, inter-sector heterogeneity of fluctuations in China's stock prices are mainly reflected in secular trends, rather than the cyclical rotation. The empirical mode decomposition and principal component analysis are introduced to build the analyzing framework of this study.

### 1. INTRODUCTION

The value of rotation is that investors can predict how the security prices vary according to the economic environment. In the 1970s, Merrill Lynch put forward the investment clock theory, which is an important analytical paradigm guiding asset allocation with the thought of rotation. The investment clock theory describes the macroeconomic environment through two dimensions, price level and economic development. Then it constructs four quadrants of the environment: recovery, overheating, stagflation and recession. In these four quadrants, the assets that outperform other ones are stocks, commodities, currencies and bonds, respectively. According to the cyclical changes in the macro environment, investors should withdraw their capital from the advantageous assets of the previous stage and invest in the one of the next stage. By doing so, they can earn a solid return on their investment.

In the practice of the capital market, investors often first determine which industries are more promising, and then invest funds in the securities of these promising industries. The development of the industry is closely related to the business cycle. Therefore, based on the business cycle theory, the rotation of the industry is derived. Classical business cycle theory also includes the following: The first one is the Kitchen cycle, which was put forward by the American economist Kitchen. The cycle of inventory from excess to shortage is the basic logic of the Kitchen cycle. Specifically, overproduction creates inventory, which reduces production and leads to shortages. Thus, the Kitchen cycle is called the inventory cycle as well. Empirical analysis shows that the length of the Kitchen cycle ranges from 2 to 4 years.

The second one is the Juglar cycle, which is approximately a cycle of ten years. The Juglar cycle highlights that the economic cycle consists of three stages, namely prosperity, crisis and depression. This cycle is also known as the "real estate cycle" because it is marked by booms and busts in real estate industry.

The last one with the largest time scale is the Codrulieff cycle, which is a kind of 50- to 60-year economic cycle in the macro economy. The Codrulieff cycle attributes the economic cycle to the replacement of large production facilities. At the same time, it emphasizes the reconstruction of production relations and market order by technological progress.

In the capital market, the price fluctuation of securities has both long-term fluctuation and short-term fluctuation. It indicates that price fluctuations are influenced by both long-term factors, such as macroeconomic factors, and short-term factors, such as cash liquidity. The price change of the stock market is a complex multi-factor system and it is necessary to analyze the motivation of the changes systematically in different time scales.

Most economic cycles are longer than five years. That is a bit too long for today's capital markets, where institutional investors are increasingly becoming the main players. Almost all institutional investors such as mutual funds and trusts are evaluated and ranked according to their investment performance within a year. The time span from establishment to liquidation for these institutional investors also tends to be less than five years. In other words, most players tend to stay in the market for no longer than a single economic cycle. Therefore, it is still an open question as to whether cyclical changes or long-term trends can reflect the performance differences of securities in different sectors.

Another analytical framework is to introduce industry factors into the Fama-French multi-factor model to consider price changes caused by industry risk factors. Unlike the theory based on economic cycles, the theory based on risk factors believes that market factors are always the most important drivers of price changes. In contrast, the explanatory power of industry factors to price fluctuations is less than that of market factors. For example, some literature showed that capital flow dominates short-term price changes (Conover, Jensen, Johnson, & Mercer, 2008).

Changes in the prices of securities in different industries can form a leading and lagging system in a statistical sense, which has also been proved to be the driver of short-term fluctuations in stock prices (Granger & Morgenstern, 1963; Kanas & Kouretas, 2007). Besides, the idea that price changes themselves lead to periodic price changes is known as momentum and reversal theory. Jegadeesh and Titman (1993); De Bondt and Thaler (1985) first confirmed that stock price changes have 12-month momentum and 3-year reversal, respectively. At the industry level, the frequency of periodic price fluctuations will be higher, roughly less than one month in the European and American capital market (Hong, Torous, & Valkanov, 2007).

It is reasonable that these studies do not rely entirely on macroeconomic cycles. The periodicity of price changes is so short that the instability of macroeconomic cycles will critically interfere with the analysis of rotation. The ideal periodic motion like sine wave does not exist in reality, and the length and phase changes of the economic cycle are often unpredictable random events. Therefore, it is necessary to introduce appropriate analysis tools to explore the real reasons for the differences in securities price fluctuations in different industries.

This is the motivation for this article.

First, empirical mode decomposition (EMD) is introduced to decompose price sequences of stocks (index) of different industries in China's stock market into sets of subsequences with different frequencies. Then through signal-noise separation, each set of subsequences is reconstituted into the components representing the component of long-term trend, low frequency fluctuation and high frequency fluctuation of the stock index prices change. After that, we can explore the law of industry rotation on different time scales and try to find relevant prediction and decision-making mechanisms.

Capital markets are subject to both fundamentals and liquidity. We believe that the price fluctuation driven by capital flow features of the rise and fall of the whole market in the same direction, while the one driven by industry fundamentals shows the asynchronism of the volatility. The key to our discussion is whether out-of-sync price changes exist in long-term trends or in cyclical fluctuations. In this paper, principal component analysis (PCA), a very important tool in volatility analysis, was used for this discussion.

We found that in the overall fluctuation of the 28 SWS industry index prices, the component that makes all index prices fluctuate in the same direction accounts for more than 75%. The one that makes the fluctuations in index prices varying with the industries accounts for 7%. The former component mainly features the periodic fluctuation, the later one mainly features the aperiodic, trending fluctuations. Our conclusion will serve as a reminder to investors and researchers concerned about China's stock market to pay more attention to analyzing the long-term trend of industry development, especially the industrial development under the guidance of the government's industrial policy. At the same time, the findings of this paper also enrich the research on rotation. Based on EMD, we proposed a framework to adaptively transform the long-period cycle that is difficult to observe in trend motion in a short-term. These are the contributions of this article.

In section 2, we introduce the EMD method and its improved version. In part 3, we present the raw data and some pre-processing after EMD decomposition. In part 4, we show how we reorganized the price subsequences decomposed by EMD into the trends, low-frequency fluctuations and high-frequency fluctuations, and the results of principal component analysis (PCA) are further discussed.

## 2. METHODOLOGY

### 2.1. Empirical Mode Decomposition

The time-frequency analysis methods such as Empirical Mode Decomposition (EMD) and their improvements have been gradually introduced into the study of economic problems. EMD was first proposed by Huang, et al. (1998) with the original purpose of decomposing non-stationary sequences for further frequency-domain analysis of Hilbert transform (therefore, EMD method and Hilbert transform are collectively called Hilbert-Huang transform, a.k.a. HHT). According to EMD, signals can be decomposed into a series of intrinsic mode functions (IMF) to separate fluctuations or trends at different scales in the original information.

Applications of EMD initially focused on natural science, such as measuring frequency-domain characteristics of seismic waves (Lei, He, & Zi, 2009; Wang, Chau, Xu, & Chen, 2015). In these areas, a large portion of the raw data is still a stationary sequence and is generated from linear systems. Therefore, EMD can be used in parallel with Fourier analysis and wavelet analysis. As it was gradually introduced into the study of economic and financial data, its advantages were further developed. This method is involved in data analysis of various capital markets (Al-Hnaity & Abbod, 2015; Goel & Hatzinakos, 2014).

According to Huang, et al. (1998) multiple components of different frequencies can be extracted from the original sequence by EMD method, which are called intrinsic mode function (IMF). IMF should be a locally symmetric sequence, and its extremum should differ from zero by at most one. EMD is an algorithm based on cubic spline function. Specifically, the algorithm steps of EMD method can be expressed as follows:

Step1: with a given the original  $x_0$ , find its upper and lower extrema  $x_{max}$  and  $x_{min}$ .

Step2: through cubic spline interpolation of  $x_{max}$  and  $x_{min}$ , obtain the upper and lower envelope of  $x_0$ , note as  $e_{max}$  and  $e_{min}$ :

Step3 : take the mean value of the upper and lower envelope as  $m = 0.5 * (e_{max} + e_{min})$ , then obtain a new sequence  $h_{1,1} = x_0 - m$ . Note that the  $h_{i,j}$  is one of the candidates of  $IMF_i, \forall j$ , and in order to find out the real  $IMF_i$ , we have repeated from step 1 to step 3 for  $j$  times. Thus, according to the definition, the  $x_i$  is equivalent to  $h_{i+1,0}$ :

Step4 : if  $h_{1,j}$  does not meet the conditions for being an IMF, repeat from steps 1 to step 3 to obtain  $h_{1,j+1}$ . Until a  $J$ , which makes  $h_{1,J}$  satisfies the condition of IMF, is found. Then let  $IMF_1 = h_{1,J}$ , and obtain  $x_1 = x_0 - IMF_1$ ;

Step5 : repeat from step 1 to step 4 to decompose  $IMF_{i+1}$  from  $x_i$ . Until  $x_n$  is a monotone sequence, or at most one local extremum. Then  $r = x_n$  can be viewed as residual, or the trend.

Now as shown in Equation 1 the raw sequence  $x_0$  has been decomposed into the sum of n IMFs and one trend.

$$x_0 = \sum_{i=1}^n IMF_i + r \quad (1)$$

## 2.2. The Limit and the Corresponding Improvement of EMD

In recent years, the research results of EMD are being rapidly enriched. First of all, traditional EMD focuses on time-frequency analysis of a single dimension, while the research of Rilling, Flandrin, Gonçalves, and Lilly (2007) extends EMD technology to joint analysis of two-dimensional and even multiple variables. Secondly, the basic algorithm of EMD technology treats the equal weight of different components, and can only conduct FM and AM within components, not between components. Zheng, Pan, Liu, and Liu (2018) and Wang, Lingyu, Yuyan, and Peng (2017) used different methods to weight different components after EMD decomposition to improve data prediction performance based on EMD.

The improvements to the EMD mainly focus on several key issues. The first one is the stopping criterion of the selection of the IMFs. According to Huang et al. (2003) IMF with ideal properties may not be found when searching by repeating from step1 to step 3 described in section 2.1. Therefore, a proper stop criterion for the IMF selection becomes the key of the EMD algorithm designing. In practice,  $J = 3$  or  $J = 5$  have been set as default stop criterion. This criterion gives a good balance between the preserving the properties of IMFs and improving the computational effectiveness.

The second one is intermittent change. The resonance of different periodic oscillations will generate huge energy, and then lead to intermittent changes. Intermittency would mislead us into spotting changes from the IMF

that do not actually exist. To avoid the pitfalls of intermittence, an interval threshold is required. Only when the distance between a pair of successive extrema is less than the interval threshold, the envelope derived from extrema can be used to extract IMF. Otherwise, we should replace the extrema point by average value of the former extrema and the later one. However, such a threshold is difficult to determine unless there is a solid theoretical basis. Therefore, some alternative statistical methods were introduced, such as Box-Cox transformation. Taking the natural logarithm of the original sequence is the most common transformation, which was also used in this paper.

The final one is the mode mixing. Sampling refers to obtaining discrete samples at a frequency, which is lower than the original sequence frequency, to approximate the original sequence. A unique discrete signal sequence can be obtained by sampling from a continuous sequence at a specific frequency. However, it is not necessarily to restore the unique continuous original sequence through a discrete sample at a certain frequency. Then the mode mixing happens. Theoretically, sampling stock markets that generate nearly a thousand trades per millisecond at any common interval, such as per minute or per day, can lead to mode mixing. In order to solve the existing mode mixing, Wang., Chau, and Xu (2007) further improved the EMD and proposed the ensemble empirical mode decomposition (EEMD). By adding white noises to the original sequences, EEMD can realize the integrated averaging of the disturbance factors and eliminate the influence of mode mixing. The algorithm of EEMD is as follows:

Step1 : with a given  $K, k = \{1, 2, \dots, K\}$ . For each  $k$ , generate a white noise series  $\varepsilon_k$ , get the  $x_{0,k} = x_0 + \varepsilon_k$  as the ensemble of original sequence and white noise sequence;

Step2 : decompose each  $x_{0,k}$  through EMD, then obtain  $x_{0,k} = \sum_{i=1}^n IMF_{i,k} + r_k$ .

Step3 : obtain the decomposition result of original sequence  $x_0$  as Equation 2:

$$x_0 = \sum_{i=1}^n \overline{IMF}_{i,k} + \bar{r}_k \quad (2)$$

The fluctuations of the noise series  $\varepsilon_k$  should be correlated with those of the original series  $x_0$ . The relationship between variances of  $\varepsilon_k$  and  $x_0$  satisfy the following rule shown in Equation 3:

$$\text{var}(\varepsilon_k) = \frac{\text{var}(x_0)}{\sqrt{K}} \quad (3)$$

### 3. DATA

#### 3.1. The SWS Industry Index

Through the wind database, we obtained the daily closing prices of the SWS Industry Index for 28 different industries. The observation period we chose was in a range of 3,046 trading days, from July 2, 2007 to December 30, 2019. When SWS Industry Index decides which industry a listed company belongs to, it is mainly based on the main business income and main business profit composition of this company in the last two years. When the main business income and profit ratio is not consistent, the profit prevails. SWS Industry Index contains information on the prices of securities of listed companies in various sectors in China's capital markets. The specific industries are listed in Table 1.

Table-1. The industries and their index codes of SWS Industry Indexes.

Industry	Index code	Industry	Index code
Agriculture	801010.SI	Trading	801200.SI
Mining	801020.SI	Leisure	801210.SI
Chemical	801030.SI	Comprehensive	801230.SI
Iron & Steel	801040.SI	Building Materials	801710.SI
Non-ferrous	801050.SI	Building Decoration	801720.SI
Electronics	801080.SI	Electrical Equipment	801730.SI
Household Appliance	801110.SI	Defense Industry	801740.SI
Food	801120.SI	Computer	801750.SI
Textile and Apparel	801130.SI	Media	801760.SI
Light Industry	801140.SI	Communication	801770.SI
Biomedical	801150.SI	Banking	801780.SI
public utilities	801160.SI	Non-bank Financial	801790.SI
Transportation	801170.SI	Motor	801880.SI
Real Estate	801180.SI	Mechanical	801890.SI

Source: Wind database.

### 3.2. Data Pre-Processing

According to the statement in section 2.2, we first took the natural logarithm of 28 price sequences, and then used EEMD to decompose these logarithmic price sequences. For each logarithmic price, 100 white noise sequences were generated. According to (3), the variance of these white noises generated for a logarithmic price sequence should be 0.1 times of the variance of this sequence. Through EEMD, each of the 28 SWS Industry Index logarithmic price sequences were decomposed into 9 IMFs and a trend sequence. In this paper, only the Biomedical Index (801150.SI) was taken as an example to demonstrate the result of EEMD as follows:

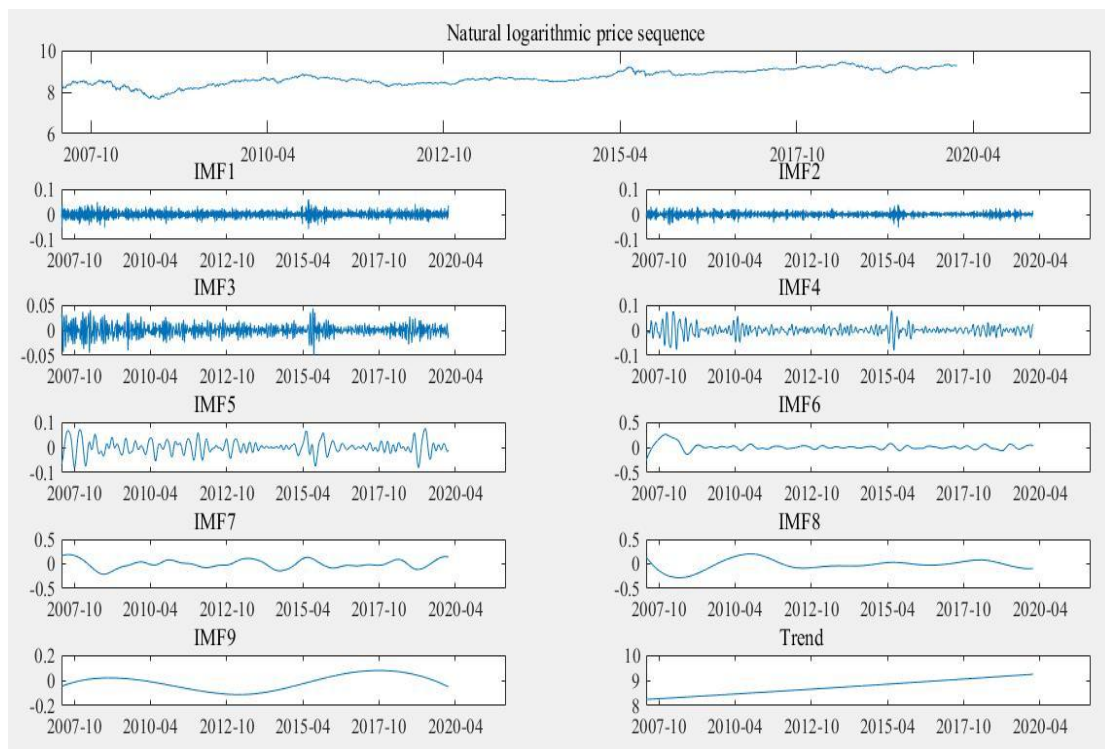


Figure-1. The EEMD result of logarithmic price sequence of 801150.SI (biomedical industry).

Source: Wind database.

As shown in Figure 1, IMF frequencies obtained from the decomposition index SWS Industry Index are arranged from high to low, and the trend term was a monotonically increasing smooth sequence without any periodic changes. At the same time, IMFs with different time scales of fluctuations had different energy levels.

Next, we performed a spectrum analysis of IMFs of all index prices. The main method of spectrum analysis was the periodogram method, which periodically decomposes a time-series  $x(t)$  according to Equation 4:

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos \frac{n\pi}{l} t + b_n \sin \frac{n\pi}{l} t) \tag{4}$$

where

$$a_n = \frac{1}{l} \int_{-l}^l x(t) \cos \frac{n\pi}{l} t dt$$

$$b_n = \frac{1}{l} \int_{-l}^l x(t) \sin \frac{n\pi}{l} t dt$$

in which  $a_0$  represents the trend of  $x(t)$ . This method was essentially the Fourier decomposition with given basis function. Since IMFs are locally symmetric sequences, it was feasible to use the periodogram method for spectral analysis.

If the random sequence  $x(t)$  is a band-limited signal sequence, its variance  $P(t)$ , or energy  $P(\omega)$  satisfies

$$P(t) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{|F_T(\omega)|^2}{T} d\omega = P(\omega) \tag{5}$$

in which  $t \in [-T/2, T/2]$ , and  $F_T(\omega) = F(x_T(t))$  represents the Fourier transform of  $x(t)$  according to (4).

Thus the power of  $x(t)$  satisfies

$$p(\omega) = \frac{dP(\omega)}{d\omega} = \lim_{T \rightarrow \infty} \frac{|F_T(\omega)|^2}{2\pi T} \tag{6}$$

The  $p(\omega)$  is called the power spectrum density function (PSDF) of  $x(t)$ . We still took Index 801150.SI (biomedical industry) for the example to observe the distribution of PSDFs of its IMFs according to Equation 6.

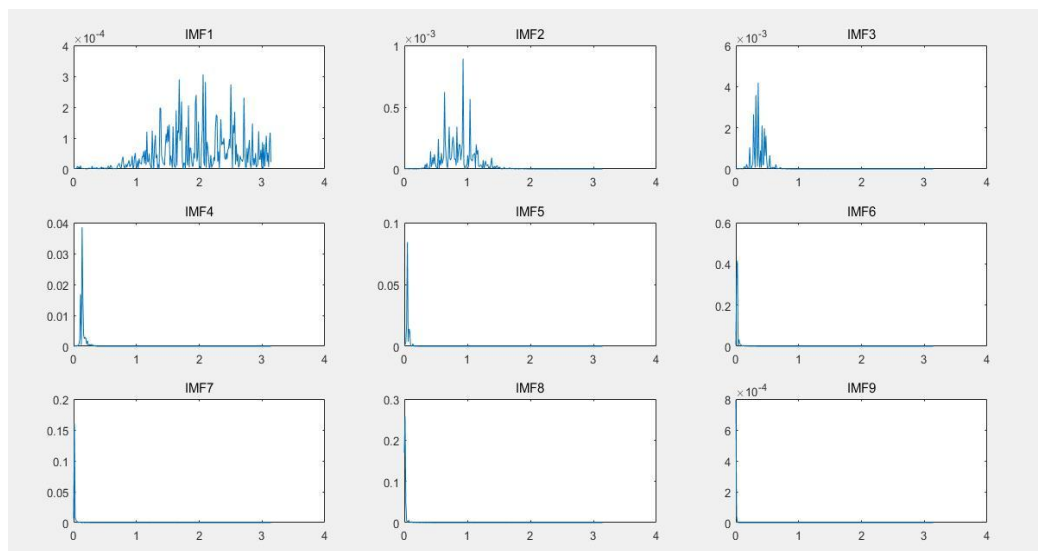


Figure-2. The PSDFs of IMFs of Index 801150.SI logarithmic price.

Source: Wind database.

According to Equation 5 the distribution of an IMF power shows the frequency range in which its fluctuations are concentrated. In Figure 2, the power spectrum distribution of the first three IMFs of logarithmic price of 801150.SI was relatively dispersed. From IMF4 to IMF9, their power distributions were highly concentrated around a narrow frequency bound. In fact, the more concentrated the energy distribution, the more cyclical the IMF is in the time domain. In contrary, the more dispersed the energy distribution, the closer it is to the noise.

If we take the power  $p(\omega)$  as the weight and average the frequency  $\omega$  in frequency domain, we obtain

$$\bar{\omega} = \frac{\int_{-\infty}^{\infty} p(\omega) \omega d\omega}{\int_{-\infty}^{\infty} p(\omega) d\omega} = \frac{\int_{-\infty}^{\infty} p(\omega) \omega d\omega}{\text{var}(x(t))} \quad (7)$$

as mean frequency of  $x(t)$ . The inverse of  $\bar{\omega}$  is the mean period of  $x(t)$ . According to Equation 6 and 7 we further calculated the power spectral density of IMFs of the logarithmic price sequences of all indexes and converted them into mean period on a daily basis, which can be summarized as follows:

Table-2. Mean period of IMFs of SWS Industry Indexes (trading day).

Industry	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9
Agriculture	4.2	10.2	22.1	58.5	162.4	218.0	282.1	319.5	768.2
Mining	4.2	10.4	22.5	56.1	106.4	411.7	350.0	660.5	3907.0
Chemical	4.4	9.9	22.4	57.9	144.4	418.5	359.5	585.7	9299.1
Iron & Steel	4.1	10.0	21.4	52.7	88.5	158.0	278.7	336.1	5134.5
Non-ferrous	4.1	9.9	23.8	61.5	123.3	434.1	365.2	522.4	1911.1
Electronics	4.2	9.9	21.9	55.5	120.5	238.3	361.9	524.9	267.1
Household Appliance	4.1	9.7	20.9	56.2	119.2	218.6	327.1	420.9	796.4
Food	4.3	10.3	20.3	57.9	129.2	400.3	431.4	376.6	482.1
Textile and Apparel	4.2	9.9	21.4	56.6	124.7	206.7	350.0	276.7	2837.9
Light Industry	4.3	10.0	21.4	58.3	112.0	181.4	340.1	440.0	3011.9
Biomedical	4.0	9.6	20.3	53.2	118.1	149.4	283.3	1386.0	2984.5
public utilities	4.2	9.8	21.1	56.9	114.8	148.6	451.0	318.5	1048.1
Transportation	4.3	10.0	21.4	56.6	114.4	137.3	256.3	354.9	3178.2
Real Estate	4.3	9.7	22.0	55.2	120.5	454.7	357.2	497.3	1607.7
Agriculture	4.3	9.9	23.0	56.3	134.9	153.3	236.1	354.9	588.8
Mining	4.3	10.2	22.5	52.7	121.3	231.3	433.8	260.0	1868.6
Chemical	4.2	10.0	21.1	54.8	128.6	176.6	502.5	393.2	2708.6
Iron & Steel	4.2	10.2	21.5	62.2	126.3	208.4	516.5	410.5	567.0
Non-ferrous	4.2	10.1	21.8	52.2	141.1	184.7	516.7	370.6	832.9
Electronics	4.1	10.0	21.1	55.7	143.5	244.7	1123.3	266.1	636.2
Household Appliance	4.0	9.9	22.0	58.8	115.6	171.0	384.2	384.6	516.1
Food	4.2	9.2	19.9	52.8	109.2	181.3	277.3	420.0	427.0
Textile and Apparel	4.3	9.7	21.7	54.0	96.4	209.8	614.0	2000.5	1126.8
Light Industry	4.4	9.7	21.5	57.3	155.3	154.1	438.5	361.2	1009.7
Biomedical	4.2	10.2	23.2	48.4	117.1	413.4	393.6	598.9	2276.0
public utilities	4.1	10.6	21.8	54.2	121.4	255.8	191.1	333.0	7543.8
Transportation	4.3	10.2	20.6	55.1	119.6	508.3	238.9	1028.6	1052.6
Real Estate	4.1	10.4	20.6	57.2	143.3	259.5	349.6	473.1	2299.7
Average	4.2	9.9	21.6	55.9	124.0	254.6	393.1	524.1	2167.4
Standard deviation	0.1	0.3	0.9	2.9	14.8	114.0	175.0	380.0	2180.0

Source: Wind database.



According to the results shown in Table 2, the average mean periods of fluctuation of IMF1s, IMF2s, IMF3s, IMF4s and IMF5s of the SWS Industry Index price sequences are approximately equivalent to one week, two weeks, one month, three months and half a year, respectively. Thus, on a one-year time scale, the volatility of SWS Industry Index prices is consistent with the trading rules of China's stock market exchanges. From IMF6s to IMF9s, the mean periods of IMFs of SWS Industry Index price sequences are longer than one year. From the perspective of inter-industry differences, IMFs with a mean period of less than one year have relatively small inter-industry differences, while IMFs with a mean period of more than one year have relatively large ones.

After analyzing the frequency domain characteristics of the industry index IMFs, we further summarized the distribution of their means and variances from the time domain.

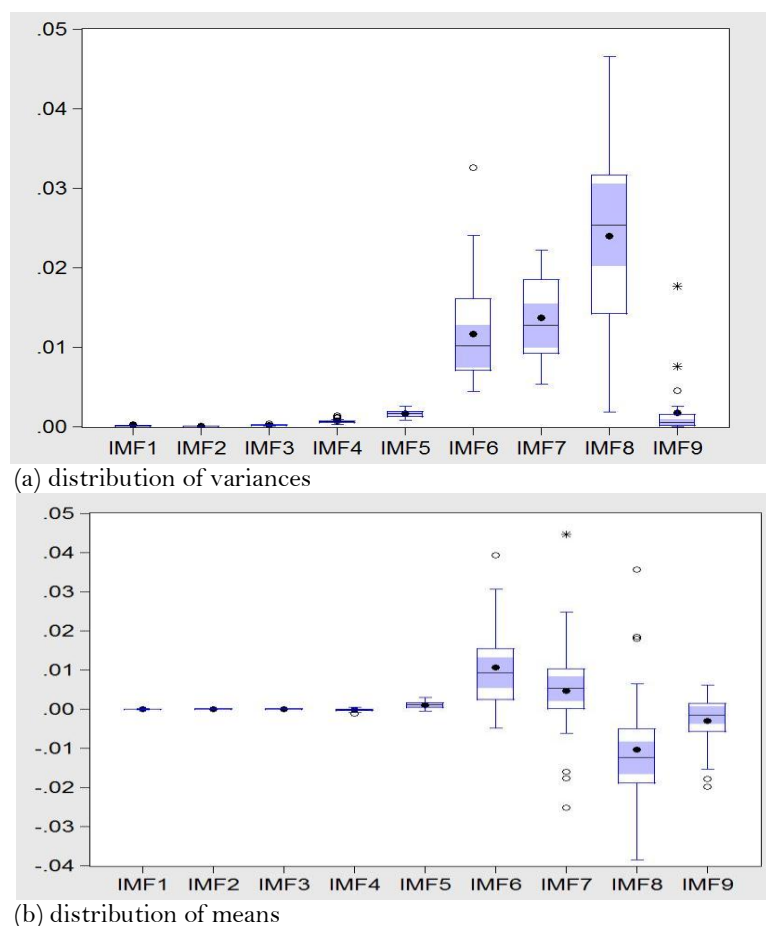


Figure-3. Distribution of means and variances of IMFs.

Source: Wind database.

According to Figure 3, the means of IMFs with mean periods less than one year are close to zero, and the differences between industries are relatively small. For IMFs with mean periods longer than one year, their means far deviated from zero, and the differences between industries are relatively large. Similarly, the variance of IMFs with mean periods more than one year is significantly higher than that of IMFs with mean periods less than one year. So do their inter-industry differences (the variances of IMFs' variances). In fact, a lot of the volatility in the price series is reflected in the long-term trend. We discuss our further analysis in the next section.

## 4. NUMERICAL RESULT

### 4.1. The Signal-Noise Separation

The basic idea of reconstructing a set of IMFs is to distinguish them into signals and noises according to their statistical characteristics. If the PSDF of a sequence is close to a constant, the energy of this sequence varies only

with the length of interval. Then the sequence can be regarded as noise. According to Equation 5 the variance of a sequence in time domain is equivalent to its energy in frequency domain. Therefore, the statistical definition of noise is a sequence with stationarity. In time domain, stationarity means that the distribution of a time-series does not change with time.

In addition to the general definition of signal and noise, we also need to consider the particularity of financial data. The statistical methods used to distinguish noise from signals actually reflect the limitations of human signal processing. When investors observe a sequence of prices, they tend to concentrate on sustained and significant price movements. Frequent oscillations are therefore relatively less noticeable (Black, 1986). Thus, on the one hand, we want to keep the signal rather than to keep noise, on the other hand, we want to maintain the IMFs with lower frequencies rather than to keep the ones with higher frequencies.

For each price sequence, our signal-noise separation scheme was to test the stationarity of each IMF, starting with its IMF1, until we found a non-stationary IMF, say, IMF6. Then, the IMFs from IMF1 to IMF5 were classified as high-frequency noise, and the ones from IMF6 to IMF9 as low-frequency signal.

**Table-3.** Result of signal-noise separation.

Industry	Low-Frequency Signal	Industry	Low-Frequency Signal
Agriculture	IMF6,7 ,8,9	Trading	IMF6,7 ,8,9
Mining	IMF6,7 ,8,9	Leisure	IMF6,7 ,8,9
Chemical	IMF6,7 ,8,9	Comprehensive	IMF6,7 ,8,9
Iron & Steel	IMF6,7 ,8,9	Building Materials	IMF6,7 ,8,9
Non-ferrous	IMF8,9	Building Decoration	IMF6,7 ,8,9
Electronics	IMF6,7 ,8,9	Electrical Equipment	IMF6,7 ,8,9
Household Appliance	IMF6,7 ,8,9	Defense Industry	IMF6,7 ,8,9
Food	IMF6,7 ,8,9	Computer	IMF7,8,9
Textile and Apparel	IMF6,7 ,8,9	Media	IMF6,7 ,8,9
Light Industry	IMF6,7 ,8,9	Communication	IMF6,7 ,8,9
Biomedical	IMF6,7 ,8,9	Banking	IMF7,8,9
public utilities	IMF7 ,8,9	Non-bank Financial	IMF6,7 ,8,9
Transportation	IMF6,7 ,8,9	Motor	IMF6,7 ,8,9
Real Estate	IMF6,7 ,8,9	Mechanical	IMF6,7 ,8,9

Source: Wind database.

As shown in Table 3, after the separation, IMF6, 7, 8 and 9 were classified as low frequency signals for the majority of SWS Industry Index price sequences. The price of few industry indexes, such as the utilities index, the banking index, and the computer index, contained less IMFs in their low-frequency signals. According to Table 2, since the mean period of IMF6 was one year on average, we believed that when observing the price sequence in a 10-year interval, IMFs with periods less than one year could be regarded as noise.

#### 4.2. The PCA Result

In this chapter, principal component analysis (PCA) was used for further discuss the fluctuation characteristics of the price sequences, the trend sequences that derived from EEMD and the low-frequency signal sequences obtained by signal-noise separation.

Price fluctuations are similar to the movement of people on a train. On a moving train, even though the people on board are moving in all directions, from the outside, everyone is moving in the same direction, showing consistency. The structure of price series fluctuation was obtained by PCA, which was represented by several pairs of one-to-one corresponding eigenvalues and eigenvectors. We focused on the largest few eigenvalues (in this paper, we focus on the largest two) and observed the sign of their corresponding eigenvectors. If the signs of elements in the eigenvector were exactly in same direction, it meant that all index prices change in the same

direction under this fluctuation mode. Otherwise, the index price fluctuations of different industries were non-synchronous.

**Table-4.** The eigenvalues of the first and second principal components of the logarithmic price sequences.

Eigenvalues	Proportion (%)	Cumulative Proportion (%)
20.765	75.973	75.973
4.376	16.011	91.884

Source: Wind database.

Table 4 shows that two principal components could explain more than 90% of the price fluctuations. Specifically, the first component and the second one explained 76% and 16% of the total volatility, respectively. Then we went further to observe the corresponding feature vectors of the two components:

**Table-5.** The eigenvectors of the first and second principal components of the logarithmic price sequences.

Industry	The first component	The second component	Industry	The first component	The second component
Agriculture	0.906	-0.341	Trading	0.645	0.239
Mining	0.272	0.872	Leisure	0.789	-0.160
Chemical	0.896	0.737	Comprehensive	0.810	-0.119
Iron & Steel	0.334	0.853	Building Materials	0.643	0.033
Non-ferrous	0.738	0.638	Building Decoration	0.752	0.145
Electronics	0.708	-0.571	Electrical Equipment	0.904	-0.119
Household Appliance	0.779	-0.438	Defense Industry	0.922	-0.124
Food	0.500	-0.568	Computer	0.725	-0.367
Textile and Apparel	0.708	-0.002	Media	0.998	-0.261
Light Industry	0.741	-0.040	Communication	0.946	0.129
Biomedical	0.560	-0.860	Banking	0.558	0.182
public utilities	0.713	0.279	Non-bank Financial	0.830	0.515
Transportation	0.868	0.897	Motor	0.792	-0.120
Real Estate	0.576	0.199	Mechanical	0.762	-0.100

Source: Wind database.

According to Table 5, in the two principal components that dominated the price movements of the index, there were components that represented both movements showing homogeneity between industries and those showing heterogeneity. The first principal component was the factor that caused all index prices to move in the same direction, while the second one was a volatility component with heterogeneity in relative fluctuations between industries. We needed to confirm the one that was the main cause of inter-industry differences.

**Table-6(a).** The eigenvalues of the first and second principal components of the low frequency signal sequences.

Eigenvalues	Proportion (%)	Cumulative Proportion (%)
23.764	82.332	82.332
2.019	6.995	89.327

**Table-6(b).** The eigenvalues of the first and second principal components of the trend sequences.

Eigenvalues	Proportion (%)	Cumulative Proportion (%)
20.754	81.460	81.460
4.726	18.540	100

Source: Wind database.

According to Table 6, both the volatility of the low-frequency signal and the one of the trend could be explained by their first principal components for more than 80%, and were the dominant components of the two systems. Further observation of the eigenvectors was as follows.

**Table-7(a).** The eigenvectors of the first and second principal components of the low-frequency signal sequences.

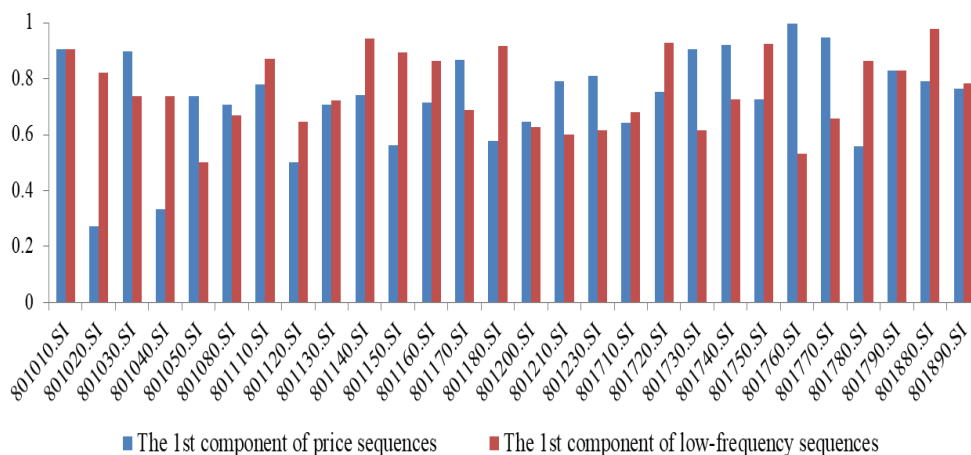
Industry	The first component	The second component	Industry	The first component	The second component
Agriculture	0.906	0.697	Trading	0.627	0.402
Mining	0.822	-0.529	Leisure	0.600	0.444
Chemical	0.738	0.308	Comprehensive	0.614	0.824
Iron & Steel	0.736	-0.537	Building Materials	0.682	0.723
Non-ferrous	0.501	-0.475	Building Decoration	0.926	0.666
Electronics	0.667	0.255	Electrical Equipment	0.616	0.434
Household Appliance	0.869	0.639	Defense Industry	0.728	0.714
Food	0.644	0.263	Computer	0.926	0.519
Textile and Apparel	0.723	0.567	Media	0.530	0.711
Light Industry	0.945	0.640	Communication	0.659	0.255
Biomedical	0.893	0.437	Banking	0.863	0.241
public utilities	0.863	0.285	Non-bank Financial	0.828	0.281
Transportation	0.689	0.020	Motor	0.976	0.598
Real Estate	0.918	0.414	Mechanical	0.782	0.509

**Table-7(b).** The eigenvectors of the first and second principal components of the trend sequences.

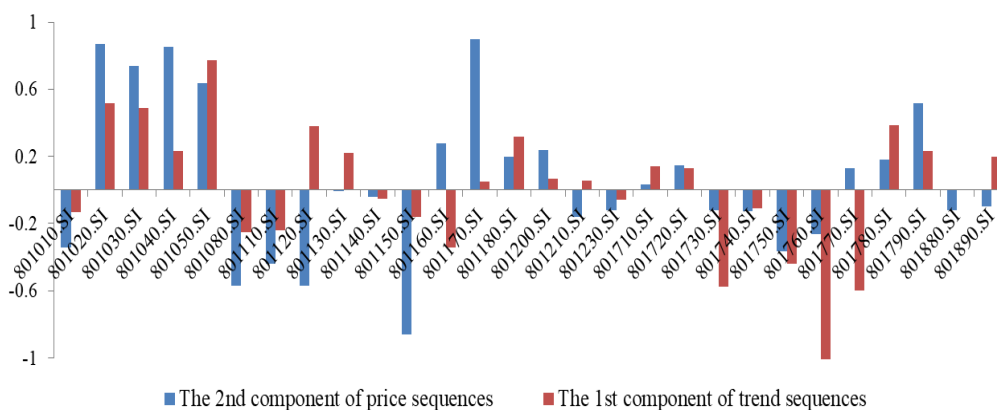
Industry	The first component	The second component	Industry	The first component	The second component
Agriculture	-0.132	-0.092	Trading	0.066	0.159
Mining	0.520	0.104	Leisure	0.058	-0.067
Chemical	0.490	0.336	Comprehensive	-0.055	0.021
Iron & Steel	0.234	0.921	Building Materials	0.139	-0.135
Non-ferrous	0.775	0.744	Building Decoration	0.129	0.370
Electronics	-0.251	-0.167	Electrical Equipment	-0.577	-0.410
Household Appliance	-0.239	-0.252	Defense Industry	-0.110	0.081
Food	0.377	-0.388	Computer	-0.438	-0.100
Textile and Apparel	0.220	0.119	Media	-1.143	0.111
Light Industry	-0.055	0.218	Communication	-0.599	0.432
Biomedical	-0.159	-0.391	Banking	0.384	0.259
public utilities	-0.341	0.646	Non-bank Financial	0.232	0.913
Transportation	0.049	0.972	Motor	0.005	-0.195
Real Estate	0.320	0.173	Mechanical	0.199	-0.267

Source: Wind database.

According to Table 7, the first principal component of low-frequency sequences caused the sequences to move in the same direction. The first principal component of a trend causes the heterogeneity. Compare the eigenvectors of the first principal component of low-frequency fluctuations and that of trends with the one of the first and second principal components of the original sequences, respectively.



(a) The first principal component of low-frequency sequences versus the first principal component of price sequences



(b) The first principal component of trend sequences versus the second principal component of price sequences.

Figure-4. Comparison of eigenvectors of different systems.

Source: Wind database.

Through the pairwise comparison of the eigenvectors shown in Figure 4, it was not difficult to find that the first principal component of the price sequence and the first principal component of the low-frequency signal had similar feature vector structures. The second principal component of the price series had a similar feature vector structure as the first principal component of the trend.

This result was instructive and economically significant. It showed that some long-term structural changes in China's economic are reflected in the long-term price trend of the Chinese stock market. Within the period we have chosen, the Chinese government has issued a policy called "cutting overcapacity" to change the status of supply exceeding demand. The policy reshaped the structure of many sectors like steel, cement and shipbuilding. At the same time, the long-term trends in the prices of related industry indexes in the capital market, such as the mining, steel, chemical and non-ferrous metals became different from those in other industries. In contrast, China's high-tech industry is growing rapidly due to strong demand and emerging technology. The government's support policies like "Made in China 2025" have further made computer, defense and communications become attractive industries for investors in capital market. In general, fundamental trend changes in development of industries, resulting in the heterogeneity of changes in securities prices in different industries. However, heterogeneity reflects in the trend of price fluctuations, rather than in a short cycle.

### 5. CONCLUSION

Through (ensemble) empirical mode decomposition, the closing price sequence of each one of the 28 SWS industry index in China's stock market was decomposed into a set of subsequences, which consisted of intrinsic

mode functions (IMFs) and trend. The obtained IMFs were further separated into two groups, the low-frequency signal and the high-frequency noise. Through the principal component analysis, we found that in the overall fluctuation of the 28 SWS industry index prices, the component that made all index prices fluctuate in the same direction accounts for more than 75%. The one that made the fluctuations in index prices vary depending on the industries accounts for 7%. The former component was mainly represented by the fluctuation in low-frequency signal sequence, the latter one was mainly represented by the fluctuation in trend sequences.

It means that in a time interval on a decade basis, the inter-industries difference of securities performance should be viewed as the difference of long-term trend rather than the periodic fluctuation (which is usually called as rotation). Therefore, investors who are interested in China's capital market should pay more attention to analyzing the long-term development trend of China's industrial economy, especially the major industrial policy issued by the government.

**Funding:** This study received no specific financial support.

**Competing Interests:** The author declares that there are no conflicts of interests regarding the publication of this paper.

**Acknowledgement:** The author wishes to thank the reviewers and the editor for providing feedback for improving the quality of this paper.

## REFERENCES

- Al-Hnaity, B., & Abbod, M. (2015). *A novel hybrid ensemble model to predict FTSE100 index by combining neural network and EEMD*. Paper presented at the Control Conference IEEE, Linz, AUSTRIA, 15-17JUL, 2015.
- Black, F. (1986). Noise. *The Journal of Finance*, 41(3), 528-543.
- Conover, C. M., Jensen, G. R., Johnson, R. R., & Mercer, J. M. (2008). Sector rotation and monetary conditions. *The Journal of Investing*, 17(1), 34-46. Available at: <https://doi.org/10.3905/joi.2008.701955>.
- De Bondt, W. F., & Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, 40(3), 793-805. Available at: <https://doi.org/10.1111/j.1540-6261.1985.tb05004.x>.
- Goel, G., & Hatzinakos, D. (2014). *Ensemble empirical mode decomposition for time series prediction in wireless sensor networks*. Paper presented at the International Conference on Computing, Networking and Communications, Honolulu, HI, 03-06FEB, 2014.
- Granger, C. W., & Morgenstern, O. (1963). Spectral analysis of New York stock market prices 1. *Kyklos*, 16(1), 1-27. Available at: <https://doi.org/10.1111/j.1467-6435.1963.tb00270.x>.
- Hong, H., Torous, W., & Valkanov, R. (2007). Do industries lead stock markets? *Journal of Financial Economics*, 83(2), 367-396. Available at: <https://doi.org/10.1016/j.jfineco.2005.09.010>.
- Huang, N. E., Wu, M.-L. C., Long, S. R., Shen, S. S., Qu, W., Gloersen, P., & Fan, K. L. (2003). A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 459(2037), 2317-2345. Available at: <https://doi.org/10.1098/rspa.2003.1123>.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., . . . Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971), 903-995. Available at: <https://doi.org/10.1098/rspa.1998.0193>.
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65-91. Available at: <https://doi.org/10.1111/j.1540-6261.1993.tb04702.x>.
- Kanas, A., & Kouretas, G. P. (2007). Regime dependence between the official and parallel foreign currency markets for US dollars in Greece. *Journal of Macroeconomics*, 29(2), 431-449. Available at: <https://doi.org/10.1016/j.jmacro.2005.11.003>.

- Lei, Y., He, Z., & Zi, Y. (2009). Application of the EEMD method to rotor fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, 23(4), 1327-1338. Available at: <https://doi.org/10.1016/j.ymssp.2008.11.005>.
- Rilling, G., Flandrin, P., Gonçalves, P., & Lilly, J. M. (2007). Bivariate empirical mode decomposition. *IEEE Signal Processing Letters*, 14(12), 936-939.
- Wang, J., Lingyu, T., Yuyan, L., & Peng, G. (2017). A weighted EMD-based prediction model based on TOPSIS and feed forward neural network for noised time series. *Knowledge-Based Systems*, 132, 167-178. Available at: <https://doi.org/10.1016/j.knosys.2017.06.022>.
- Wang, W.-C., Chau, K.-W., Xu, D.-M., & Chen, X.-Y. (2015). Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resources Management*, 29(8), 2655-2675. Available at: <https://doi.org/10.1007/s11269-015-0962-6>.
- Wang, W. C., Chau, K. W., & Xu, D. M. (2007). On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences*, 104(38), 14889-14894. Available at: <https://doi.org/10.1073/pnas.0701020104>.
- Zheng, J., Pan, H., Liu, T., & Liu, Q. (2018). Extreme-point weighted mode decomposition. *Signal Processing*, 142, 366-374. Available at: <https://doi.org/10.1016/j.sigpro.2017.08.002>.

*Views and opinions expressed in this article are the views and opinions of the author(s), Asian Economic and Financial Review shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.*