Exploring the underlying factors of single financial time series using EEMD-ICA based analysis method

Lu Xian
International Business School, Shaanxi Normal University, Xi'an, China
Tel: (+86) 029-8531-0457, Email: xianlu2010@snnu.edu.cn

Kaijian He
Business School, Hunan University of Science and Technology, Xiangtan, China
School of Economics and Management, Beijing University of Chemical Technology, Beijing, China
Tel: (+86) 010-6443-8793, Email: kaijian.he@my.cityu.edu.hk

George Mang
China Fortune Financial Group, Hong Kong
Email:georgemang@ffc.com.hk

Kin Keung Lai†
Department of Management Sciences, City University of Hong Kong, Hong Kong
International Business School, Shaanxi Normal University, Xi'an, China
Tel: (+852) 2788-8563, Email: mskklai@cityu.edu.hk

Abstract. Analyses of financial time series and exploring its underlying characteristic factors are longstanding research objectives. It received increasing attention from scientists and practitioners studying the financial market. Ensemble Empirical model decomposition (EEMD) and independent component analysis (ICA) are developed to deal with nonlinear and non-stationary time series, and recently, a new model integrated the two methods (called EEMD-ICA) has proposed for single-channel signal processing. For better exploration of the underlying factors of single financial time series, this article attempts to use an analysis approach based on EEMD-ICA model for this task. In the proposed approach, the single financial time series to be analyzed is decomposed into several statistically independent components. The decomposed components have a lot of potential information, and they respectively caused by the supply and demand, cycle, economical development and other factors. Especially, we can find the related economic variable for every decomposed component by careful analysis and comparison. The crude oil price is analyzed for illustration and verification. The empirical results show that EEMD-ICA based analysis approach is a vital technique for exploring the underlying factors of single financial time series.

Keywords: ensemble empirical model decomposition; independent component analysis; underlying factors; financial time series

1. INTRODUCTION

One of the most important goals of financial time series analysis is source extraction and dimensionality reduction, which can help extracting features and exploring the underlying factors in high-dimensional and complex data. Especially, understanding its underlying characteristic features has been the recent focus of scientists and practitioners studying the financial time series. There have been lots of studies on mining potential factors in financial time series. The approaches mostly can be grouped into two categories: fundamental analysis approaches and mathematical statistics methods. Fundamental analysis approaches are mainly based on the point of causal relationship theory. They always investigate the various factors affecting the financial market, research the behavior characteristics of the financial market's variables, explore its inherent law and predict its future change trend (e.g.

Bacon, 1991; Yang et al., 2002; Coleman, L., 2011). Mathematical statistics methods include many linear and nonlinear models, such as Autoregressive Moving Average (ARMA), Autoregressive Conditional Heteroscedasticity type models (ARCH) (e.g. Sadorsky, 2002; Morana, 2001), Artificial Neural Network (e.g. Mirmirani and Li, 2004; Moshiri, 2004; Nelson et al., 1994; Yu et al., 2006), Support Vector Regression (Kim, Kyoung Jae, 2003), etc.

The financial time series is always with high-frequency, nonlinear, a non-stationary and long memory property. Therefore it cannot be analyzed and forecasted by traditional method. In recent years, Ensemble Empirical model decomposition (EEMD) and independent component analysis (ICA) have been developed to deal with nonlinear and non-stationary time series. They have made many successes in financial time series analysis (Huang et al. 2003b; X. Zhang et al, 2008; Back and Weigend, 1997; Kiviluoto and Oja, 1998; Oja et al,2000). For further analysis, a new technique combining EMD and ICA (called EMD-ICA) and its modified EEMD-ICA have been introduced (Mijović, et al. 2010, L. Xian, et al.2016).

In this paper, we apply the EEMD-ICA model to single financial time series analysis and formulate an innovative EEMD-ICA based approach to identify the underlying factors. In the proposed approach, the original time series is firstly decomposed into statistically independent components (ICs), and then the change trend of each component is investigated. Usually, for each component, we can find a dominant economic factor that caused this oscillation. And these economic factors are very important for the formation of the original time series.

The organization of this paper is given as below: Section 2 introduces EEMD and ICA briefly. Simultaneously, the proposed EEMD-ICA based analysis approach is presented in details. For illustration and verification, the crude oil price is analyzed in section 3. Section 4 concludes.

2. METHODOLOGY

2.1 Ensemble empirical mode decomposition

EMD is a method proposed by Huang et al. (1998) for decomposing a nonlinear, non-stationary time series into several components referred to as Intrinsic Mode Functions (IMFs). We always extract the IMFs and a residual series from the original data by sifting process (Huang et al. 1998). The total number of IMFs is limited to $\log_2 N$, where N is the length of original data series $x(t)$. If we denote $c_i(t), i = 1, \cdots, N$ to be the resultant set of IMFs and the residual series is $r(t)$, the original time series can be expressed as:

$$x(t) = \sum_{i=1}^{N} c_i(t) + r(t) \quad (1)$$

The advantages of EMD are clear and have been demonstrated by many researchers (Huang et al. 2003; X. Zhang et al.2008). However, the original EMD has a drawback — the frequent appearance of mode mixing, defined as a single IMF consisting of either signals of widely disparate scales, or a signal of a similar scale residing in different IMF components. Wu and Huang (2008) proposed EEMD to solve this problem. The basic idea of EEMD is that observed data are amalgamations of the true time series and noise. Thus even if data are collected by separate observations, each with a different noise level, the ensemble mean is close to the true time series. Therefore, an additional step is taken by adding white noise that can help extract the true signal in the data.

The effect of the added white noise can be controlled by the well-established statistical rule, calculated as in (2).

$$\varepsilon_n = \varepsilon / \sqrt{N} \quad (2)$$

Where N is the number of ensemble members, ε is the amplitude of the added noise and ε_n is the final standard deviation of error, defined as the difference between the input signal and the corresponding IMFs.In practice, the number of ensemble members is often set to 100 and the standard deviation of white noise series is set to 0.1 or 0.2.

2.2 Independent component analysis

ICA is a statistical and computational technique used for identifying hidden factors that underlie sets of random variables, measurements or signals. We assume that the set of random samples $\boldsymbol{x}_i$ of size $n \times 1$, $i = 1, 2, ..., m$, $m \le n$ is generated by a linear mixture of unknown factors which we denote as $\boldsymbol{s}_i$ of size $n \times 1$. In matrix notation, we can obtain (Hyvärinen & Oja, 2000)

$$\boldsymbol{X} = \boldsymbol{AS} = \sum_{i=1}^{m} \boldsymbol{a}_i \boldsymbol{s}_i^T \quad (3)$$

Here $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_m]^T$, $\boldsymbol{A}$ is the assumed $m \times m$ static mixing matrix, $\boldsymbol{a}_i$ is the i th column of $\boldsymbol{A}$ and $\boldsymbol{S} = [\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_m]^T$ of size $m \times n$. The statistical model in the above equation is called independent components analysis, or ICA model. The goal of ICA model is to estimate both $\boldsymbol{A}$ and $\boldsymbol{S}$ using the observed data $\boldsymbol{X}$. Several existing algorithms (such as Infomax, FastICA, FJADE and CICA) can be used for ICA modeling (Bell & Sejnowski, 1995; Hyvärinen & Oja., 2000). In this paper, the FastICA algorithm proposed by Hyvärinen & Oja (2000) is adopted to estimate the de-mixing matrix $\boldsymbol{A}$ and $\boldsymbol{S}$.

2.3 The improved EEMD-ICA model

Mijović et al. (2010) proposed the EMD-ICA model for single-channel signal processing. Recently, this model was improved in two aspects: (1) EEMD replaces the original EMD model to decompose the financial time series for better performance; (2) For reducing the influence of unimportant IMFs, a procedure of recombination is added to in the new method (L. Xian, et al. 2016).

The improved EEMD-ICA methodology generally comprises of the following three steps (L. Xian, et al. 2016):

1. The financial time series $x(t), t=1,\cdots,T$ is decomposed into N IMFs, $c_i(t), i=1,\cdots,N$ and the residual series $r(t)$.

2. Evaluate the contribution coefficient of the k th IMF (CCI_k) and the residual series by the transformative r elative hamming distance (RHD):

$$CCI_k = 1 - \frac{1}{T-1}\sum_{t=1}^{T-1} R(t) \tag{4}$$

Where $R(t)=1$ if $(x(t+1)-x(t))(\hat{x}(t+1)-x(t)) \geq 0$, or else $R(t)=0$; $\hat{x}(t)=\sum_{i=1, i\neq k}^{N+1} c_i(t)$.

3. Compare all the contribution coefficients with a hard threshold λ, which always be set to a fixed small value (such as 0.2 or 0.3). All the IMFs with smaller contribution coefficients are merged into a new data series. The reserved and merged data series formed a new data set, called VIMFs, $v_j(t), j=1,\cdots,M$ and $M \leq N+1$.

4. Apply the ICA to VIMFs and get statistically independent components $s_k(t), k=1,\cdots,L$ and $L \leq M \leq N+1$.

Through linear transformation, we can reconstruct of f inancial time series in terms of the estimated ICs as

$$\hat{x}(t) = \sum_{i=1}^{N} c_i(t) + r(t) = \sum_{j=1}^{M} v_j(t) = \sum_{k=1}^{L} b_k s_k(t) \tag{5}$$

Where b_k is the sum of the kth column of mixing matrix $\boldsymbol{A}$, and it is called transformation coefficient of the kth IC.

2.4 EEMD-ICA based analysis approach

The new EEMD-ICA model gives us a new way for exploring the underlying factors of single financial time series. After we identify the single financial data of interest, the overall process of exploring its underlying factors is described as follows:

1) Determining the time period of research

The first step is to determine the period of the single financial time series to be analyzed. For the complex of the financial time series, there are many different driven factors in different time (Hamilton, 2009). Determining the interest period firstly is of great help to explore potential factors more accurately. The period is always determined by the goal of our research. Of course, we can use some special methods to determine the period with certain characters, such as structure breakpoint test (Bai and Perron, 2003), event study methods (Mackinlay, 1997) etc.

2) Fundamental analysis

Fundamental analysis method is a common way based on economic supply and demand theory and the experience analysis, but to overcome the pure data driven analysis method with utility. Therefore, we can preliminarily analyze the factors which drive the fluctuations by investigating the structural characteristics of the single financial time series.

3) Decomposing data by EEMD-ICA

After the determining the time period and the preliminary understanding of the data, the single financial time series of interest is decomposed into several statistically independent components.

4) Compare and economic analysis

As mentioned before, each IC has a concrete implication in economic. This step is comparing each IC with economic variables referred in the second step and identifying the potential correspondence. There are two main ways of comparing, one is trend contrast, and the other is observing the significant change in major point.

In the ICA model, since both S and A are unknown, we cannot determine the variances of the ICs (A. Hyviirinen et al, 2001), the matrix A is adapted to make the variance of each IC 1. Therefore, our compare is based on each IC's changing trend rather than its amplitude.

5) Verification

In this step, we investigate the correlation between economic factors and the extracted independent components to test and confirm the derived economic meaning of each independent component.

3. EMPIRICAL ANALYSIS: EXPLORING TH E UNDERLYING FACTORS OF CRUDE OIL PRICE

In this section, the crude oil price is analyzed to illustrate and verify the proposed EEMD-ICA based analysis approach. We use the monthly data of West Texas Intermediate (WTI) crude oil spot price for our analysis.

1) Determining the time period of research

The time period in our research is from January 1986 to December 2015. We select the time period is based on following purposes. On the one hand, the oil price changes more intensely in recent 30 years. On the other hand, the long time range of this data set helps extract more information. We can also choose other shorter time period, if we interested in features of a certain time period of crude oil price.

2) Fundamental analysis

Different financial time series is always considered with different background and different driven factors. Petroleum has many different attributes, such as resource attribute, commodity attribute, financial attribute, and political attribute, etc. These different attributes make the oil price fluctuations become more complicated. The relationships between the crude oil price and the driven factors that come from different attributes have been investigated in many literatures (Ferson and Harvey, 1995; Basher and Sadorsky, 2006; Dees et al., 2007; Zhang et al.,2008; Narayan et al., 2010; Zhang and Wei, 2010;Fan and Xu, 2011; Mayukha et al. 2014). The main factors determining the oil price include: supply-demand, world economic development, stock market, gold market, US dollar, speculation, geopolitical circumstances, and so on.

3) Decomposing data by EEMD-ICA

EEMD-ICA is applied to decompose WTI crude oil spot price, with a total of 360 data points. The number of ensemble members is set to 100, the standard deviation of white noise series is set to 0.2, and the hard threshold is set to 0.3 in our analysis. The crude oil price is decomposed into 5 ICs. Fig.1 shows the visualization.

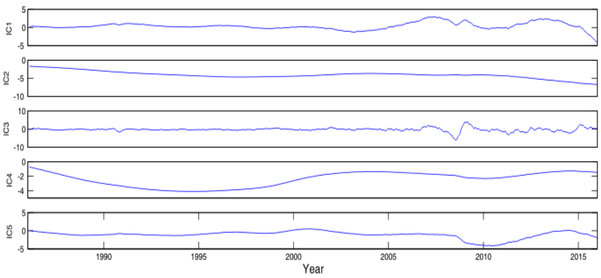

Figure 1: The ICs for the crude oil price from Jan. 1986 to Dec. 2015 by EEMD–ICA.

4) Compare and economic analysis

The mixing matrix A and its inverse are also obtained during the decomposing process. From Eq.(5), the reconstruction of crude oil price can be shown as

$$\hat{x}(t)=\sum_{k=1}^{L} b_k s_k(t)=12.72s_1-14.26s_2-13.53s_3+15.52s_4-11.95s_5 \quad (6)$$

The next work is to explore the economic meaning of each IC by careful comparison and analysis. Analyses of the statistical characteristics of ICs can help us to finish this work. Table 1 shows related statistical information about the ICs, include mean, skewness, kurtosis and Hurst exponent, correlation between each IC and the crude oil price, Jarque-Bera(J-B) test for normality, and so on. The transformation coefficients of the ICs are also listed.

From Table 1, we can get some important information. First, all ICs are distinctly non-Gaussian, as the J-B test are all significant at the 5% confidence level and the ICs' kurtosis are far from 3. Second, the correlation coefficients between ICs and crude oil price are similar. This result shows that ICs' corresponding economic variables are of almost equal importance for crude oil price fluctuations. Third, all the Hurst exponents are greater than 0.5, it means all ICs have long-term memory.

US dollar is the major invoicing currency in the international crude oil markets, and the US dollar exchange rate always play contrary effect on crude oil price fluctuation (Zhang et al.,2008; Lizardo and Mollick. 2010). Through the observation of IC1's fluctuation in some periods (such as 1993 to 2002, 2003 to 2007, 2014 to present), we can find that the change trend of IC1 is always contrary to the US dollar index. Hence, IC1 can be seen as the factor of US dollar.

Table 1: Descriptive statistics of the estimated ICs.

	b_i	Mean	Skewness	Kurtosis	Correlation coefficient	Hurst exponent	J-B test Statistic
IC1	12.72	0.40	-0.20	5.56	0.42(0.00)	0.79	100.6(0.00)
IC2	-14.26	-4.08	-0.01	3.88	-0.47(0.00)	0.95	11.7(0.01)
IC3	-13.53	-0.28	-0.91	11.88	-0.44(0.00)	0.65	1232.3(0.00)
IC4	15.52	-2.40	-0.47	1.74	0.51(0.00)	0.99	37.2(0.00)
IC5	-11.95	-1.10	-1.51	5.30	-0.39(0.00)	0.98	217.1(0.00)

Note: p-values in parenthesis; b_i is the transformation coefficient of the ith IC.

The second IC changes slowly, and remains at a certain high absolute value; it represents the change of oil supply and demand situation, as both IC2 and the value of oil production minus consumption have same descending trend. There are many factors which affect the oil supply and demand, include oil reserves, production, consumption, the development of alternative energy sources and so on.

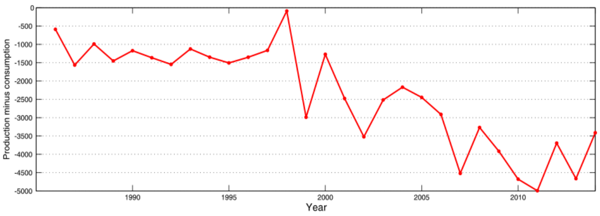

Figure 2: The crude oil production minus consumption from 1995 to 2013.

Petroleum is always a critical strategic policy tool of international politics and IC3 reflects these political and other extreme events effects perfectly. It has the highest change frequency among all the ICs. Events such as the Gulf War in 1991, the global financial crisis in 2008 represent the corresponding changes in IC3. The behavior of speculative funds exacerbated these impacts.

IC4 reveals the cycle fluctuations of crude oil prices. As important energy resource and commodities, crude oil always shows closely relationship with economic growth (Ghalayini, 2011). World and some important countries economic development cycles and seasonal consumption impetus are main cause of the oil's fluctuant cycle.

For the financial attribute, there usually exists a very strong relationship between the stock market, gold market and crude oil market. IC5 seems to be consistent with these influences, especially at some important time point, such as around 2000 and 2008 to 2009.

5) Verification

From the above analysis, some main underlying factors of the crude oil price have been found. Next we will test and confirm our analysis by regression analysis. Before regression, we should illustrate the follow three points. First, in our analysis, we use the robust regression instead of the simple ordinary least-squares (OLS) for its sensitivity. Second, these economic variables are not always quantifiable, such as economical cycle, geopolitics and peculation. Thus we use some proxy variables, and the specific contents are shown in Table 2. Third, since data for some economic variables such as GDP, oil Production and consumption, etc. are only available ate yearly frequency, data for their corresponding ICs are converted to the annual frequency.

The regression results are reported in Table2, and provide the statistical evidence about the linear relationship between ICs and related economic variables.

Table 2: Robust regression of the estimated ICs on the relevant economic variables.

	Intercept	Coefficients	R^2	Adjusted R^2	F-stat
IC1 on US dollar	3.77^{***}	0.11^{***}/-0.04^{***}	0.29	0.29	72.5^{***}
IC2 on Supply & Demand	2.21	3.8e-04**/3.25e-04*/2.38e-03*	0.53	0.48	9.43^{***}
IC3 on Geopolitics & Speculation	-0.98^{**}	4.60e-05**/-3.25e-08***	0.34	0.29	6.77^{***}
IC4 on Cycle	-2.72^{***}	0.52^{**}/-0.46^{**}	0.2	0.14	3.24^{*}
IC5 on Finance	-0.76^{***}	3.1e-04***/1.16e-03***	0.45	0.45	145^{***}

Note: US dollar includes the consumer price index (CPI) in US and US dollar index; Supply & Demand includes crude oil production, consumption and proved reserves in world; Geopolitics & Speculation includes the world's annual oil trade and GDP; Cycle includes the world's GDP growth rate and US GDP growth rate (annual %); Finance includes the nasdaq composite index and the gold price. * $p<0.1$, ** $p<0.05$, $^{***}p<0.01$.

4. CONCLUSIONS

In this paper, we introduce an EEMD-ICA based analysis approach to explore the underlying factors of single financial time series, and the crude oil price is analyzed for illustration and verification. The significance of our research work can be summarized as follows.

(1) A complete single financial time series analysis approach has been established. And this method can be easily extended to other areas, such as signal processing, forecast, and feature extraction.

(2) The decomposition method separates the complex financial time series into several simple independent parts. This way helps us to analyze the underlying information on various scales accurately.

(3) Our approach change previous way of analysis of time series. Previous way is, setting a model and fixed variables, and then estimating the effects of each variable. Our way is, getting the effects of each IC by automatically decomposition, and then exploring each IC's according economic variable.

(4) Crude oil is perhaps the most important commodity in the world. We analyze its price for illustration and verification. The empirical results show that our EEMD-

ICA based analysis approach is vital and effective.

REFERENCES

A.D. Back, A.S. Weigend, 1997. A first application of independent component analysis to extracting structure from stock returns, *Int. J. Neural Syst.* 8 (04) 473–484.

A. Hyviirinen, J. Karhunen, E. Oja, Independent Component Analysis, *Wileyand Sons,* 2001.

Bacon, R.(1991). Modelling the price of oil. *Oxford Review of Economic Policy* 7 (2), 17–34.

Bai, J., & Perron, P.. 2003. Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics,* 18(1), 1–22.

Basher, S.A., Sadorsky, P., 2006. Oil price risk and emerging stock markets. *Global Finance Journal* 17, 224–251.

B. Mijović, M. De Vos, I. Gligorijević, J. Taelman, S. Van Huffel, 2010. Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis, *IEEE Trans. Biomed. Eng.* 57 (9) 2188–2196.

C.-J. Lu, T.-S. Lee, C.-C. Chiu, 2009. Financial time series forecasting using independent component analysis and support vector regression, *Decis. Support Syst.* 47 (2) 115–125.

Coleman, L., 2011. Explaining crude oil prices using fundamental measures. *Energy Policy* 40(1), 318-324.

Dees, S., Karadeloglou, P., Kaufmann, R.K., Sanchez, M., 2007. Modeling the world oil market: assessment of a quarterly econometric model. *Energy Policy* 35 (1), 178–191.

E. Oja, K. Kiviluoto, S. Malaroiu,(2000) Independent component analysis for financial time series, in: Adaptive Systems for Signal Processing, *Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pp. 111–116.

Ferson, W.W., Harvey, C.R., 1995. Predictability and time-varying risk in world equity markets. *Research in Finance* 13, 25–88.

Ghalayini, L. (2011). The interaction between oil price and economic growth. *Review of Middle East Economics & Finance* (13), 127-141.

Hamilton J D. Understanding Crude Oil Prices. *Energy Journal -Cambridge Ma then Cleveland Oh-*, 2008, 30(2):179-206.

Kim, Kyoung Jae, 2003. Financial time series forecasting using support vector machines. *Neurocomputing* 55(s 1–2), 307-319.

K. Kiviluoto, E. Oja, Independent component analysis for parallel financial time series, in: *The Fifth International Conference on Neural Information Processing, ICONIP'R98, Kitakyushu, Japan, October* 21–23, 1998, Proceedings, 1998, pp. 895–898.

L. Xian, K. He, K. K. Lai, 2016. Gold price analysis based on ensemble empirical model decomposition and independent component analysis. *Physica A: Statistical Mechanics and its Applications.* 454 11-23.

MacKinlay, A. C.. 1997. Event Studies in Economics and Finance. *Journal of Economic Literature,* 35(1), 13–39.

Narayan, P.K., Narayan, S., Zheng, X.W., 2010. Gold and oil futures markets: Are markets efficient? *Applied Energy* 87 (10), 3299–3303.

N.E. Huang, M.-L. Wu, W. Qu, S.R. Long, S.S.P. Shen, 2003. Applications of Hilbert–Huang transform to non-stationary financial time series analysis, *Appl. Stoch. Models Bus. Ind.* 19 (3) 245–268.

N.E. Huang, Shen, Z., Long, S.R., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Process of the Royal Society of London. A* 454, 903–995.

P. Mayukha and P. Madhusudana Rao and P. Manimaran,2014. Multifractal detrended cross-correlation analysis on gold, crude oil and foreign exchange rate time series. *Physica A: Statistical Mechanics and its Applications* 416, 452-460.

Radhamés A. Lizardo and André V. Mollick.(2010). Oil price fluctuations and U.S. dollar exchange rates. *Energy Economics* 32(2), 399 – 408.

X. Zhang, K. Lai, S.-Y. Wang, 2008. A new approach for crude oil price analysis based on empirical mode decomposition, *Energy Econ.* 30 (3) 905–918.

Yang, C.W., Hwang, M.J., Huang, B.N., 2002. An analysis of factors affecting price volatility of the US oil market. *Energy Economics* 24, 107–119.

Ying Fan, Jin-Hua Xu, 2011. What has driven oil prices since 2000? A structural change perspective, *Energy Economics,* 33(6) 1082-1094

Y.J. Zhang, Y.M. Wei.2010. The crude oil market and the gold market: Evidence for cointegration, causality and price discovery. *Resources Policy* 35(3),168-177.

Y.J. Zhang, Y. Fan, H.T. Tsai, Y.M. Wei, 2008. Spill over effect of US dollar exchange rate on oil prices. *Journal of Policy Modeling* 30, 973–991.