Physica A 454 (2016) 11–23

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Gold price analysis based on ensemble empirical model decomposition and independent component analysis

Lu Xian [a], Kaijian He [b,c,*], Kin Keung Lai [a,d]

[a] International Business School, Shaanxi Normal University, Xi'an, 710119, China
[b] School of Economics and Management, Beijing University of Chemical Technology, Beijing, 100029, China
[c] Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong
[d] Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong

HIGHLIGHTS

- We propose a new improved model based on EEMD and ICA.
- We decompose gold price into statistically independent components (ICs).
- The regression analysis is used to analyze the economic meanings of different ICs.
- The proposed model is effective in the identification of main influencing factors for gold price movement.

ARTICLE INFO

Article history:
Received 28 September 2015
Received in revised form 21 January 2016
Available online 23 February 2016

Keywords:
Ensemble empirical model decomposition
Independent component analysis
Gold price

ABSTRACT

In recent years, the increasing level of volatility of the gold price has received the increasing level of attention from the academia and industry alike. Due to the complexity and significant fluctuations observed in the gold market, however, most of current approaches have failed to produce robust and consistent modeling and forecasting results. Ensemble Empirical Model Decomposition (EEMD) and Independent Component Analysis (ICA) are novel data analysis methods that can deal with nonlinear and non-stationary time series. This study introduces a new methodology which combines the two methods and applies it to gold price analysis. This includes three steps: firstly, the original gold price series is decomposed into several Intrinsic Mode Functions (IMFs) by EEMD. Secondly, IMFs are further processed with unimportant ones re-grouped. Then a new set of data called Virtual Intrinsic Mode Functions (VIMFs) is reconstructed. Finally, ICA is used to decompose VIMFs into statistically Independent Components (ICs). The decomposition results reveal that the gold price series can be represented by the linear combination of ICs. Furthermore, the economic meanings of ICs are analyzed and discussed in detail, according to the change trend and ICs' transformation coefficients. The analyses not only explain the inner driving factors and their impacts but also conduct in-depth analysis on how these factors affect gold price. At the same time, regression analysis has been conducted to verify our analysis. Results from the empirical studies in the gold markets show that the EEMD–ICA serve as an effective technique for gold price analysis from a new perspective.

* Corresponding author at: School of Economics and Management, Beijing University of Chemical Technology, Beijing, 100029, China. Tel.: +86 010 64438793; fax: +86 010 64438793.
E-mail addresses: xianlu2010@snnu.edu.cn (L. Xian), kaijian.he@my.cityu.edu.hk (K. He), mskklai@cityu.edu.hk (K.K. Lai).

http://dx.doi.org/10.1016/j.physa.2016.02.055
0378-4371/

1. Introduction

As one of the most important industry inputs and the medium of exchange, gold has been playing an important role in global commodity and financial markets. In recent years, the gold market has become very active because of its profit-making prospects and remarkable risk-avoidance features. As a result, gold price has received increasing attention and research interest. In contrast to many other financial assets, the price of gold follows unusual paths. There are numerous researches on the statistical characteristics of gold price and its movement in the literature. For instance, both Tschoegl [1] and Michael E. Solt [2] found that fluctuations in gold price are time-dependent. Cheung and Lai [3] believed that after 1979, generation of long-term memory in the gold market was attributed to several factors. As these factors gradually disappear, long-term memory will vanish. Mills [4] statistically investigated the behavior of gold price on the London market from 1971 to 2002 and found significant sharp peaks and heavy tails which was non-normal. Tully and Lucey [5] investigated macroeconomic influences on gold using the asymmetric power GARCH model and the results suggested that this new model provided a very good description of gold price. Parisi et al. [6] analyzed recursive and rolling neural network to forecast one-step-ahead sign variations in gold price and tried to dispose the sharp peaks and heavy tails. Sjaastad [7] examined the theoretical and empirical relationships between the major exchange rates and the price of gold using forecast error data. Batten and Lucey [8] examined gold futures contract on the Chicago Board of Trade (CBOT) using intraday data from 1999 to 2005 by means of GARCH model and nonparametric estimation method. They explored the characteristics of price volatility and concluded that effect of trading volume on price of gold futures was almost negligible. Shafiee and Topal [9] applied a modified econometric version of the long-term trend reverting jump and dip diffusion model to forecast gold price. The proposed models or methods can help understand characteristics of gold price and recognize each possible factor's influence on gold price. However, some very important topics have not been discussed in the existing literature. For example, how do these factors affect gold price exactly? How can we identify and describe each factor's influence? Can we quantify the importance of these factors? These questions are of crucial practical significance in further analysis of the features of gold price.

Independent Component Analysis (ICA) can be used to identify the underlying factors. ICA was for the first time introduced in early 1980s in the context of neural network modeling [10,11]. ICA can identify some independent and hidden sources from the mixtures without any prior knowledge of the mixing mechanism. The hidden information is called the Independent Components (ICs) which provide insights into the structure of the observable data set. Due to its generality, the ICA model has been applied in many different areas, such as signal processing, face recognition, feature extraction and quality control [12–15].

The ICA model has been used in financial data as well. There are many situations in which financial time series are closely related. Typical examples may include currency exchange rates, daily returns of stocks, crude oil price and so on. These data may have some underlying factors in common. Back and Weigend [16] applied ICA to extract the features of daily returns of stocks. Their results showed that the dominant ICs can reveal more details of the underlying structure and information of stock prices than Principal Components Analysis (PCA). Kiviluoto and Oja [17] tried to find the common fundamental factors among 40 stores under the same retail chain by ICA. They found that the cash flow of the retail stores was mainly affected by holidays, seasons and competitors' strategies. Oja et al. [18] applied ICA in foreign exchange rate time series prediction. Lu et al. [19] constructed a two-stage modeling approach using ICA and support vector regression in financial time series forecasting and found that this method could alleviate the influence of noise effectively. Nevertheless, ICA has rarely been applied in gold price analysis. These models are mostly proposed in the multivariate financial data analysis.

In the case of univariate data analysis, recently we have witnessed some initial efforts, most from the perspective of combinations of both wavelet analysis and ICA. For example, the method combining wavelet and ICA (called WICA) has been proposed by Lin and Zhang [20], which was used for fault feature separation. However, the wavelet transform imposes strict assumptions of using particular wavelet basis and encounter serious limitations in modeling the practical data. These limitations include the interference terms, border distortion and energy leakage, etc. [21]. During the wavelet transformation process, this would produce a lot of small undesired spikes all over the frequency scales and make the results confusing and difficult to be interpreted. What is more, the choice of mother wavelet and decomposition level has a significant impact on decomposition results [22]. Meanwhile, EMD is an empirical, intuitive, direct and self-adaptive data processing method. This means that it decomposes a signal without prior knowledge about the signal of interest embedded in the data series [23]. Moreover, the decomposition results of the EMD can reflect the physical properties of the original system more accurately, not restricted by the particular basis employed, and offer much better temporal and frequency resolutions [24]. In the literature EMD model is commonly perceived to be more effective in analyzing nonlinear and non-stationary data. Based on this, Mijović et al. [25] introduced the technique combining EMD and ICA (called EMD–ICA), and compared with WICA, the result showed that EMD–ICA outperforms WICA, especially for high noise-to-signal ratios. In this paper we will propose the modified EEMD–ICA models beyond that of Mijović et al. [25] for analyzing the gold price analysis.

The rest of the paper is organized as follows. Section 2 gives a brief introduction of ICA and EEMD. The improved EEMD–ICA model is proposed and described in Section 3. Section 4 describes the descriptive statistics of the gold price from 1969 to 2013. Detailed analyses based on the decomposed data using the EEMD–ICA model are presented. Section 5 concludes.