Physica A 433 (2015) 17–25

Contents lists available at ScienceDirect

Physica A

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

Modified cross sample entropy and surrogate data analysis method for financial time series

Yi Yin*, Pengjian Shang

Department of Mathematics, Beijing Jiaotong University, No. 3 of Shangyuan Residence Haidian District, Beijing, 100044, PR China

HIGHLIGHTS

- We propose a novel modified cross sample entropy (MCSE).
- Surrogate data analysis is employed to generate surrogate series.
- We compute MCSE differences between original dynamics and surrogate series (MCSDiff).
- MCSDiff is applied to simulated signals and US and Chinese stock markets.
- There are synchrony containing in the original financial time series by MCSDiff.

ARTICLE INFO

Article history:
Received 31 December 2014
Received in revised form 1 February 2015
Available online 28 March 2015

Keywords:
Modified cross-sample entropy (MCSE)
Surrogate data analysis
Multifractal detrended cross-correlation analysis (MF-DCCA)
Financial time series

ABSTRACT

For researching multiscale behaviors from the angle of entropy, we propose a modified cross sample entropy (MCSE) and combine surrogate data analysis with it in order to compute entropy differences between original dynamics and surrogate series (MCSDiff). MCSDiff is applied to simulated signals to show accuracy and then employed to US and Chinese stock markets. We illustrate the presence of multiscale behavior in the MCSDiff results and reveal that there are synchrony containing in the original financial time series and they have some intrinsic relations, which are destroyed by surrogate data analysis. Furthermore, the multifractal behaviors of cross-correlations between these financial time series are investigated by multifractal detrended cross-correlation analysis (MF-DCCA) method, since multifractal analysis is a multiscale analysis. We explore the multifractal properties of cross-correlation between these US and Chinese markets and show the distinctiveness of NQCI and HSI among the markets in their own region. It can be concluded that the weaker cross-correlation between US markets gives the evidence for the better inner mechanism in the US stock markets than that of Chinese stock markets. To study the multiscale features and properties of financial time series can provide valuable information for understanding the inner mechanism of financial markets.

1. Introduction

Economics has become an active research area for physicists, and a number of studies have used statistical mechanics to investigate the nature of an economy. Econophysics [1,2] is the term used to denote the application of statistical mechanics to economic systems. Information is an important keyword in analyzing the market or in estimating the stock price of a given company. A key measure of information is known as entropy, which is usually expressed by the average number of

* Correspondence to: Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing 100044, PR China. Tel.: +86 10 51684410; fax: +86 15210586330.

E-mail addresses: 09271087@bjtu.edu.cn (Y. Yin), pjshang@bjtu.edu.cn (P. Shang).

http://dx.doi.org/10.1016/j.physa.2015.03.055
0378-4371/

bits needed to store or communicate one symbol in a message. Entropy is a measure of degree of uncertainty to detect the system complexity from financial time series and has a wide application. Richman et al. developed sample entropy (SampEn) and it was widely applied in clinical cardiovascular studies [3,4]. In order to measure the similarity of two distinct time series, cross sample entropy (Cross-SampEn) was introduced. Comparison with correlation coefficient shows that Cross-SampEn is superior to describe the correlation between time series [5].

Nonadditive statistics [6] is a novel theoretical formulation, which is more suitable for complex systems, where the classical Boltzmann–Gibbs statistics is not quite adequate. This formulation generalizes Boltzmann–Gibbs entropy, introducing a nonadditive parameter (q). In the nonadditive formalism, the entropy of composed systems is not the simple sum of each subsystem entropy, which is in agreement with complex systems, once the composed systems interact among themselves. The essential feature of Cross-SampEn is to measure the degree of the asynchrony of two time series. The higher the value of Cross-SampEn, the more asynchronous is the pair of series [3,7,8]. As is known to us, the time series in the real world generally show multiscale behaviors. Inspired by the generalized sample entropy [9], we desire to investigate the multiscale behavior between time series hidden in the Cross-SampEn. Thus, we propose a modified cross sample entropy (MCSE), as parametric statistics, consistent with the nonadditive formalism. MCSE has the capacity of detecting the multiscale features of asynchrony for the financial time series. Surrogate data generation is a useful technique for nonlinearity hypothesis testing for time series analysis [10]. In surrogate data tests, a simple null hypothesis is that the series is described by an independent and identically distributed (IID) random variable. Here, the surrogate data are generated by simply shuffling the original time series, yielding in a time series with exactly the same time distribution but with no time correlations. In this paper, surrogate data are used to compute entropy differences between original dynamics and surrogate series.

Besides, multifractal analysis is a multiscale analysis and many sequences do not exhibit a simple monofractal behavior, instead, multifractal behavior is often observed. Simply speaking, a time series is multifractal if its fluctuations of different magnitude have different scaling exponents, and such a wide range of scaling exponents is preserver for every reasonably long part of the time series. Kantelhardt et al. [11] interpreted this multifractal phenomenon by two reasons: one is the broad probability density function for the values of the series, another one is the different long-range correlations of the small and large fluctuations. They also proposed multifractal detrended fluctuation analysis (MF-DFA) method for analyzing the multifractal characteristics. Zhou [12] extended this idea for two time series for the purpose of obtaining the multifractal features in the power law cross-correlations between two time series or higher-dimensional quantities recorded simultaneously, and named it multifractal detrended cross-correlation analysis (MF-DCCA). Since then, a lot of studies about the analysis of multifractal properties between the time series have been made [13–15]. Here we also apply MF-DCCA to study the multifractal behavior for cross-correlations between financial time series.

The remainder of this paper is organized as follows. In the following section, we propose the MCSE method and describe the MF-DCCA method briefly. Section 3 shows data description and Section 4 provides the results and analysis of applying MCSE and MF-DCCA methods to US and Chinese stock markets. Finally, we conclude the paper in Section 5.

2. Methodology

2.1. Modified cross-sample entropy (MCSE)

Tsallis entropy is a generalization of Boltzmann–Gibbs–Shannon (BGS) entropy and was proposed by Tsallis [6]. It is suitable for complex and multifractal systems, which exceeds domain of applicability for classical BGS entropy [16]. The discrete form of Tsallis entropy (S_q) is given by

$$S_q = \frac{1 - \sum_{i=1}^{W} p_i^q}{q - 1} \tag{1}$$

where p_i is the probability that the system is in state i, W is the number of possible states of the system, and q is the entropic index. In the limit $q \to 1$, S_q recovers the classical BGS entropy. Eq. (1) can also be written as

$$S_q = \sum_i p_i \log_q(1/p_i) \tag{2}$$

and derive the general form for logarithm function, namely, q-logarithm, defined as

$$\log_q(x) = \frac{x^{1-q} - 1}{1 - q} \tag{3}$$

($x \in R_+^*$; $q \in R$; $\log_1(x) = \log(x)$).

The essential feature of Cross-SampEn is to measure the degree of the asynchrony of two time series. The higher the value of Cross-SampEn, the more asynchronous is the pair of series [3,7,8]. Because of the multiscale nature contained in real time series, we are interested in exploring the multiscale behavior between time series hidden in the Cross-SampEn.