Physica A 400 (2014) 151–158

Contents lists available at ScienceDirect

Physica A

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

Classifying of financial time series based on multiscale entropy and multiscale time irreversibility

Jianan Xia *, Pengjian Shang, Jing Wang, Wenbin Shi

Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing 100044, PR China

HIGHLIGHTS

- We apply the methods and succeed to classify the financial markets.
- We confirm that the asymmetry is an inherent property of financial series.
- Loss of time irreversibility has been detected in high noise added series.

ARTICLE INFO

Article history:
Received 7 August 2013
Received in revised form 8 October 2013
Available online 11 January 2014

Keywords:
Multiscale time irreversibility
Multiscale entropy
Financial time series

ABSTRACT

Time irreversibility is a fundamental property of many time series. We apply the multiscale entropy (MSE) and multiscale time irreversibility (MSTI) to analyze the financial time series, and succeed to classify the financial markets. Interestingly, both methods have nearly the same classification results, which mean that they are capable of distinguishing different series in a reliable manner. By comparing the results of shuffled data with the original results, we confirm that the asymmetry property is an inherent property of financial time series and it can extend over a wide range of scales. In addition, the effect of noise on Americas markets and Europe markets are relatively more significant than the effect on Asia markets, and loss of time irreversibility has been detected in high noise added series.

1. Introduction

Time irreversibility is a characteristic feature of many observed time series and has recently gained increasing attention. A time series $\{x(j) : 1 \leq j \leq N\}$ is said to be time irreversible if its statistical properties change after its time reversed (asymmetry with respect to time reversal). It provides an effective way to detect the complexity of non-equilibrium systems and reflects the arrow of time. Indeed, time irreversibility analysis is capable of detecting a specific class of nonlinear dynamics, that is, those characterized by a temporal asymmetry [1].

In previous studies, a number of works focused on the analysis of physiologic signals such as heart rate variability with different pathology [2,3], and also in healthy subjects during physical activity [3,4]. Traditionally, time irreversibility is assessed by the asymmetric distribution of all the points with regard to the main diagonal line in the Poincaré plot. Several indices were proposed to measure the asymmetry, such as *Porta's index*, and *Guzik's index*, [1,5,6]. In order to increase the information gained from irreversibility analysis, the multiscale irreversibility measures have been developed [4,7–11]. However, relatively little work has been focused on the asymmetry analysis of other complex signal (i.e. financial time series).

The aim of this study is twofold. The first aim is to detect the temporal asymmetries of financial time series from multiscale irreversibility analysis, and examine the capability of data distinguishing. We apply the MSE [12–14] to quantify

* Corresponding author. Tel.: +86 15120074546.
E-mail address: 08271110@bjtu.edu.cn (J. Xia).

0378-4371/$ – see front matter
http://dx.doi.org/10.1016/j.physa.2014.01.016

the complexity of financial system and distinguish different area markets simultaneously. The second purpose is to confirm that the asymmetry property is an intrinsic characteristic of financial series. We suspect that if it is possible to find similar behavior in the challenge of testing the temporal asymmetries of shuffled data and investigating the effect of noise on this method.

The rest part of this paper is organized as follows. In the following section, we describe the methodology in detail. In Section 3, we show the results of the MSTI analysis on financial time series and discuss the performance, including the comparison of original results and the results of shuffled data, and the effect of noise. The conclusion is shown at the last section.

2. Methods

2.1. Theory of MSE

The MSE method is based on the application of sample entropy (SampEn), which is proposed by Costa et al. [12–14]. And our group has applied this method to analyze the daily records of 10 stock indices [15] and traffic time series [16].

Here we briefly review the procedure of MSE analysis. In the MSE method, the first is to construct consecutive coarse-grained time series from the original series $\{x(j) : 1 \leq j \leq N\}$ with the scale factor τ, $\{y^{(\tau)}\}$. Each point $y_j^{(\tau)}$ is defined as

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leq j \leq \frac{N}{\tau}. \tag{1}$$

For scale one ($\tau = 1$), the time series $\{y^{(1)}\}$ is the original series. Then, for each given τ, the original series is divided into N/τ coarse-grained series. The entropy of each coarse-grained time series can be calculated as follows.

Given a time series of N points: $\{u(j) : 1 \leq j \leq N\}$. Fix the vector length m and the tolerance for accepting matches r, r is defined in Refs. [17,18]. Let $x_m(i)$ with $\{i : 1 \leq i \leq N - m + 1\}$ where $x_m(i) = \{u(i+k) : 0 \leq k \leq m-1\}$. Two vectors $x_m(i)$ and $x_m(j)$ are called similar if the distance $d(x_m(i), x_m(j)) = \max\{|u(i+k) - u(j+k)| : 0 \leq k \leq m-1\}$ is smaller than r. Let $n_i^{(m)}$ be the number of vectors similar to $x_m(i)$, $n_i^{(m+1)}$ presents the number of matches of length $m+1$. Then SampEn is calculated with the equation:

$$\mathrm{SampEn}(m, r, N) = -\ln\left(\sum_{i=1}^{N-m} n_i^{(m+1)} \Big/ \sum_{i=1}^{N-m} n_i^{(m)}\right). \tag{2}$$

Finally, we plot the sample entropies over multiple scale factor τ.

2.2. Theory of MSTI

In order to extract the multiscale irreversibility from a time series, we assess the asymmetric distribution of all the points on multiple timescales. For a given time series, denoted by $x = \{x(i),\ i = 1, \ldots, N\}$, we construct the consecutive coarse-grained time series $\{x_\tau(j)\},\ 1 \leq j \leq N/\tau$ by averaging the data inside non-overlapping windows of τ points as Eq. (1), where τ is the scale factor.

The corresponding Δx_τ series can be obtained by the difference of two adjacent points (i.e. $x_\tau(k+1) - x_\tau(k)$) from the coarse-grained time series $\{x_\tau\}$. We can analyze the asymmetry from the Δx_τ series, because the number of increments ($\Delta x_\tau > 0$) of a symmetric series is equal to the number of decrements ($\Delta x_\tau < 0$). There are several indices proposed to measure the asymmetry. We consider the following indices:

(1) *Porta's index* [1,5], based on evaluating the percentage of negative Δx (i.e. Δx^-) with respect to the total number of $\Delta x \neq 0$. This index can be calculated as

$$P\% = \frac{N(\Delta x^-)}{N(\Delta x \neq 0)} \cdot 100. \tag{3}$$

This index ranges from 0 to 100. A $P\% > 50$ implies that the number of Δx^- is larger than Δx^+.

(2) *Guzik's index* [5,6], based on the evaluation of the percentage of the sum of the square values of positive Δx (i.e. Δx^+) to the cumulative square values of all Δx. It is expressed as

$$G\% = \frac{\sum_{i=1}^{N(\Delta x^+)} \Delta x^+(i)^2}{\sum_{i=1}^{N(\Delta x)} \Delta x(i)^2} \cdot 100. \tag{4}$$

This index ranges from 0 to 100. A $G\% > 50$ implies that the averaged magnitude of $\left|\Delta x^+\right|$ is larger than that of $\left|\Delta x^-\right|$ and the distribution of Δx is skewed towards positive values.