Applied Mathematics and Computation 289 (2016) 98–110

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

Modified multiscale cross-sample entropy for complex time series

Yi Yin, Pengjian Shang*, Guochen Feng

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

ARTICLE INFO

Keywords:
Refined composite multiscale cross-sample entropy (RCMCSE)
Composite multiscale cross-sample entropy (CMCSE)
Multiscale cross-sample entropy (MCSE)
Artificial time series
Stock indices

ABSTRACT

In this paper, we introduce the composite multiscale cross-sample entropy (CMCSE) which may induce undefined entropies and then further propose the refined composite multiscale cross-sample entropy (RCMCSE) which modifies CMCSE. First, we apply multiscale cross-sample entropy (MCSE), CMCSE and RCMCSE methods to three types of artificial time series in order to test the validity and accuracy of these methods. Results show that RCMCSE reduces not only standard deviation, but also the probability of inducing undefined entropy effectively, which can provide better robustness and more accurate entropies. Then, these three methods are employed to investigate financial time series including US and Chinese stock indices. For the study between stock indices in the same region, some conclusions which are consistent with previous study are drawn by the RCMCSE results. Meanwhile, it can be found that undefined entropies are induced and the numbers of inducing undefined entropy by three methods for investigation between three US stock indices and two Chinese mainland stock indices are given. Compared with MCSE and CMCSE, RCMCSE method is capable of reducing the number of undefined entropy and providing more accurate entropies. Moreover, the differences on inducing undefined entropy between results for US stock indices & two Chinese mainland stock indices and results for US stock indices & HSI demonstrate a much closer relation between US stock markets and HSI than between US stock markets and two Chinese mainland stock markets. Hence, it can be concluded that RCMCSE is more applicable for the study between US and Chinese stock markets.

1. Introduction

Recently, numerous techniques have focused on the analysis of nonlinear time series [1–3]. Meanwhile, many approaches have been proposed and applied to measure the complexity of time series in the complex systems [4–7]. As a measure of degree of uncertainty to detect the system complexity from time series, entropy has a wide application. Approximate entropy (ApEn) was introduced by Pincus to quantify the concept of changing complexity [8–11] and had been used to measure the biologic time series [11,12]. Furthermore, the shortcomings of the ApEn method were analyzed and sample entropy (SampEn) was developed by Richman et al., which agreed with theory results much more closely than ApEn over a broad range of conditions and had a wide applications in clinical cardiovascular studies [13,14]. Cross-sample entropy (Cross-SampEn) was introduced to measure the similarity of two distinct time series. Compared with correlation coefficient, Cross-SampEn is superior to describe the correlation between time series [15]. Generally, the larger the entropy is, the more random and

* Corresponding author. Tel.: +86 10 51684410.
E-mail address: pjshangdu@163.com (P. Shang).

http://dx.doi.org/10.1016/j.amc.2016.05.013
0096-3003/

complex a system is. However, an increase in the entropy may not always be associated with an increase in dynamical complexity. For example, when entropy-based algorithms are applied to real-world datasets obtained in health and disease states, we obtain contradictory findings [16]. These results may be attributed to the factor of scale. Thus, Costa et al. introduced the multiscale entropy (MSE) to calculate SampEn over a range of scales which can represent the complexity more comprehensively. MSE can resolve the contradiction and has been successfully applied to measure the complexity of time series generated from various dynamical systems including physiological signals [17–19] and vibrational signals [20,21]. Similarly, multiscale cross-sample entropy (MCSE) is proposed to show Cross-SampEn at different time scales and applied in the financial time series [22,23].

However, in the MCSE method, the coarse-graining procedure construct coarse-grained time series whose length is equal to N/τ, from the original N-points time series with the scale factor τ, which may result in an inaccurate entropy value for the coarse-grained time series at large scales may not be adequately long. Besides, the Cross-SampEn is undefined in some cases because no template vectors are matched to one another. The reliability of MCSE reduces due to the inaccurate or undefined Cross-SampEn. These issues in accuracy and validity challenge the application of MCSE method. As a result, inspired by [24] we propose the composite multiscale cross sample entropy (CMCSE) to address the accuracy concern of the MCSE method. CMCSE algorithm calculates the Cross-SampEns of all coarse-grained time series at a scale factor τ and then defines the CMCSE value as the means of τ Cross-SampEn values. CMCSE method provides more accurate entropy values but increases the probability of inducing undefined entropy. Hence, we modify the CMCSE algorithm and further propose the refined composite multiscale cross sample entropy (RCMCSE).

The remainder of this paper is organized as follows. Section 2 introduces the MCSE, CMCSE and RCMCSE methods briefly. In Section 3, three types of artificial time series are used to evaluate the effectiveness of these methods. Section 4 presents the application to the financial time series. A conclusion is drawn in Section 5.

2. Methodologies

2.1. Composite multiscale cross-sample entropy

MSE is based on the application of SampEn, which is proposed by Costa et al. [16,25] used MSE to analyze the biological time series, and succeed in separating healthy and pathologic groups. It has been indicated the difficulty in distinguishing the inter-beat interval time series of different diseased and healthy states if only a single-scale SampEn is used [16,25]. As a result, MSE is proposed. Similarly, MCSE is based on MSE and defined as using Cross-SampEn to solve the same problem in the study of two time series and analyze two time series over scale factor. The MCSE method is also based on Cross-SampEn. Thus, the MCSE algorithm consists of two procedures: 1) a coarse-graining procedure, which can be used to obtain the representations of the original time series on different time scales, and 2) the Cross-SampEn, which is capable of measuring the degree of the asynchrony of two time series. First, we review Cross-SampEn procedure briefly. Given two time series of N points: $u:\{u(j): j=1,\ldots,N\}$ and $v:\{v(j): j=1,\ldots,N\}$. From vector length m sequences

$$x_m(i) = (u(i), u(i+1), \ldots, u(i+m-1)), \{i: 1 \le i \le N-m+1\}, \tag{1}$$

$$y_m(j) = (v(j), v(j+1), \ldots, v(j+m-1)), \{j: 1 \le j \le N-m+1\}, \tag{2}$$

from u and v respectively. Let $n_i^{(m)}$ be the number of vectors $y_m(j)$ whose distance of $x_m(i)$

$$d(x_m(i), y_m(j)) = max\{|u(i+k) - v(j+k)| : 0 \le k \le m-1\} \tag{3}$$

is within the tolerance r. Similarly, $n_i^{(m+1)}$ is the number of matches of length $m+1$. Finally, Cross-SampEn is calculated with the equation:

$$Cross - SampEn(u, v, m, r) = -ln\left(\sum_{i=1}^{N-m} n_i^{(m+1)} \Big/ \sum_{i=1}^{N-m} n_i^{(m)}\right) \tag{4}$$

The essential feature of Cross-SampEn is to measure the degree of the asynchrony of two time series. The value of Cross-SampEn is higher, if the pair of series is more asynchronous [16,26,27].

Meanwhile, the coarse-graining procedure need construct coarse-grained time series from the original series u and v with the scale factor τ, respectively. Then we get $\{x^{(\tau)}\}$ and $\{y^{(\tau)}\}$. Each point of the coarse-grained time series is defined as

$$x_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} u_i, 1 \le j \le N/\tau \tag{5}$$

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} v_i, 1 \le j \le N/\tau \tag{6}$$

For scale one ($\tau = 1$), the time series $\{x^{(1)}\}$ and $\{y^{(1)}\}$ are the original series. Then, for each given τ, the original series is divided into N/τ coarse-grained series. MCSE method is to obtain the coarse-grained time series and then calculate the Cross-SampEn for each pair of coarse-grained series by Eq. (4). The plot of entropies over scale factor τ can be obtained to observe the change of Cross-SampEn with the scale more visually.

However, for MCSE method, the length of the time series is reduced by a factor of τ when applying the coarse-graining procedure to the original time series. For example, the time series with 1000 data points can be coarse-grained to a time series containing 50 data points at a scale factor of 20, which violates the length requirements in [18,28]. As a result, analyzing the short time series by MCSE method may lead to inaccurate Cross-SampEn at large time scales. As known to us, for solving the same problem in the multiscale sample entropy (MSE), Wu et al. [24] proposed the CMSE method which improves the accuracy of the MSE. Inspired by the CMSE method, we believe that the problem existing in the MCSE method can also be addressed by the similar way and then we propose the CMCSE algorithm, which calculates the Cross-SampEns of all coarse-grained time series first and then define the CMCSE value as the means of τ Cross-SampEns. Thus, the CMCSE is defined as follows:

$$CMCSE(x, y, \tau, m, r) = \frac{1}{\tau}\sum_{k=1}^{\tau} Cross - SampEn(x_k^{(\tau)}, y_k^{(\tau)}, m, r) = \frac{1}{\tau}\sum_{k=1}^{\tau} -ln\left(\frac{n_{k,\tau}^{m+1}}{n_{k,\tau}^{m}}\right) \tag{7}$$

where $n_{k,\tau}^{m}$ represents the total number of m-dimensional matched vector pairs and is obtained from the two k-th coarse-grained time series at a scale factor of τ.

In this paper, we calculate the entropies from scale 1 to scale 20, and the cross-sample entropy of each pair of coarse-grained series is calculated with $m = 2$ and $r = 0.15\sigma$ [18], where σ denotes the standard deviation (SD) of original time series.

2.2. Refined composite multiscale cross-sample entropy

As described above, the logarithms of the ratio of $n_{k,\tau}^{m+1}$ to $n_{k,\tau}^{m}$ for all τ coarse-grained series method are calculated first and the average of these logarithms is defined as the entropy value in the CMCSE method. However, if one of the values of $n_{k,\tau}^{m+1}$ or $n_{k,\tau}^{m}$ is zero, the CMCSE value is undefined. Hence, the probability of inducing undefined entropy is higher when employing CMCSE to the short time series than employing MCSE. This disadvantage makes the application of CMCSE method limited especially for short time series analysis. Therefore, we propose the RCMCSE method to address the validity issue. The procedure of RCMCSE method can be described as follows:

Step 1: The coarse-grained time series on different time scales can be obtained by the coarse-graining procedure described in Eq. (5,6).

Step 2: For a scale factor τ, the number of matched vector pairs, $n_{k,\tau}^{m+1}$ and $n_{k,\tau}^{m}$, is calculated for all coarse-grained series.

Step 3: Let $\bar{n}_{k,\tau}^{m}$ ($\bar{n}_{k,\tau}^{m+1}$) represent the mean of $n_{k,\tau}^{m}$ ($n_{k,\tau}^{m+1}$) for $1 \le k \le \tau$. The RCMCSE value at a scale factor τ is defined as the logarithm of the ratio of $\bar{n}_{k,\tau}^{m+1}$ to $\bar{n}_{k,\tau}^{m}$. In other words, the RCMCSE at a scale factor τ is given as Eq. (8)

$$RCMCSE(x, y, \tau, m, r) = -ln\left(\frac{\bar{n}_{k,\tau}^{m+1}}{\bar{n}_{k,\tau}^{m}}\right) \tag{8}$$

where $\bar{n}_{k,\tau}^{m+1} = \frac{1}{\tau}\sum_{k=1}^{\tau} n_{k,\tau}^{m+1}$ and $\bar{n}_{k,\tau}^{m} = \frac{1}{\tau}\sum_{k=1}^{\tau} n_{k,\tau}^{m}$. Eq. (8) can be simplified as

$$RCMCSE(x, y, \tau, m, r) = -ln\left(\frac{\bar{n}_{k,\tau}^{m+1}}{\bar{n}_{k,\tau}^{m}}\right) = -ln\left(\frac{\frac{1}{\tau}\sum_{k=1}^{\tau} n_{k,\tau}^{m+1}}{\frac{1}{\tau}\sum_{k=1}^{\tau} n_{k,\tau}^{m}}\right) = -ln\left(\frac{\sum_{k=1}^{\tau} n_{k,\tau}^{m+1}}{\sum_{k=1}^{\tau} n_{k,\tau}^{m}}\right) \tag{9}$$

According to Eq. (9), we find that the RCMCSE value is undefined only when all $n_{k,\tau}^{m+1}$ or $n_{k,\tau}^{m}$ are zeros. As a result, RCMCSE method reduces the probability of inducing undefined entropy, which is similar to the refined composite multiscale entropy [29].

3. Numerical experiments for artificial time series

Before discussing multiscale analysis of financial time series and its interpretation, we illustrate MCSE, CMCSE and RCMCSE methods to simulated examples where a judgement can be made about the validity and accuracy of these methods. In this section, we use three types of artificial time series, the two-exponent ARFIMA processes [30], multifractal binomial measures [31] and the NBVP time series [32] to test the effectiveness of these methods.

3.1. Two-component ARFIMA process

The power-law auto-correlations in stochastic variables can be modeled by an ARFIMA process [33]:

$$z(t) = Z(d, t) + \varepsilon_z(t), \tag{10}$$

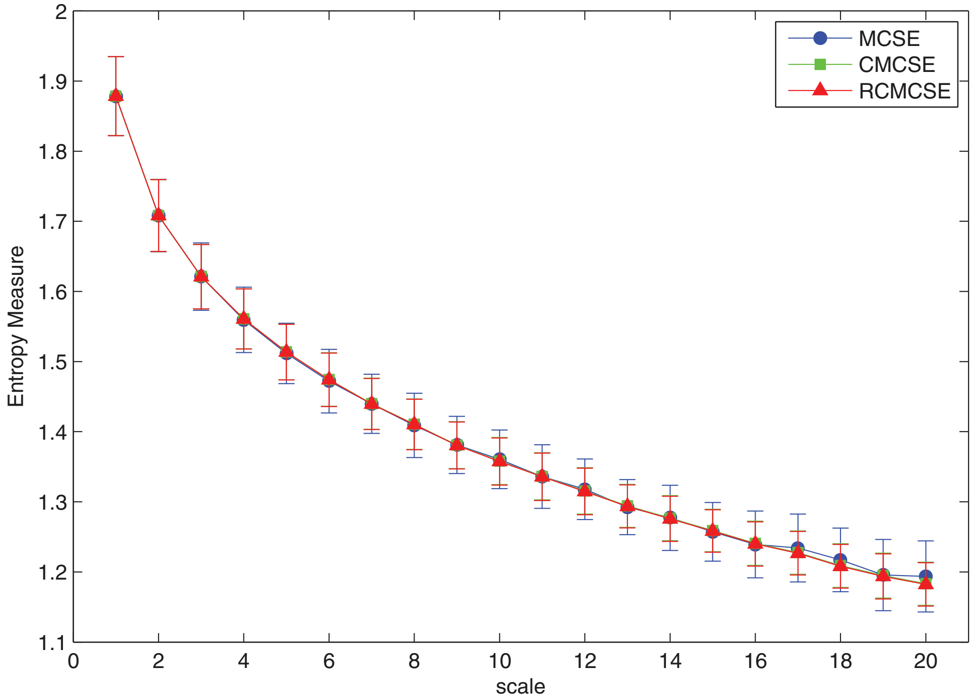

Fig. 1. MCSE, CMCSE and RCMCSE results between the series simulated by two-component ARFIMA process with $d_1 = d_2 = 0.4$. RCMCSE outperforms the other two methods in this case due to its smaller standard deviation.

where $d \in (0, 0.5)$ is a memory parameter, ε_z is an independent and identically distributed Gaussian variable, and $Z(d, t) = \sum_{n=1}^{\infty} a_n(d) z(t-n)$, in which $a_n(d)$ is the weight $a_n(d) = d\Gamma(n-d)/[\Gamma(1-d)\Gamma(n+1)]$. The Hurst index H_{zz} is related to the memory parameterized by [34,35]. For the two-component ARFIMA processes discussed below, we take $Z = X$ or Y. The two-component ARFIMA process is defined as follows [30]:

$$\begin{cases} x(t) = WX(d_1, t) + (1-W)Y(d_2, t) + \varepsilon_x(t) \\ y(t) = (1-W)X(d_1, t) + WY(d_2, t) + \varepsilon_y(t) \end{cases}, \tag{11}$$

where $W \in [0.5, 1]$ quantifies the coupling strength between the two processes $x(t)$ and $y(t)$. When $W = 1$, $x(t)$ and $y(t)$ are fully decoupled and become two separate ARFIMA processes as defined in Eq. (10). The cross-correlation between $x(t)$ and $y(t)$ increases when W decreases from 1 to 0.5 [30]. Besides, it is known to us that the DCCA exponent and DFA exponent hold $\alpha = (H_1 + H_2)/2$ [34].

Without loss of generality, we choose $W = 0.8$ and the parameters (d_1, d_2) of ARFIMA as $d_1 = d_2 = 0.4$ and $d_1 = 0.1$, $d_2 = 0.4$ separately, and corresponding two error terms $\varepsilon_x(t)$ and $\varepsilon_y(t)$ share one independent and identically distributed Gaussian variable with zero mean and unit variance. We apply MCSE, CMCSE and RCMCSE methods to the series simulated by two-component ARFIMA process. Figs. 1 and 2 show MCSE, CMCSE and RCMCSE results between the series simulated by two-component ARFIMA process with $d_1 = d_2 = 0.4$, and between the series simulated by two-component ARFIMA process with $d_1 = 0.1$, $d_2 = 0.4$, respectively. The numerical results are quantified by calculating 100 pairs of independent ARFIMA series with $d_1 = d_2 = 0.4$ and $d_1 = 0.1$, $d_2 = 0.4$ separately. Because 100 pairs of independent ARFIMA series are used and each series contains 4096 data points, we can obtain the means and standard deviations of entropies which have been shown in Figs. 1 and 2 and Tables 1 and 2. For ARFIMA series, there is no undefined entropy induced by these three methods, which means these three methods are applied to calculate the entropy values of scale 1–20 successfully. The entropy values decrease as the time scale increases which indicates the asynchrony of two time series decreases with the increasing scale. It can be seen that the means of these three methods are nearly equivalent, while the standard deviations of CMCSE and RCMCSE become smaller than that of MCSE as the scale increases indicating the validity of CMCSE and RCMCSE increase when the scale increases. Moreover, it can be found from Figs. 1 and 2 that the standard deviation of RCMCSE is a little smaller than that of CMSE for each time scale. Therefore, RCMCSE outperforms the other two methods and is the most reliable and applicable method for the analysis of ARFIMA series.

3.2. Multifractal binomial measures

We construct several binomial measures from the p-model with known analytic multifractal properties as a first example [36]. Each multifractal signal is obtained in an iterative way. We start with the zeroth iteration $g = 0$, where the data set

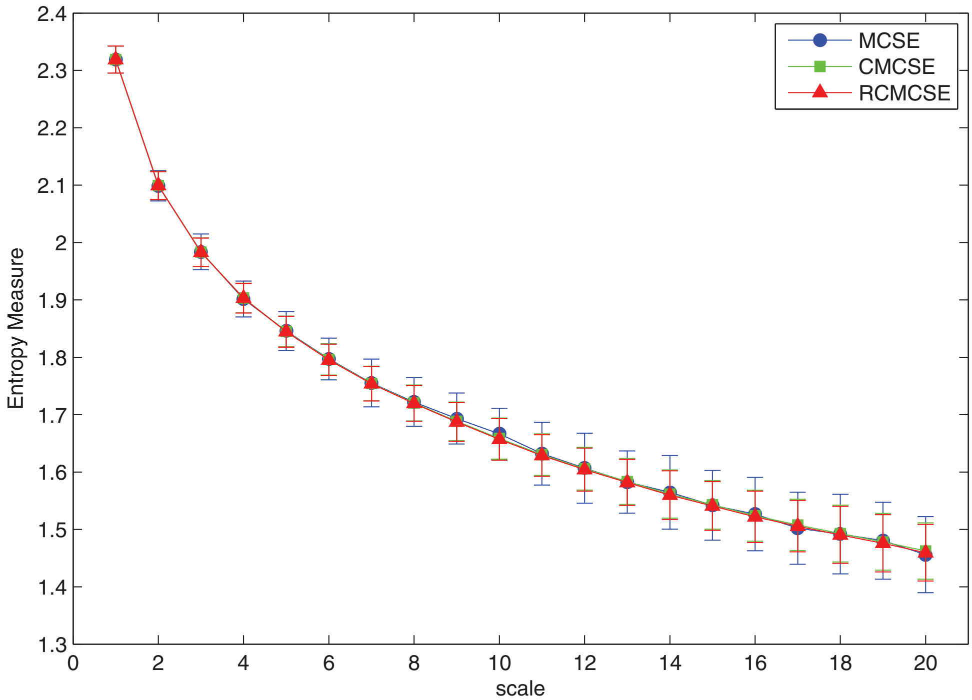

Fig. 2. MCSE, CMCSE and RCMCSE results between the series simulated by two-component ARFIMA process with $d_1 = 0.1, d_2 = 0.4$. RCMCSE outperforms the other two methods in this case due to its smaller standard deviation.

Table 1
Means and standard deviations of the MCSE, CMCSE and RCMCSE between the series simulated by two-component ARFIMA process with $d_1 = d_2 = 0.4$.

Scale	MCSE	CMCSE	RCMCSE
1	1.879 ± 0.056	1.879 ± 0.056	1.879 ± 0.056
3	1.621 ± 0.048	1.621 ± 0.046	1.621 ± 0.046
5	1.511 ± 0.043	1.514 ± 0.040	1.513 ± 0.040
7	1.440 ± 0.042	1.440 ± 0.036	1.439 ± 0.036
9	1.381 ± 0.041	1.381 ± 0.033	1.380 ± 0.033
11	1.336 ± 0.045	1.336 ± 0.034	1.336 ± 0.034
13	1.292 ± 0.039	1.294 ± 0.031	1.293 ± 0.030
15	1.257 ± 0.042	1.259 ± 0.030	1.258 ± 0.030
17	1.234 ± 0.048	1.228 ± 0.031	1.227 ± 0.031
19	1.196 ± 0.051	1.195 ± 0.032	1.194 ± 0.032
20	1.194 ± 0.051	1.183 ± 0.031	1.182 ± 0.031

Table 2
Means and standard deviations of the MCSE, CMCSE and RCMCSE between the series simulated by two-component ARFIMA process with $d_1 = 0.1, d_2 = 0.4$.

Scale	MCSE	CMCSE	RCMCSE
1	2.319 ± 0.023	2.319 ± 0.023	2.319 ± 0.023
3	1.984 ± 0.031	1.983 ± 0.025	1.983 ± 0.025
5	1.846 ± 0.034	1.845 ± 0.027	1.845 ± 0.027
7	1.755 ± 0.042	1.755 ± 0.030	1.754 ± 0.030
9	1.693 ± 0.044	1.688 ± 0.034	1.687 ± 0.034
11	1.632 ± 0.055	1.630 ± 0.036	1.629 ± 0.036
13	1.583 ± 0.054	1.584 ± 0.040	1.582 ± 0.040
15	1.542 ± 0.061	1.543 ± 0.042	1.541 ± 0.042
17	1.502 ± 0.063	1.508 ± 0.045	1.505 ± 0.045
19	1.480 ± 0.067	1.478 ± 0.049	1.476 ± 0.050
20	1.456 ± 0.066	1.462 ± 0.049	1.459 ± 0.049

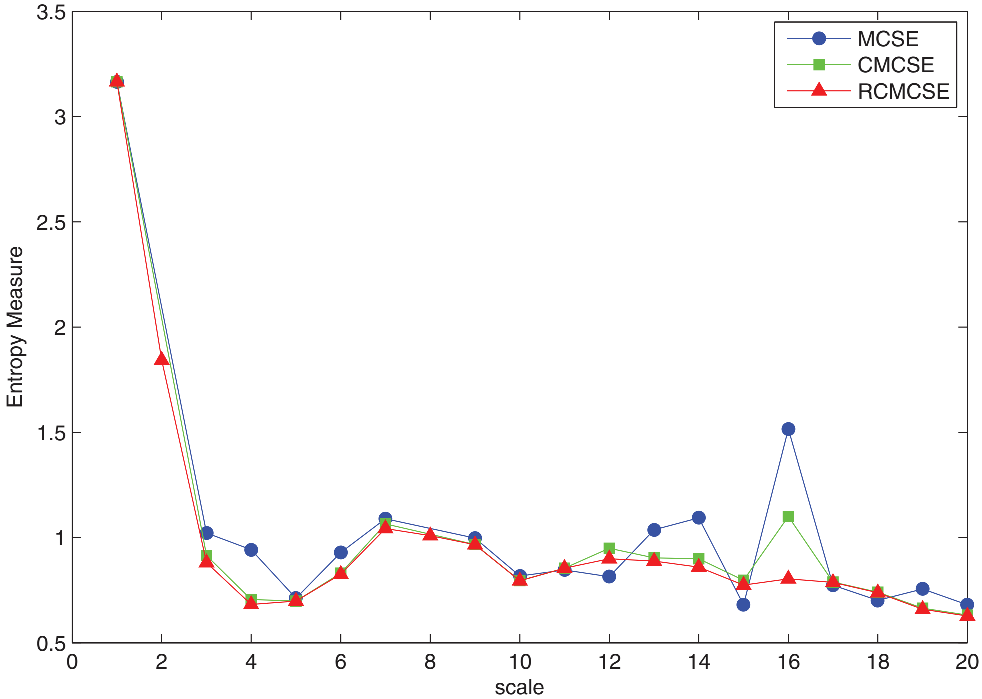

Fig. 3. MCSE, CMCSE and RCMCSE results between the series $p_x = 0.2$ for $\{x(i)\}$ and $p_y = 0.3$ for $\{y(i)\}$. The number of inducing undefined entropy can be reduced by RCMCSE in this case. The results of CMCSE and RCMCSE are more stable than those of MCSE, while the entropy values of RCMCSE are very close to those of CMCSE for the scales with defined entropy.

Table 3
The number of undefined entropy when MCSE, CMCSE and RCMCSE are applied to $\{x(i)\}$ & $\{y(i)\}$, $\{x(i)\}$ & $\{z(i)\}$ and $\{y(i)\}$ & $\{z(i)\}$ from scale 1 to 20.

Time series	MCSE	CMCSE	RCMCSE
$\{x(i)\}$ & $\{y(i)\}$	2	2	0
$\{x(i)\}$ & $\{z(i)\}$	5	5	1
$\{y(i)\}$ & $\{z(i)\}$	5	5	1

$z(i)$ consists of one value, $z^{(0)}(1) = 1$. In the gth iteration, the data set $z^{(g)}(i) : i = 1, 2, \ldots, 2^g$ is obtained from $z^{(g)}(2k+1) = pz^{(g-1)}(2k+1)$ and $z^{(g)}(2k) = (1-p)z^{(g-1)}(2k)$ for $k = 1, 2, \ldots, 2^{g-1}$. When $g \to \infty$, $z^{(g)}(i)$ approaches to a binomial measure, whose scaling exponent function $h_{zz}(q)$ has an analytic form [36,37]

$$H_{zz}(q) = 1/q - log_2[p^q + (1-p)^q]/q \tag{12}$$

In our simulation, we have performed $g = 12$ iterations with $p_x = 0.2$ for $\{x(i)\}$, $p_y = 0.3$ for $\{y(i)\}$ and $p_z = 0.4$ for $\{z(i)\}$. Each pair of these time series are correlated, which is originated from the fact that the two sequences are constructed according to the same rules [38]. We present the MCSE, CMCSE and RCMCSE results of the series $\{x(i)\}$ & $\{y(i)\}$, $\{x(i)\}$ & $\{z(i)\}$, $\{y(i)\}$ & $\{z(i)\}$ in Figs. 3, 4, 5, respectively. For the three pairs of binomial measures, the number of inducing undefined entropy is reduced by RCMCSE which not only can be found in these three figures, but also is shown in Table 3. Compared with the results of ARFIMA series which are monofractal, the results of these binomial measures with multifractality show more fluctuations as the time scale increases. The results of CMCSE and RCMCSE are more stable than those of MCSE, while the entropy values of RCMCSE are very close to those of CMCSE for the scales with defined entropy. Thus, we can conclude that the RCMCSE results are more accurate to reflect the real property between the binomial measures implying the higher validity and accuracy for RCMCSE algorithm.

3.3. NBVP time series

The Bonhoeffer-van der Pol (BVP) oscillator has a wide application and rich nonlinear behavior. Xue et al. [39] investigated Bonhoeffer-van der Pol (BVP) model time series and found significant multifractal features. Using the forward Euler

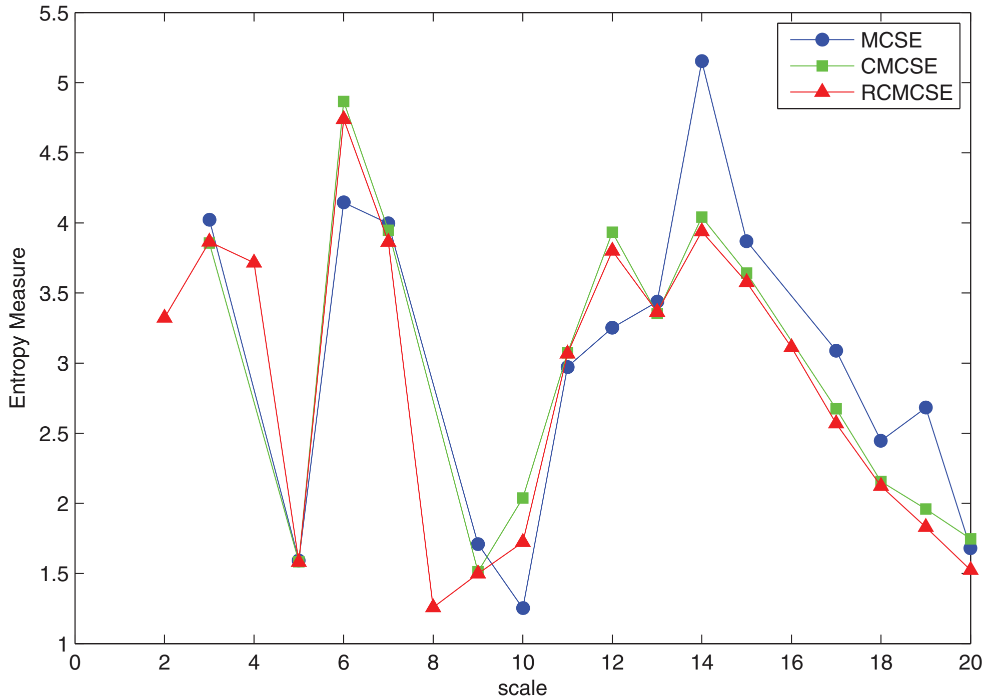

Fig. 4. MCSE, CMCSE and RCMCSE results between the series $p_x = 0.2$ for $\{x(i)\}$ and $p_z = 0.4$ for $\{z(i)\}$. The number of inducing undefined entropy can be reduced by RCMCSE in this case. The results of CMCSE and RCMCSE are more stable than those of MCSE, while the entropy values of RCMCSE are very close to those of CMCSE for the scales with defined entropy.

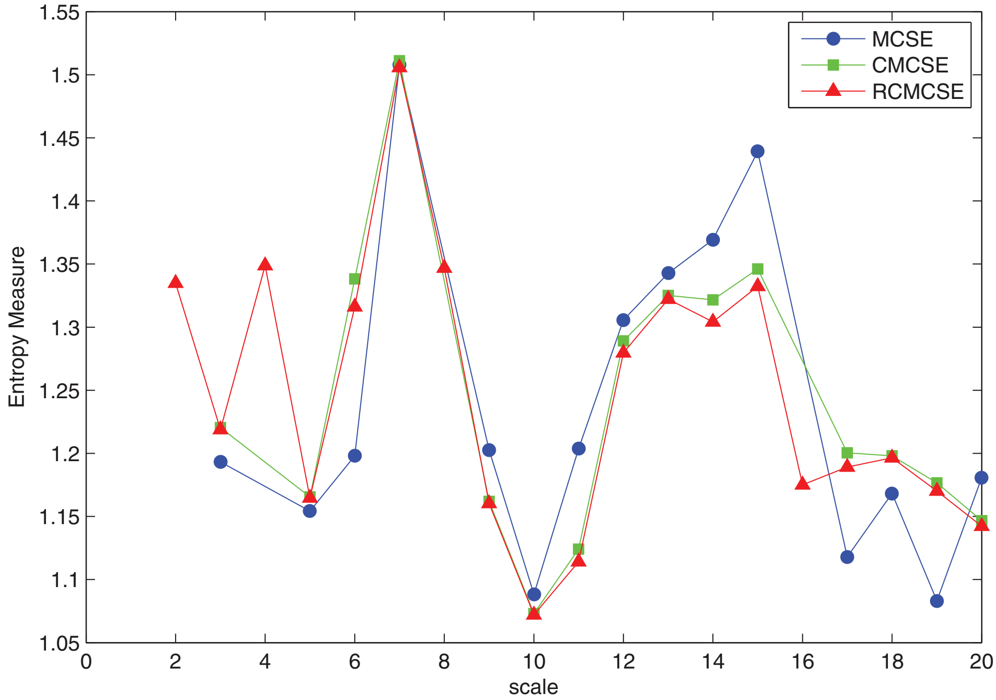

Fig. 5. MCSE, CMCSE and RCMCSE results between the series $p_y = 0.3$ for $\{y(i)\}$ and $p_z = 0.4$ for $\{z(i)\}$. The number of inducing undefined entropy can be reduced by RCMCSE in this case. The results of CMCSE and RCMCSE are more stable than those of MCSE, while the entropy values of RCMCSE are very close to those of CMCSE for the scales with defined entropy.

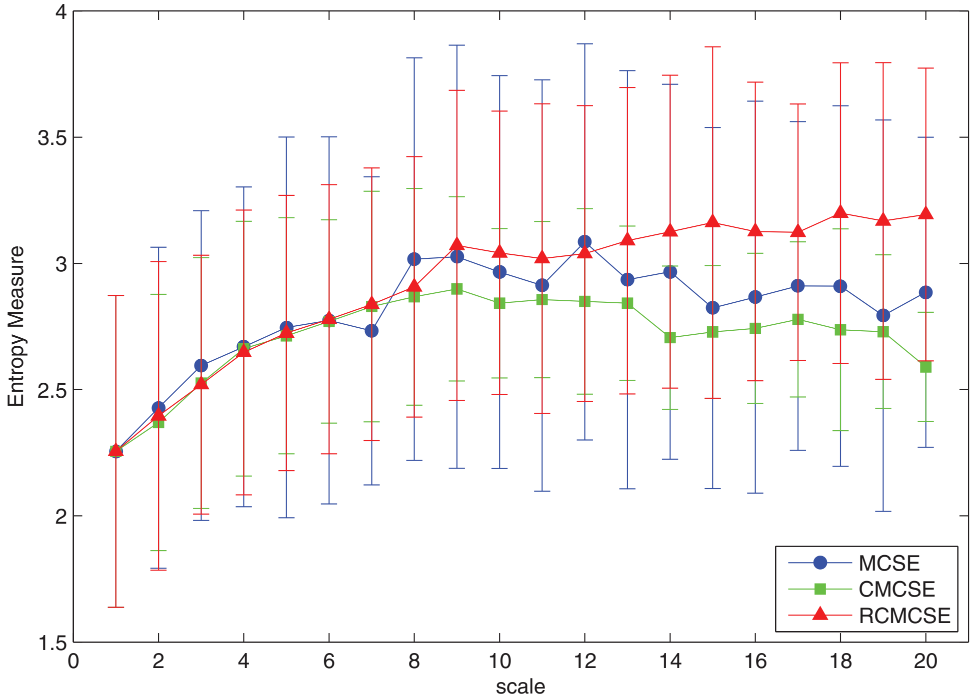

Fig. 6. MCSE, CMCSE and RCMCSE results between the NBVP time series. The standard deviation of CMCSE is the smallest among that of the three methods for all the scales, while standard deviation of RCMCSE is smaller than that of MCSE.

discrete scheme, a two-dimensional map derived from the classical continuous BVP oscillator is written as:

$$\begin{cases} x_{n+1} = x_n + \delta(y_n - \frac{1}{3}x_n^3 + x_n + \mu) \\ y_{n+1} = y_n + \delta\rho(a - x_n - by_n) \end{cases} \tag{13}$$

where $0 < \rho \ll 1$, $0 < a < 1$, $0 < b < 1$, μ is a stimulus intensity and $0 < \delta < 1$ is the step size. Given a set of variables and initial values x_0 and y_0, we can get time series $\{x(i)\}$ and $\{y(i)\}$. We generate two more complex BVP series with the length of $N = 10^{12}$ which are called NBVP time series, whose initial value μ is a random variable, satisfying $\mu \sim Norm(0.2, 1)$ and other initial values are constant values, with $\delta = 0.1$, $\rho = 0.001$, $a = 0.2$, $b = 0.5$, $x_0 = 0$ and $y_0 = 0.01$, where the cross-correlation relationship is found to be multifractal [32]. Then, we employ the MCSE, CMCSE and RCMCSE methods to the NBVP series to examine the validity of methods as well. The numerical results obtained by MCSE, CMCSE and RCMCSE methods are quantified by calculating 100 pairs of independent NBVP time series separately and the means and standard deviations of entropies are also shown in Fig. 6. The standard deviation of CMCSE is the smallest among that of the three methods for all the scales, while standard deviation of RCMCSE is smaller than that of MCSE. However, more and more undefined entropies are induced by CMCSE with the increasing scale, which can be noticed by the probabilities of inducing undefined entropy when applying MCSE, CMCSE and RCMCSE methods in Fig. 7. As a result, the smaller standard deviation of CMCSE can be attributed to the existence of undefined entropy in the CMCSE results. It is obvious that RCMCSE reduces the probabilities of inducing undefined entropy greatly by Fig. 7. Hence, compared with MCSE and CMCSE, RCMCSE reduces not only standard deviation, but also the probabilities of inducing undefined entropy, which can provide better robustness and more accurate entropies.

4. Analysis for financial time series

In order to validate the applicability of the proposed RCMCSE method for real data, we then apply MCSE, CMCSE and RCMCSE methods to financial time series. The analyzed data set consists of six indices: three US stock indices, S&P500, DJI and NQCI, and three Chinese stock indices, ShangZheng, ShenCheng, and HSI. Due to the different opening dates between the stock indices we investigated, we exclude the asynchronous datum and then reconnect the remaining parts of the original series to obtain the same length time series. As a result, the total of the closing prices recorded from January 3, 1997, to December 28, 2012 is 3572 days. Let x_t denote the closing price for a stock index on day t. Generally, the daily price return r_t, which is calculated as its logarithmic difference, i.e. $r_t = log(x_t) - log(x_{t-1})$ is used to investigate the time series of stock market. The daily price returns of six stock indices are shown in Fig. 8.

Thus, we employ the MCSE, CMCSE and RCMCSE methods to investigate the US and Chinese stock indices and present the results for the study between stock indices in the same region, i.e. USA and China, in Figs. 9 and 10, respectively. In Figs. 9 and 10, the entropies decrease with the increasing scale, which indicates the degrees of the asynchrony of these stock

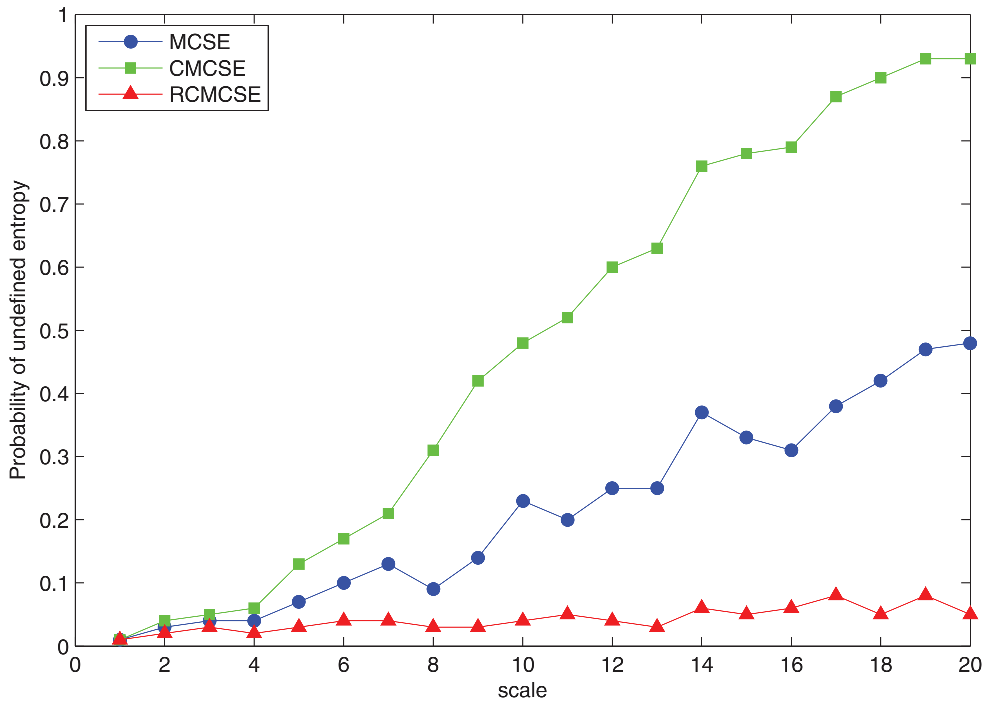

Fig. 7. Probabilities of inducing undefined entropy when MCSE, CMCSE and RCMCSE methods are applied to the NBVP time series. More and more undefined entropies are induced by CMCSE with the increasing scale, while RCMCSE reduces the probabilities of inducing undefined entropy greatly.

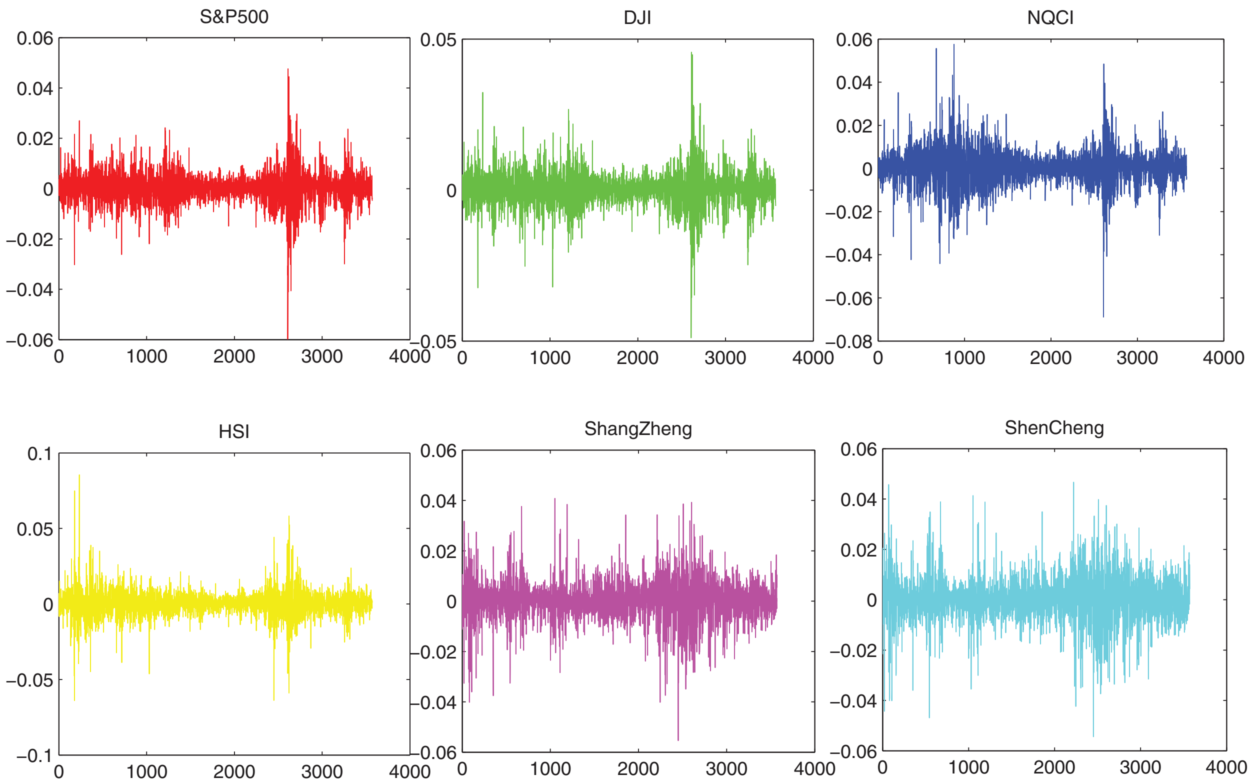

Fig. 8. Daily price returns for the S&P500, DJI, NQCI, ShangZheng, ShenCheng, and HSI stock indices.

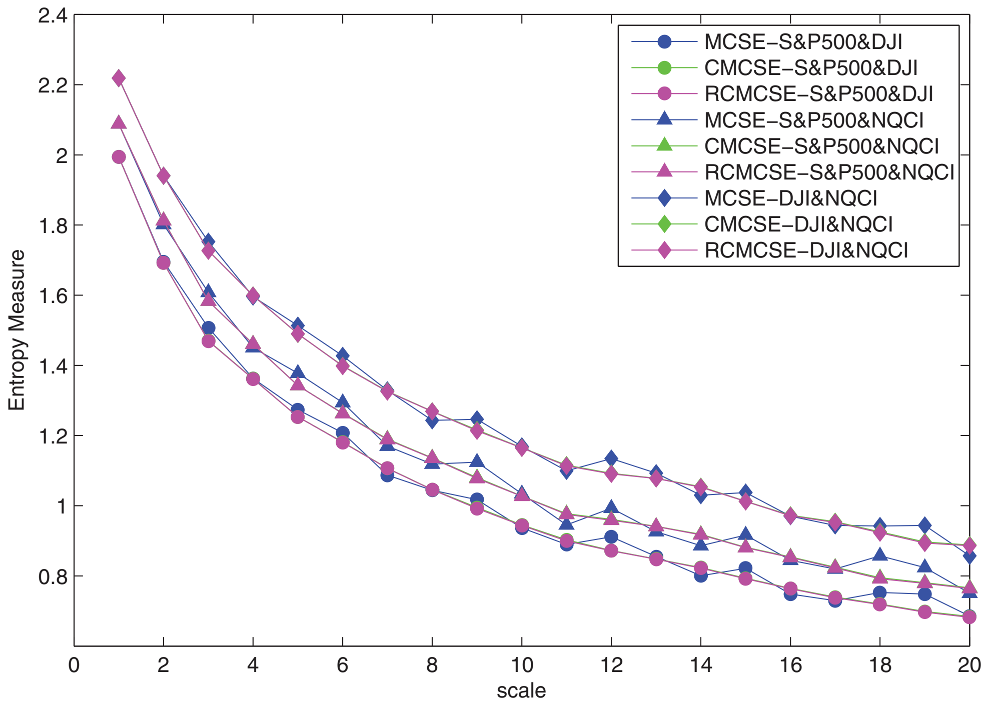

Fig. 9. MCSE, CMCSE and RCMCSE results between the time series in the US stock markets: S&P500 & DJI, S&P500 & NQCI and DJI & NQCI. The entropies decrease with the increasing scale, which indicates the degrees of the asynchrony of these stock indices reduce when the scale increases.

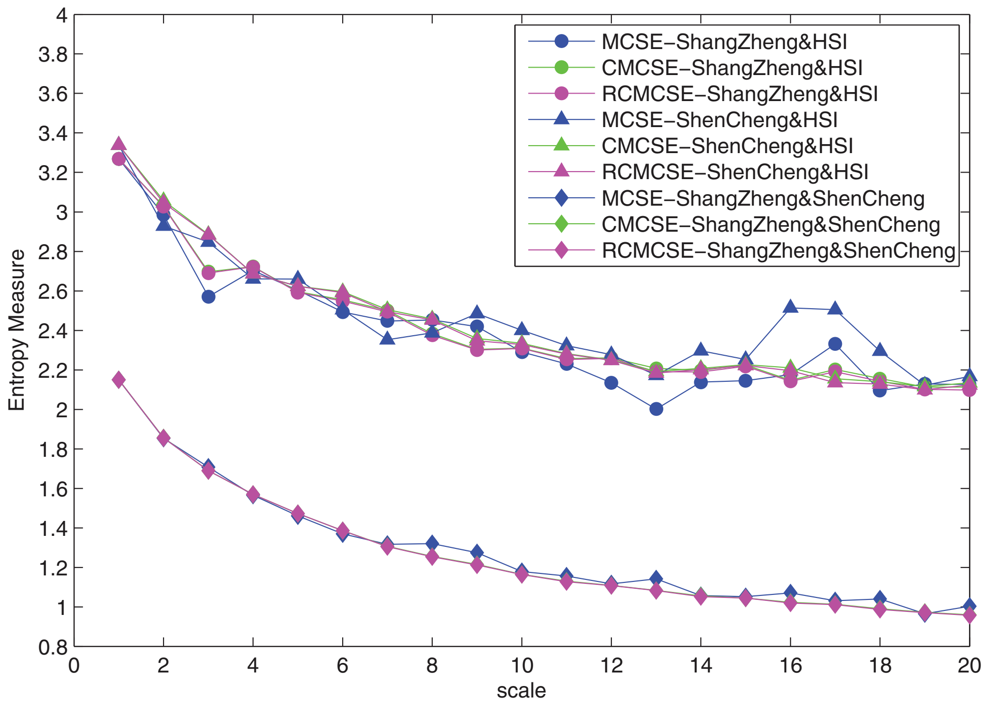

Fig. 10. MCSE, CMCSE and RCMCSE results between the time series in the Chinese stock markets: ShangZheng & HSI, ShenCheng & HSI and ShangZheng & ShenCheng. The entropies decrease with the increasing scale, which indicates the degrees of the asynchrony of these stock indices reduce when the scale increases.

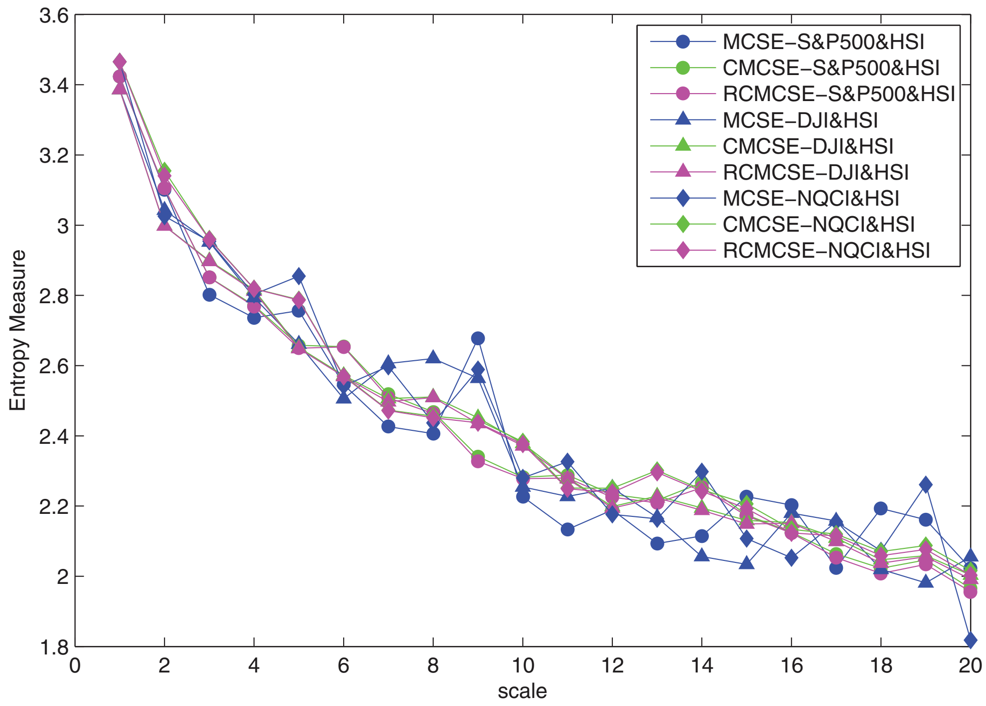

Fig. 11. MCSE, CMCSE and RCMCSE results between three US stock indices: S&P500, DJI and NQCI, and HSI. It can be seen that the results for three US stock indices and HSI are not influenced by the undefined entropies. The degrees of the asynchrony between three US stock indices: S&P500, DJI and NQCI, and HSI reduce when the scale increases, while their CMCSE and RCMCSE results are almost coincident and more stable than MCSE results.

Table 4
The number of undefined entropy when MCSE, CMCSE and RCMCSE are applied to the investigation between three US stock indices: S&P500, DJI and NQCI, and two Chinese mainland stock indices: ShangZheng and ShenCheng from scale 1 to 20.

Stock indices	MCSE	CMCSE	RCMCSE
S&P500 & ShangZheng	16	20	4
S&P500 & ShenCheng	13	17	1
DJI & ShangZheng	20	20	6
DJI & ShenCheng	14	19	1
NQCI & ShangZheng	16	18	1
NQCI & ShenCheng	11	18	0

indices reduce when the scale increases. The CMCSE and RCMCSE results are almost coincident and more stable than MCSE results. In Fig. 9, RCMCSE for DJI & NQCI has the biggest drop, while the RCMCSE values for DJI & NQCI still are the highest for all scales, implying that DJI and NQCI have the strongest asynchrony (the least cross-correlation). The RCMCSE values for S&P500 & DJI are the lowest for all scales which indicates the strong synchrony between S&P500 and DJI. Meanwhile, it can be found that the three kinds of entropies for ShangZheng & ShenCheng all separate from those of ShangZheng & HSI, ShenCheng & HSI, while the results for ShangZheng & ShenZheng are the lowest and RCMCSE results of ShangZheng & HSI are similar to those of ShenCheng & HSI. We can conclude that ShangZheng and ShenCheng have a lower asynchrony (stronger cross-correlation), while ShangZheng & HSI, ShenCheng & HSI hold higher and similar asynchronies, which means lower cross-correlations. It is worth noting that these findings by Figs. 9 and 10 are consistent with the conclusions obtained in [40].

Moreover, we also apply these three methods to the study between the stock indices from different region, i.e. US & China. However, when using MCSE, CMCSE and RCMCSE methods to the investigation between three US stock indices: S&P500, DJI and NQCI, and two Chinese mainland stock indices: ShangZheng and ShenCheng, undefined entropies are induced and the numbers of inducing undefined entropy by three methods are given in Table 4. It can be seen from Table 4 that MCSE induces some undefined entropies, while CMCSE leads to more undefined entropies when scale from 1 to 20, which implies that CMCSE is not appropriate due to inducing many undefined entropies. But RCMCSE method is capable of reducing the number of undefined entropy and providing more accurate entropies. Whereas, the results for three US stock indices and HSI are not influenced by the undefined entropies and are shown in Fig. 11. It is similar that the degrees of the asynchrony between three US stock indices: S&P500, DJI and NQCI, and HSI reduce when the scale increases, while their CMCSE and RCMCSE results are almost coincident and more stable than MCSE results. There is reason to believe that the differences on inducing undefined entropy between results for US stock indices & two Chinese mainland stock indices and results for US stock indices & HSI demonstrate a much closer relation between US stock markets and HSI than between US stock markets & two Chinese mainland stock markets. As a result, RCMCSE is more applicable for the study between US and Chinese stock markets.

5. Conclusions

In this paper, we introduce the CMCSE method to address the accuracy concern of the MCSE method. CMCSE method provides more accurate entropy values but increases the probability of inducing undefined entropy. Hence, we modify the CMCSE algorithm and further propose the RCMCSE method. To evaluate the validity and accuracy of MCSE, CMCSE and RCMCSE methods, we first apply these methods to three types of artificial time series: the two-exponent ARFIMA processes, multifractal binomial measures and the NBVP time series. For the analysis of ARFIMA series, RCMCSE has the smaller standard deviation, which outperforms the other two methods and is the most reliable and applicable method. For the three pairs of binomial measures, RCMCSE reduces the number of inducing undefined entropy and is more accurate to reflect the real property between the binomial measures implying the higher validity and accuracy for RCMCSE algorithm. When applying to NBVP time series, RCMCSE reduces not only standard deviation, but also the probabilities of inducing undefined entropy effectively, which can provide better robustness and more accurate entropies. Then we employ MCSE, CMCSE and RCMCSE methods to financial time series including three US stock indices and three Chinese stock indices. The degrees of the asynchrony of these stock indices reduce when the scale increases. For the study between stock indices in the same region, some conclusions which are consistent with the conclusions obtained in [39] can be drawn by the RCMCSE results. Meanwhile, it can be found that undefined entropies are induced and the numbers of inducing undefined entropy by three methods for investigation between three US stock indices and two Chinese mainland stock indices are given. MCSE and CMCSE are not appropriate due to inducing many undefined entropies, while RCMCSE method is capable of reducing the number of undefined entropy and providing more accurate entropies. Besides, the differences on inducing undefined entropy between results for US stock indices & two Chinese mainland stock indices and results for US stock indices & HSI suggest a much closer relation between US stock markets and HSI than between US stock markets & two Chinese mainland stock markets. Therefore, we conclude that RCMCSE is more applicable for the study between US and Chinese stock markets. In future, we may further study RCMCSE combining with other techniques, such as techniques used in [41,42] to measure the degree of the asynchrony between series with various trends in different fields and machine learning techniques [43,44] to analyze the financial time series.

Acknowledgments

This work is supported by “the Fundamental Research Funds for the Central Universities” (2016YJS160).

References

[1] S. Kodba, M. Perc, M. Marhl, Detecting chaos from a time series, Eur. J. Phys. 26 (2005) 205–215.
[2] M. Perc, The dynamics of human gait, Eur. J. Phys. 26 (2005) 525–534.
[3] A. Murks, M. Perc, Evolutionary games on visibility graphs, Adv. Complex Syst. 14 (2011) 307–315.
[4] J.S.A.E. Fouda, W. Koepf, Detecting regular dynamics from time series using permutations slopes, Commun. Nonlinear Sci. Numer. Simul. 27 (2015) 216–227.
[5] A.M. Lopes, J.A.T. Machado, Analysis of temperature time-series: embedding dynamics into the MDS method, Commun. Nonlinear Sci. Numer. Simul. 19 (2014) 851–871.
[6] A. Maus, J.C. Sprott, Neural network method for determining embedding dimension of a time series, Commun. Nonlinear Sci. Numer. Simul. 16 (2011) 3294–3302.
[7] K.P. Harikrishnan, R. Misra, G. Ambika, Combined use of correlation dimension and entropy as discriminating measures for time series analysis, Commun. Nonlinear Sci. Numer. Simul. 14 (2009) 3608–3614.
[8] S.M. Pincus, Approximate entropy as a measure of system complexity, Proc. Natl. Acad. Sci. USA 88 (1991) 2297–2301.
[9] S.M. Pincus, Approximate entropy (apen) as a complexity measure, Chaos 5 (1995) 110–117.
[10] S.M. Pincus, Quantifying complexity and regularity of neurobiological systems, Methods Neurosci. 28 (1995) 336–363.
[11] S.M. Pincus, R.R. Viscarello, Approximate entropy: a regularity measure for fetal heart rate analysis, Obstet. Gynecol. 79 (1992) 249–255.
[12] S.A. Schuckers, Use of approximate entropy measurements to classify ventricular tachycardia and fibrillation, J. Electrocardiol. 31 (1998) 101–105.
[13] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiol. Heart Circ. Physiol. 278 (2000) 2039–2049.
[14] D.E. Lake, J.S. Richman, M.P. Griffin, J.R. Moorman, Sample entropy analysis of neonatal heart rate variability, Am. J. Physiol. Regul. Integr. Comp. Physiol. 283 (2002) 789–797.
[15] L.Z. Liu, X.Y. Qian, H.Y. Lu, Cross-sample entropy of foreign exchange time series, Physica A 389 (2010) 4785–4792.
[16] M. Costa, A.L. Goldberger, C.K. Peng, Multiscale entropy analysis of biological signals, Phys. Rev. E 71 (2005) 021906.
[17] R.A. Thuraisingham, G.A. Gottwald, On multiscale entropy analysis for physiological data, Physica A 366 (2006) 323–332.
[18] M. Costa, C.K. Peng, A.L. Goldberger, J.M. Hausdorff, Multiscale entropy analysis of human gait dynamics, Physica A 330 (2003) 53–60.
[19] C.K. Peng, M. Costa, A.L. Goldberger, Adaptive data analysis of complex fluctuations in physiologic time series, Adv. Adapt. Data Anal. 1 (2009) 61–70.
[20] L. Zhang, G.L. Xiong, H.S. Liu, H.J. Zou, W.Z. Guo, Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference, Expert Syst. Appl. 37 (2010) 6077–6085.
[21] J.L. Lin, J.Y.C. Liu, C.W. Li, L.F. Tsai, H.Y. Chung, Motor shaft misalignment detection using multiscale entropy with wavelet denoising, Expert Syst. Appl. 37 (2010) 7200–7204.

[22] J. Xia, P. Shang, Multiscale entropy analysis of financial time series, Fluct. Noise Lett. 11 (2012) 1250033.
[23] Y. Yin, P. Shang, Comparison of multiscale methods in the stock markets for detrended cross-correlation analysis and cross-sample entropy, Fluct. Noise Lett. 13 (2014) 1450023.
[24] S.D. Wu, C.W. Wu, S.G. Lin, C.C. Wang, K.Y. Lee, Time series analysis using composite multiscale entropy, Entropy 15 (2013) 1069–1084.
[25] M. Costa, A.L. Goldberger, C.K. Peng, Multiscale entropy analysis of complex physiologic time series, Phys. Rev. Lett. 89 (2002) 068102.
[26] S.M. Pincus, B.H. Singer, Randomness and degrees of irregularity, Proc. Natl. Acad. Sci. USA 93 (1995) 2083–2088.
[27] S.M. Pincus, T. Mulligan, A. Iranmanesh, S. Gheorghiu, M. Godschalk, J.D. Veldhuis, Older males secrete luteinizing hormone and testosterone more irregularly, and jointly more asynchronously, than younger males, Proc. Natl. Acad. Sci. USA 93 (1996) 14100–14105.
[28] Q. Liu, Q. Wei, S.Z. Fan, C.W. Lu, T.Y. Lin, M.F. Abbod, J.S. Shieh, Adaptive computation of multiscale entropy and its application in EEG signals for monitoring depth of anesthesia during surgery, Entropy 14 (2012) 978–992.
[29] S.D. Wu, C.W. Wu, S.G. Lin, C.C. Wang, K.Y. Lee, C.K. Peng, Analysis of complex time series using refined composite multiscale entropy, Phys. Lett. A 378 (2014) 1369–1374.
[30] B. Podobnik, D. Horvatic, A.L. Ng, H.E. Stanley, P.C. Ivanov, Modeling long-range cross-correlations in two-component ARFIMA and FIARCH processes, Physica A 387 (2008) 3954–3959.
[31] H.O. Peitgen, H. Jurgens, D. Saupe, Chaos and fractals, Springer, Berlin, 1992.
[32] C. Xue, P. Shang, W. Jing, Multifractal detrended cross-correlation analysis of BVP model time series, Nonlinear Dyn. 69 (2012) 263–273.
[33] J.R.M. Hosking, Fractional differencing, Biometrika 68 (1981) 165–176.
[34] B. Podobnik, H.E. Stanley, Detrended cross-correlation analysis: a newmethod for analyzing two nonstationary time series, Phys. Rev. Lett. 100 (2008) 084102.
[35] B. Podobnik, P. Ivanov, K. Biljakovic, D. Horvatic, H.E. Stanley, I. Grosse, Fractionally integrated process with power-law correlations in variables and magnitudes, Phys. Rev. E 72 (2005) 026121.
[36] C. Meneveau, K.R. Sreenivasan, Simple multifractal cascade model for fully developed turbulence, Phys. Rev. Lett. 59 (1987) 1424–1427.
[37] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, B.I. Shraiman, Fractal measures and their singularities: the characterization of strange sets, Phys. Rev. A 33 (1986) 1141–1151.
[38] Z. Jiang, W. Zhou, Multifractal detrending moving-average cross-correlation analysis, Phys. Rev. E 84 (2011) 016106.
[39] C. Xue, P. Shang, J. Wang, Multifractal detrended cross-correlation analysis of BVP model time series, Nonlinear Dyn. 69 (2012) 263–273.
[40] Y. Yin, P. Shang, Modified DFA and DCCA approach for quantifying the multiscale correlation structure of financial markets, Physica A 392 (2013) 6442–6457.
[41] J. Wang, P. Shang, J. Xia, W. Shi, EMD based refined composite multiscale entropy analysis of complex signals, Physica A 421 (2015) 583–593.
[42] X. Zhao, P. Shang, A. Lin, G. Chen, Multifractal fourier detrended cross-correlation analysis of traffic signals, Physica A 390 (2011) 3670–3678.
[43] X. Xue, P. Yulei, A.S. Namin, Trimming test suites with coincidentally correct test cases for enhancing fault localizations, in: Proceeding of the 2014 IEEE 38th Annual Computer Software and Applications Conference (COMPSAC), IEEE, 2014, pp. 239–244.
[44] Y. Pang, F. Xiao, H. Wang, X. Xue, A clustering-based grouping model for enhancing collaborative learning, in: Proceeding of the 2014 13th International Conference on Machine Learning and Applications (ICMLA), IEEE, 2014, pp. 562–567.