J. Appl. Comput. Mech., 8(3) (2022) 1005-1013
DOI: 10.22055/jacm.2022.39470.3412

ISSN: 2383-4536
jacm.scu.ac.ir

Shahid Chamran University of Ahvaz

Journal of Applied and Computational Mechanics

Research Paper

Refined Cross-sample Entropy based on Freedman-Diaconis Rule: Application to Foreign Exchange Time Series

Javier E. Contreras-Reyes[1], Alejandro Brito[2]

[1] Instituto de Estadística, Facultad de Ciencias, Universidad de Valparaíso, Valparaíso, 2360102, Chile, Email: javier.contreras@uv.cl

[2] Instituto de Estadística, Facultad de Ciencias, Universidad de Valparaíso, Valparaíso, 2360102, Chile, Email: alejandro.brito@alumnos.uv.cl

Received December 16 2021; Revised January 27 2022; Accepted for publication January 27 2022.
Corresponding author: J.E. Contreras-Reyes (javier.contreras@uv.cl)

Abstract. Shang et al. (Commun. Nonlinear Sci. 94, 105556, 2022) proposed an efficient and robust synchronization estimation between two not necessarily stationary time series, namely the refined cross-sample entropy (RCSE). This method considered the empirical cumulative distribution function of distances using histogram estimator. In contrast to classical cross-sample entropy, RCSE only depends on a fixed embedding dimension parameter. In this paper, the RCSE is revisited as Freedman-Diaconis rule was considered to estimate the number of bins for the cumulative distribution function. Results are illustrated with some simulations based on 2D Hénon maps, the sinusoidal model, and the Lorenz attractor. In addition, a practical study of foreign exchange rate time series is presented. Specifically, the Canadian/US and Singaporean/US dollar time series were considered to compute the synchrony level between the 1995-1998 (before the 1999 Asian financial crisis) and the 1999-2003 (post-crisis) periods.

Keywords: Refined cross-sample entropy; Freedman-Diaconis rule; time series; 2D Hénon map; Lorenz attractor; foreign exchange market.

1. Introduction

Cross-sample entropy (CSE) was developed by Lake et al. [1] to measure the synchrony level between two time series. CSE is a non-negative value, under which a low CSE is obtained when two time series are highly synchronized, while CSE is larger for unsynchronized time series. This way, CSE also corresponds to an associative measure between time series, such as the correlation index [2,3]. However, CSE computation depends on two parameters: an embedding dimension m and tolerance r parameters. In general, values between 0.1 and 0.25 are considered for r and 1 or 2 for m, in time series of sample size 100-5,000 observations [1].

As discussed in [4] and [5], the selection of tolerance parameter r is ambiguous and produces several implications for sample entropy and CSE estimations. For sample entropy, [6] proposed the entropy profile to avoid the selection of r, while [5] redefined the sample and approximate entropy under a parametric approach and based on eigenvector analysis to eliminate the ambiguity arising from choosing tolerance parameter r and analyzed implications in sample entropy estimations. This method considered the empirical cumulative distribution function of distances using histogram estimator, i.e., the cumulative distribution function is often based on the number of bins. Recently, [7] proposed an efficient and robust synchronization estimation, namely the refined CSE (RCSE) measure, which extended entropy profile ideas to CSE and considered the cumulative histogram method (CHM) developed in [8]. In contrast to classical CSE, RCSE only depends on a fixed embedding dimension parameter.

Shang et al.'s work [7] motivated this paper, where number of bins was determined using the Freedman-Diaconis rule [9] as proposed in [10]. This in slight difference to Shang et al.'s approach, where number of bins were determined by the range of distance divided by parameter δ, obtained by the minimum of all distances. Another motivation of this paper was the study of foreign exchange rate time series, specifically the Canadian/US and Singaporean/US foreign exchange rate time series. Especially the synchrony level between the 1995-1998 (before the 1999 Asian financial crisis) and the 1999-2003 (post-crisis) time series were studied using the RCSE.

The paper is organized as follows. Section 2 starts by briefly presenting the profile sample entropy, then the refined cross-sample entropy based on Freedman-Diaconis rule is presented. Section 3 illustrates some simulations based on 2D Hénon maps, the sinusoidal model, and the Lorenz attractor. Section 4 presents a numerical application of RCSE on foreign exchange rate time series. Finally, Section 5 concludes the study.

2. Sample Entropy Methods

First, we present the profile sample entropy proposed by [4] and [6]. This measure considered the CHM and a multi-resolution binning data-based approach to get a potential set of values for tolerance parameter r. This set was used to construct the profile entropy as follows.

Published online: February 08 2022

2.1 Profile sample entropy

Let $\{x(t)\}$, $1 \le t \le N$, be a time series, not necessarily stationary, of length N. For a given embedding dimension parameter m, the following steps describe the construction of the profile entropy:

Step 1. Form $N-m$ vectors, each of length m and given by $\{X_i^m\}$, $1 \le i \le N-m$, where $X_i^m = x(i+k)$, $0 \le k \le m-1$.

Step 2. Analogously, form $N-m$ vectors, each of length $m+1$ and given by $\{X_i^{m+1}\}$, $1 \le i \le N-m$, where $X_i^{m+1} = x(i+k)$, $0 \le k \le m$.

Step 3. Take each vector $\{X_i^m\}$ of step 1 as template one and compute the Chebyshev distance (or infinity norm distance) for each $\{X_j^m\}$, as Chebyshev distance is given by

$$d_{ij}^m = \{\max_j |X_i^m - X_j^m| : 1 \le j \le N-m, j \ne i\} \tag{1}$$

where

$$\max_{\substack{1\le j\le N-m \\ j\ne i}} |X_i^m - X_j^m| = \lim_{\rho\to\infty} \left(\sum_{j=1}^{N-m} |X_i^m - X_j^m| \right)^{1/\rho}.$$

For any ith template vector, the distance vector for an embedding dimension parameter m is $d_i^m = d_{ij}^m$, $1 \le j \le N-m$, $j \ne i$. Analogously for an embedding dimension parameter $m+1$, the distance vector is $d_i^{m+1} = d_{ij}^{m+1}$, $1 \le j \le N-m$, $j \ne i$.

Step 4. Let $\boldsymbol{D}$ be the matrix that contains all the elements of d_i^m and d_i^{m+1}. The range is defined as the set of all unique elements of $\boldsymbol{D}$ and ordered in ascendent form. In addition, n_{bin} is the number of bins of this range.

Step 5. From distance vector d_i^m, the empirical cumulative distribution function $cdf_i^m(q)$ is computed as

$$cdf_i^m(q) = \mathbb{P}(d_i^m \le range(q)), \quad 1 \le q \le n_{bin},$$

where $\mathbb{P}$ denotes probability.

Step 6. Repeat step 5 for all d_i^m, $1 \le i \le N-m$, to compute complete matrix $\boldsymbol{C}^m$ of cumulative distributions for a given embedding dimension parameter m as

$$\boldsymbol{C}^m = \begin{pmatrix} cdf_1^m(1) & \cdots & cdf_1^m(n_{bin}) \\ \vdots & \ddots & \vdots \\ cdf_{N-m+1}^m(1) & \cdots & cdf_{N-m+1}^m(n_{bin}) \end{pmatrix}$$

Step 7. Analogously for embedding dimension parameter $m+1$, the complete matrix $\boldsymbol{C}^{m+1}$ of cumulative distributions is obtained by repeating step 4 and for all d_i^{m+1} as

$$\boldsymbol{C}^{m+1} = \begin{pmatrix} cdf_1^{m+1}(1) & \cdots & cdf_1^{m+1}(n_{bin}) \\ \vdots & \ddots & \vdots \\ cdf_{N-m}^{m+1}(1) & \cdots & cdf_{N-m}^{m+1}(n_{bin}) \end{pmatrix}$$

Step 8. For each column of matrices $\boldsymbol{C}^m$ and $\boldsymbol{C}^{m+1}$, the average of probabilities is computed, which can be represented respectively by

$$\theta^m(q) = \frac{1}{N-m} \sum_{i=1}^{N-m} cdf_i^m(q)$$

$$\theta^{m+1}(q) = \tfrac{1}{N-m} \textstyle\sum_{i=1}^{N-m} cdf_i^{m+1}(q), \quad 1 \le q \le n_{bin}.$$

Step 9. The sample entropy, which is an approximation of conditional probability of two segments that overlap in $m+1$ if they overlap in m, could be defined using θ^m and θ^{m+1} as $\log(\frac{\theta^m}{\theta^{m+1}})$. In this case, n_{bin} is the number of bins for sample entropy, because θ^m and θ^{m+1} are computed in interval $1 \le q \le n_{bin}$. The latter can be expressed as

$$SampEn(q) = \log\left(\frac{\theta^m(q)}{\theta^{m+1}(q)}\right), \quad 1 \le q \le n_{bin}.$$

Consider each value of q as a unique value of r. Therefore, n_{bin} as number of unique values of r is used for profile entropy. The algorithm described in steps 1-9 automatically selects n_{bin}, derived from signal dynamics. In this case, n_{bin} is obtained from the data (which involves all possible values of r for a given signal), and is hence the best selection to generate a complete profile of entropy. However, [4] and [6] did not present a statistical method to estimate n_{bin}. Shang et al. [7] considered the range of vectors to provide n_{bin} for the RCSE case. In this paper we considered the Freedman-Diaconis rule as follows.

2.2 Freedman-Diaconis rule

In many applications, it is easy to construct a histogram of the data, but the optimal choice of the partition is not known, especially when data length is not large, even if one restricts partitions to intervals of equal length. The Freedman-Diaconis binning

rule [9] is proposed to give an automatic, easy to program and efficient method to choose the number of bins for the partition from the data.

Let $\{y(i)\}$, $1 \le i \le N$, be any vector of observations of length N. The relative frequency of the histogram estimation is considered for $\{y(i)\}$, which is divided into n_{bin} bins. Let

$$IQR(\{y(i)\}) = Q_3(\{y(i)\}) - Q_1(\{y(i)\})$$

be the interquartile range of $\{y(i)\}$, where $Q_1(\{y(i)\})$ and $Q_3(\{y(i)\})$ are the first and third quartile of the set of observations $\{y(i)\}$. Then, n_{bin} was computed from the Freedman-Diaconis binning rule as

$$n_{bin}^{FD} = \frac{2IQR(\{y(i)\})}{N^{1/3}} \tag{2}$$

For a set of observations $\{y(i)\}$, $1 \le i \le N$, sampled from some probability distribution, the Freedman-Diaconis rule (2) minimized the theoretical integral of the squared difference between the relative frequency density and the theoretical one.

2.3 Refined cross-sample entropy

In contrast with the original CSE of [11], based on a fixed tolerance parameter r, the RCSE is based on CHM of [6] to obtain a range of entropies with several selected r in a determined range. In addition, the dissimilarity measure of the profile entropy is redefined, rather than the distance function used for CSE. The method proposed by [7] comprises the following steps:

Step 1. Let $\{u(t)\}$ and $\{v(t)\}$, $1 \le t \le N$, be two not necessarily stationary time series of length N.

Step 2. For a fixed and given embedding dimension parameter m, are extracted $N-m$ vectors of length m, $X_m(i) = \{u(i), u(i+1), \dots, u(i+m-1)\}$, $0 \le i \le N-m$, and $Y_m(j) = \{v(j), v(j+1), \dots, v(j+m-1)\}$, $0 \le j \le N-m$, for u and v, respectively.

Step 3. Permute vectors $X_m(i)$ and $Y_m(j)$ by the range of their respective elements and replace the elements of the vectors by integer numbers of set $\{0,1,2,\dots,m-1\}$. Then, $X_m(i)$ and $Y_m(j)$ are respectively permuted as $X_m^p(i)$ and $Y_m^p(j)$, where $X_m^p(i) = \{\tau_1, \tau_2, \dots, \tau_m\}$ and $Y_m^p(j) = \{\omega_1, \omega_2, \dots, \omega_m\}$, and τ_i and ω_j are the permuted integer numbers related to original elements.

Step 4. To define the distance/dissimilarity measure between both vectors, Kronecker's delta function is introduced for two variables:

$$\delta_{\tau_k\omega_k} = \begin{cases} 0, & if\ \tau_k \ne \omega_k, \\ 1, & if\ \tau_k = \omega_k. \end{cases}$$

In addition, the dissimilarity measure between two vectors extracted from two time series is defined as

$$d_{ij}^m = e^{-\theta_{ij}(m)} d[X_m(i), Y_m(j)], \quad \theta_{ij}(m) = \frac{1}{m}\sum_{k=1}^{m} \delta_{\tau_k\omega_k},$$

where the distance is based on Eq. (1) as

$$d[X_m(i), Y_m(j)] = \max_{\substack{0 \le k \le m-1 \\ 1 \le i,j \le N-m}} |u(i+k) - v(j+k)|.$$

Step 5. Distance vector d_i^m is constructed as $d_i^m = \{d_{i1}^m, d_{i2}^m, \dots, d_{i(N-m+1)}^m\}$. Analogously for vector of length $m+1$, the distance vector is constructed as $d_i^{m+1} = \{d_{i1}^{m+1}, d_{i2}^{m+1}, \dots, d_{i(N-m)}^{m+1}\}$.

Step 6. All distance vectors are reordered in ascendent form and put in a set $\{D_i\}$, $i = 1,2,\dots,n_e$, $n_e = (N-m+1)(N-m)$. Let

$$\lambda = \min_{\substack{1 \le i \le n_e-1 \\ D_i \ne D_{i+1}}} |D_i - D_{i+1}|,$$

thus the total number of all segments that are divided by the minimum length λ is

$$n_{bin} = \frac{\max\limits_{1 \le i \le n_e-1}\{D_i\} - \min\limits_{1 \le i \le n_e-1}\{D_i\}}{\lambda}. \tag{3}$$

In this paper, the Freedman-Diaconis rule given in (2) is proposed to obtain

$$n_{bin}^{FD} = \frac{2IQR(\{D(i)\})}{(N-m+1)^{\frac{1}{3}}(N-m)^{\frac{1}{3}}} \tag{4}$$

and, by matching Eq. (3) and Eq. (4), λ_{FD} parameter based on Freedman-Diaconis rule is

$$\lambda_{FD} = \frac{(N-m+1)^{\frac{1}{3}}(N-m)^{\frac{1}{3}}}{2IQR(\{D(i)\})}\left(\max_{1 \le i \le n_e-1}\{D_i\} - \min_{1 \le i \le n_e-1}\{D_i\}\right).$$

Step 7. For a distance vector d_i^m, the empirical cumulative distribution function

$$cdf_i^m(q\lambda_{FD}) = \mathbb{P}(d_i^m \le q\lambda_{FD}), \quad 1 \le q \le n_{bin}^{FD},$$

is computed.

Step 8. The complete cumulative distribution matrix is constructed by repeating step 7 for all d_i^m, $i = 1,2,\ldots,N-m+1$, as

$$\boldsymbol{C}^m = \begin{pmatrix} cdf_1^m(1\lambda_{FD}) & \cdots & cdf_1^m(n_{bin}^{FD}\lambda_{FD}) \\ \vdots & \ddots & \vdots \\ cdf_{N-m+1}^m(1\lambda_{FD}) & \cdots & cdf_{N-m+1}^m(n_{bin}^{FD}\lambda_{FD}) \end{pmatrix}.$$

Analogously, the complete cumulative distribution matrix for embedding dimension $m+1$ is

$$\boldsymbol{C}^{m+1} = \begin{pmatrix} cdf_1^{m+1}(1\lambda_{FD}) & \cdots & cdf_1^{m+1}(n_{bin}^{FD}\lambda_{FD}) \\ \vdots & \ddots & \vdots \\ cdf_{N-m}^m(1\lambda_{FD}) & \cdots & cdf_{N-m}^m(n_{bin}^{FD}\lambda_{FD}) \end{pmatrix}.$$

Step 9. Now, from $\boldsymbol{C}^m$ and $\boldsymbol{C}^{m+1}$, are respectively defined the averages

$$\Phi_{v|u}^m(q\lambda_{FD}) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} \log cdf_i^m(q\lambda_{FD}),$$

$$\Phi_{v|u}^{m+1}(q\lambda_{FD}) = \tfrac{1}{N-m}\textstyle\sum_{i=1}^{N-m} \log cdf_i^{m+1}(q\lambda_{FD}), \quad 1 \le q \le n_{bin}^{FD}.$$

Step 10. Previous steps treat $X_m(i)$ as template vector while $Y_m(j)$ is template overlapping vector (objective). To define the direction dependency, $\Phi_{v|u}^m(q\lambda_{FD})$ and $\Phi_{v|u}^{m+1}(q\lambda_{FD})$ are defined by changing $X_m(i)$ and $Y_m(j)$, and then by repeating steps 2-9. Then, are defined

$$\Gamma_m = \frac{\Phi_{v|u}^m(q\lambda_{FD}) + \Phi_{u|v}^m(q\lambda_{FD})}{2}$$

$$\Gamma_{m+1} = \frac{\Phi_{v|u}^{m+1}(q\lambda_{FD}) + \Phi_{u|v}^{m+1}(q\lambda_{FD})}{2}$$

to finally obtain

$$RCSE(N,m) = \Gamma_m - \Gamma_{m+1}.$$

The definition of RCSE is based on CHM and, considering all possible values of r, allows to observe the synchronization level between the time series in an exhaustive form (by adding one dimension) [7]. In general, the dissimilarity measures are based on spatial distances, yielding to erroneously decide that trends are similar although they have different spatial distances. A flowchart is presented in Fig. 1 to show the RCSE algorithm based on Freedman-Diaconis rule.

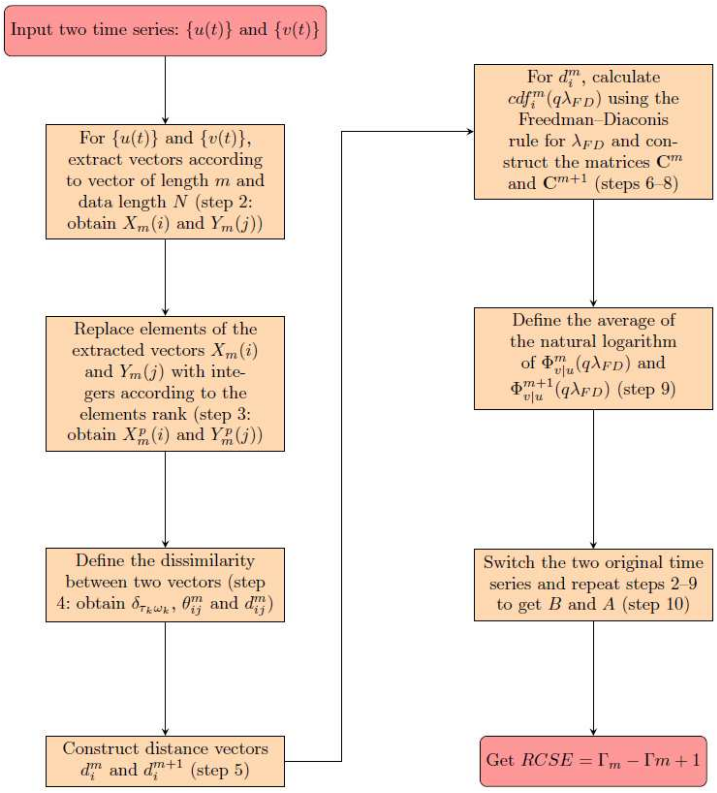

Fig. 1. Flowchart of RCSE algorithm based on Freedman-Diaconis rule.

RCSE was redefined in step 6 of Section 2.3 so that the number of bins was obtained through the Freedman-Diaconis rule. This modified the calculus of cumulative distribution function matrices (produced by CHM) and parameter λ_{FD} for the range (steps 7-10 of Section 2.3). Therefore, a new dissimilarity measure is proposed based on Kronecker's delta, Chebyshev distance, sequence of symbols and Freedman-Diaconis rule. This proportioned a reasonable and logic alternative to solve the synchronization problem of time series. This way, RCSE could estimate the synchronization of two different time series. RCSE is high under weak synchronization, while strong synchronization produces low RCSE.

3. Simulations

RCSE was implemented with R software [12] for three simulations: 2D Hénon map, the sinusoidal model, and the Lorenz attractor. All simulations were carried out with a unit featuring Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz, 2.00GHz, and 16.0GB RAM. R codes used in this paper will be made available upon reasonable request to the corresponding author.

3.1 2D Hénon map

The 2D Hénon map is considered because it provides time series to analyze the synchronization between them, using refined cross-sample entropy. The 2D Hénon map series [13] is defined as

$$\begin{cases} x(t+1) = 1 + y(t) - ax(t)^2, \\ y(t+1) = bx(t), \end{cases}$$

The values of the two parameters a and b varied for generating three sets of series: $a = 1.4$, $b = 0.3$ (Hénon 1), $a = 1.3$, $b = 0.35$ (Hénon 2), and $a = 1.1$, $b = 0.35$ (Hénon 3). For each set of time series, the initial value was $x(1) = y(1) = 0$. The schematic diagrams of the Hénon maps are shown in Fig. 2.

Results were obtained by applying RCSE on simulated times series produced by Hénon 1, 2 and 3, for $m = 3, 4$ and 5. Specifically, RCSE was computed between $\{x(t)\}$ and $\{y(t)\}$ for the three cases, which varied according to parameters a and b. Hénon 1 time series were chaotic, thus the scale of Fig. 3 is based on RCSE of this map. RCSE based on Hénon 3 was more fluctuant and showed the biggest difference with respect to Hénon 2 for all r. This indicates the synchrony level of $\{x(t)\}$ and $\{y(t)\}$ from Hénon 1 and 2 was the most similar, in contrast to synchrony level of $\{x(t)\}$ and $\{y(t)\}$ from Hénon 1 and 3, where the synchrony level of Hénon 3 was stronger (lower RCSE, implying more synchrony) than of Hénon 1, for all r. For the three cases, the time series could be considered as irregular perturbations, so the effect of r on RCSE was faint ($r > 1.5$). When m increased, in general RCSE tended to decrease for all Hénon maps.

3.2 Sinusoidal model

The sinusoidal model involves multiple time series, which combine the normally distributed errors $\{\varepsilon(t)\}$ with sinusoidal functions to obtain noisy and periodical components [10]. For ith time series, the sinusoidal model is

$$x_i(t) = \beta_i \sin(\frac{2\pi t}{T_i}) + \varepsilon(t),$$

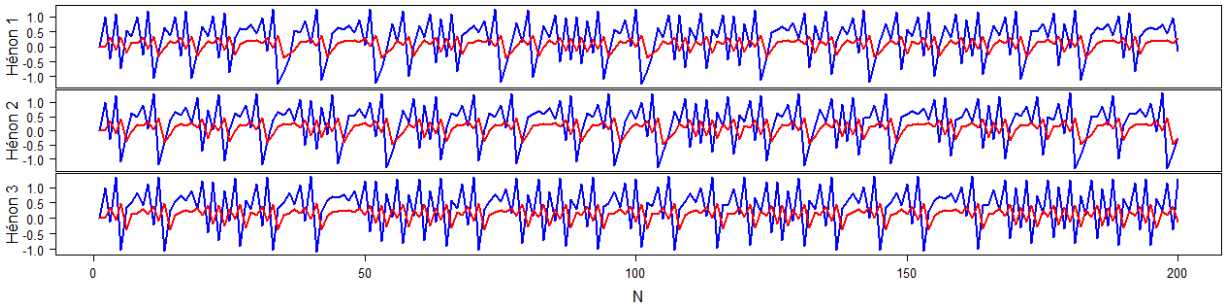

Fig. 2. Schematic diagrams of time series produced by the 2D Hénon map, where blue and red lines correspond respectively to $\{x(t)\}$ and $\{y(t)\}$ time series.

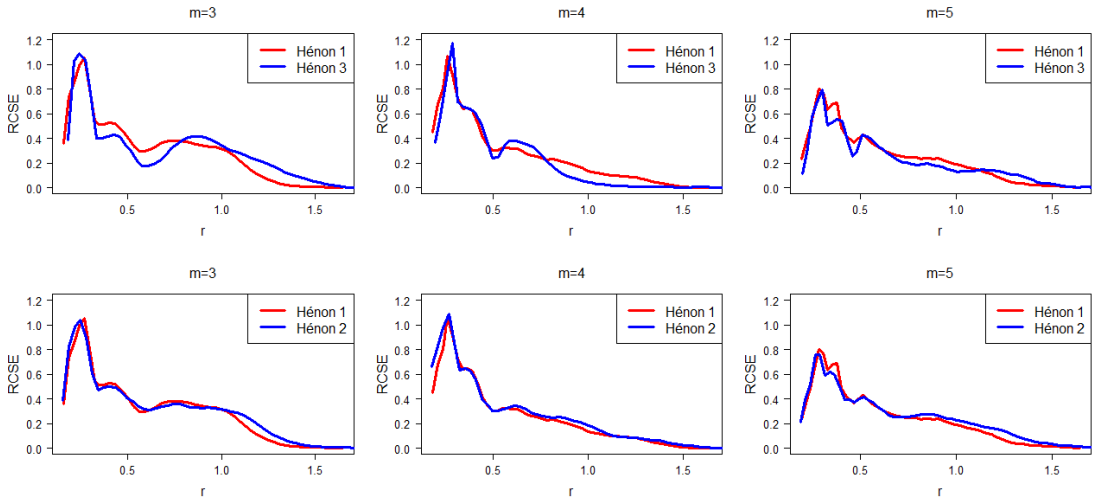

Fig. 3. The relationship between RCSE and r of the time series produced by the 2D Hénon map.

with $\varepsilon(t)\sim N(0,1)$, $i=1,\ldots,L$, $t=1,\ldots,N$. In addition, $T_i=5i$ and

$$\beta_i=\begin{cases}10, & if\ i\le 20,\\ 0, & if\ i>20.\end{cases}$$

The simulated sinusoidal model considered L=50 and N=100. Fig. 4 illustrates three time series: the periodical type $\{x_1(t)\}$, the periodical type with a larger period $\{x_5(t)\}$, and the noisy type $\{x_{50}(t)\}$ with irregular vibrations. RCSE was applied to simulated times series produced for $m=2,3,4$ and 5. Specifically, RCSE was computed between $\{x_1(t)\}$ and $\{x_5(t)\}$, and between $\{x_1(t)\}$ and $\{x_{50}(t)\}$. Results are shown in Fig. 5. In general, the synchrony level of $\{x_1(t)\}$ and $\{x_5(t)\}$ was the weakest given the difference of periods, in contrast to synchrony level of $\{x_1(t)\}$ and $\{x_{50}(t)\}$, where the synchrony level was stronger (lower RCSE) for all r.

3.3 Lorenz attractor

The Lorenz model (see e.g. [14, 15]) is a chaotic system which has a robust singular strange attractor. The Lorenz attractor is given by the following system of differential equations:

$$\begin{cases}\dfrac{dx}{dt}=a(y-x),\\ \dfrac{dy}{dt}=cx-y-xz,\\ \dfrac{dz}{dt}=xy-bz,\end{cases}$$

where a, b, and c are three real positive parameters; and x, y and z generate the tridimensional space. Classical chaotic behavior is obtained by letting $a=10$, $b=8/3$, and $c=150$.

Numerical solutions were developed using the Euler-Maruyama method over trajectories of size $N=1000$. $\Delta t=t_i-t_{i-1}=T/N=0.1$ is the length of the time discretization subinterval $[t_i,t_{i-1}]$, $i=1,\ldots,T$, with $T=100$. Fig. 6 illustrates the simulated trajectories for axes x and y, where both are characterized by periodical behavior. For practical purposes, trajectory z was excluded from the analysis because it is on another scale, different from x and y. RCSE between x and y is given for $m=2,3,4$ and 5, which indicates that synchronization of trajectories increases when m increase.

The computational cost in the calculus of RCSE based on Freedman-Diaconis rule is considerably lighter related to classical CSE. This is because CSE searches for coincidences of distances less than tolerance parameter r using minmax procedure, while RCSE estimates directly the number of bins using the Freedman-Diaconis rule. Besides, calculating minmax distance is not trivial from a computational perspective, especially for smaller values of m [1]. Consider that several methods to estimate the number of bins exist, such as Sturges or Doane, which were considered but discarded because they produced too similar results.

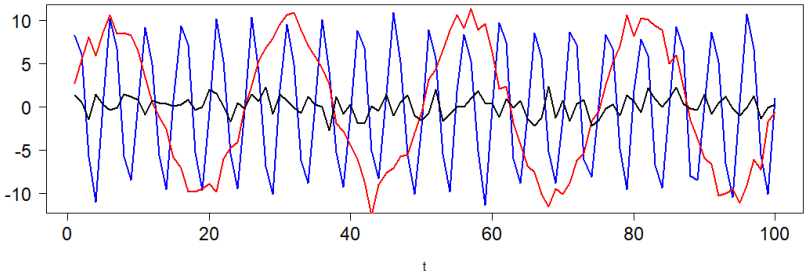

Fig. 4. Components $\{x_1(t)\}$ (blue line), $\{x_5(t)\}$ (red line), and $\{x_{50}(t)\}$ (black line) of sinusoidal model.

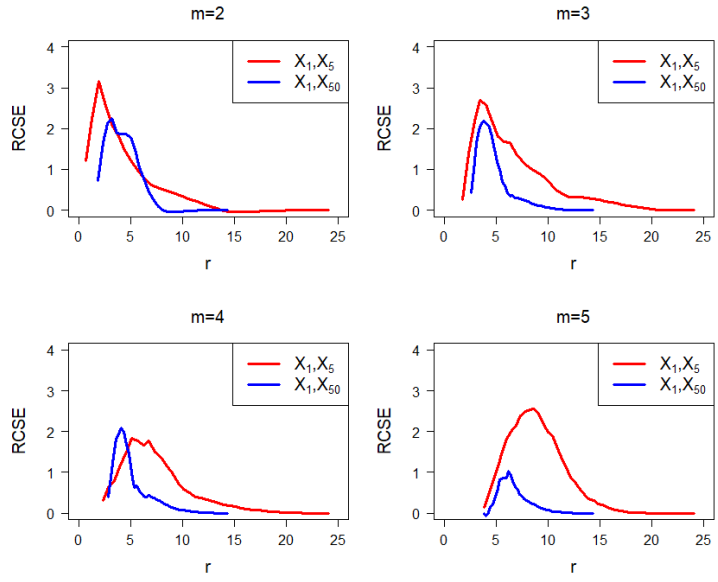

Fig. 5. The relationship between RCSE and r of components $\{x_1(t)\}$, $\{x_5(t)\}$, and $\{x_{50}(t)\}$ produced by the sinusoidal model.

4. Application

Financial markets are complex systems where several elements interact with one another to determine asset prices in dynamical form. This section considers RCSE based on Freedman-Diaconis rule to analyze the synchronization between Canadian to US (CAD/USD) and Singaporean to US (SGD/USD) foreign exchange time series, and compare values from before and after the Asian crisis. Among several foreign exchange rate time series analyzed in the literature [16-18,5], we specifically considered the SGD/USD time series because the CSE values reported by [16,5] after the Asian currency crisis increased significantly compared to those before this crisis. In contrast, as [17,5] reported, the CSE values related to CAD/USD show less synchronism before the Asian currency crisis. Therefore, the daily exchange rates of CAD/USD and of SGD/USD time series were considered to contrast an Asian and a North American market. Specifically, the synchrony level of the 1995-1998 time series (before the 1999 Asian financial crisis) was studied and the 1999-2003 time series, just after the crisis, using the RCSE, i.e., four time series were considered. The RCSE was applied to standardized log-returns of CAD/USD and SGD/USD time series, quantifying the impact of policies in response to the 1999 Asian financial crisis' effects on foreign exchanges.

Let $\{P(t)\}$ be a foreign exchange rate time series, where the log-return of $P(t)$ is given by

$$R(t) = \log(\frac{P(t)}{P(t-1)}), \quad t = 1, \ldots, N.$$

$R(t)$ is the daily appreciation/depreciation rate on concurrences with respect to USD. Fig. 4 illustrates the log-returns of both foreign exchange rate time series, where the 1999 Asian financial crisis is highlighted as producing high volatility. It becomes clear that high SGD/USD variability was produced in the middle of 1997. The same occurred at the end of 1998 for the CAD/USD case.

RCSE are illustrated in Fig. 5 for embedding dimension parameters $m = 2, 3, 4$ and 5. The blue lines represent higher values than the red lines, implying that RCSE of foreign exchange rate log-returns after the Asian crisis were higher than those produced before the crisis. Results reveal that foreign exchange rate log-returns before crisis had larger synchronization than post-crisis ones, in line with the fact that in the period posterior to the crisis foreign exchange volatility increased. This result as the exchange rate regime of Singapore moved around the US dollar before the Asian crisis and had moved to manage floating after the crisis [16]. This is consistent with the findings of [5]. Indeed, asynchrony was produced after the Asian crisis, given that exchange rate regimes or liquidity and the policies changed after the shock.

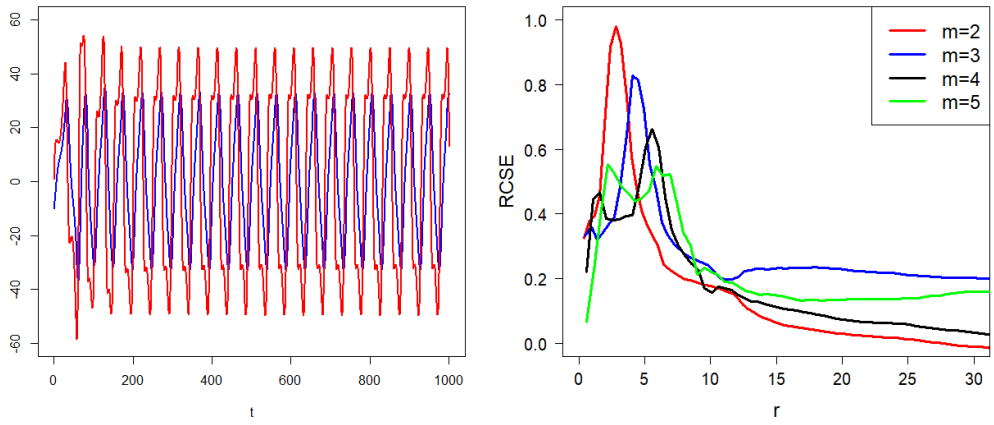

Fig. 6. Left: Trajectories x (blue) and y (red) produced by Lorenz attractor. Right: Relationship between RCSE and r of trajectories x and y.

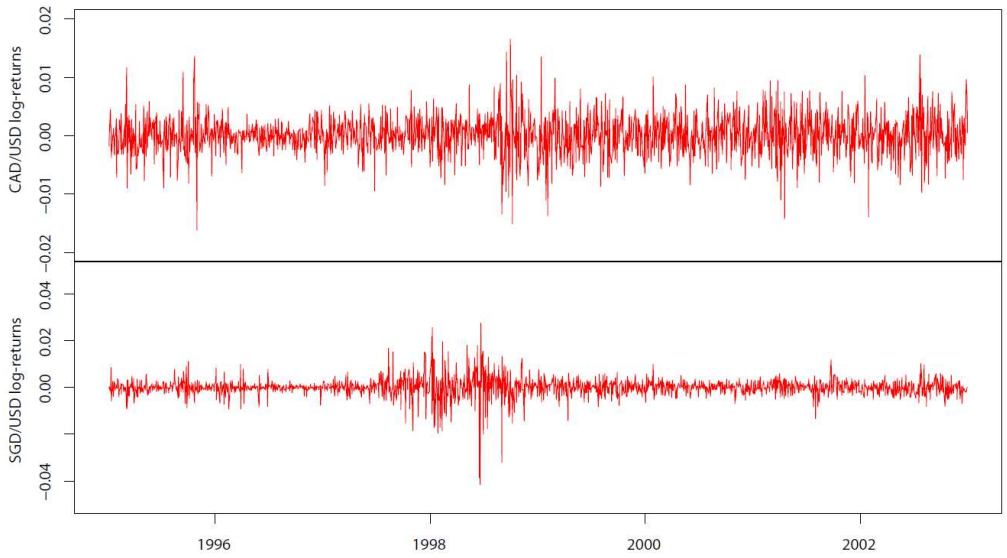

Fig. 7. Log-returns of CAD/USD and SGD/USD foreign exchange rate time series.

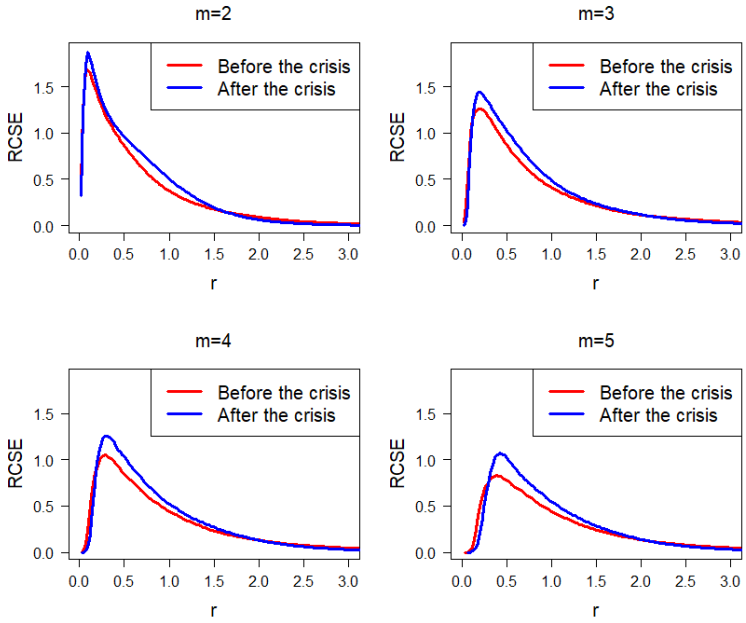

Fig. 8. The relationship between RCSE and r of log-returns of CAD/USD and SGD/USD foreign exchange rate time series. RCSE were obtained by comparing the log-returns before the Asian crisis (blue lines) and after the crisis (red lines).

5. Concluding Remarks and Further Works

In this paper, the profile entropy and refined cross-sample entropy have been revised. Specifically, the refined cross-sample entropy was redefined by the histogram estimator based on Freedman-Diaconis rule, which allowed finding the number of bins and posteriorly the complete cumulative distribution matrices for any embedding dimension. As further work, more complex simulated time series could be obtained by the use of chaotic attractors in a three-dimensional space [15] or with more dimensions [14], especially when time series are perturbed with Gaussian [19, 20] or non-Gaussian [15, 21] noise. Besides, the synchronization of asymmetric chaotic maps [20] and its characterization in the Fisher-Shannon information plane [21] could be studied. The construction of refined cross-sample entropy based on time series affected by common external factors [22] could provide an important insight to analyze refined cross-sample entropy properties. Finally, researchers may consider the proposed methodology for further investigations, including the study of nonlinear interactions between biological and other climatic variables [23, 24].

Author Contributions

J.E. Contreras-Reyes planned the scheme, initiated the project, and suggested the experiments; A. Brito conducted the experiments and analyzed the empirical results; J.E. Contreras-Reyes developed the mathematical modeling and examined the theory validation. The manuscript was written through the contribution of all authors. All authors discussed the results, reviewed, and approved the final version of the manuscript.

Acknowledgments

The authors thank the editor and three anonymous referees for their helpful comments and suggestions.

Conflict of Interest

The authors declared no potential conflicts of interest concerning the research, authorship, and publication of this article.

Funding

This study was funded by FONDECYT (Chile) grant No. 11190116.

Data Availability Statements

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

[1] Lake, D.E., Richman, J.S., Grin, M.P., Moorman, J.R., Sample entropy analysis of neonatal heart rate variability, *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 283(3), 2002, R789R797.
[2] Wang, F., Zhao, W., Jiang, S., Detecting asynchrony of two series using multiscale cross-trend sample entropy, *Nonlinear Dynamics*, 99(2), 2020, 1451-1465.
[3] Contreras-Reyes, J.E., Idrovo-Aguirre, B.J., Backcasting and forecasting time series using detrended cross-correlation analysis, *Physica A: Statistical Mechanics and its Applications*, 560, 2020, 125109.
[4] Karmakar, C., Udhayakumar, R., Palaniswami, M., Entropy Profiling: A Reduced-Parametric Measure of Kolmogorov-Sinai Entropy from Short-Term HRV Signal, *Entropy*, 22, 2020, 1396.
[5] Ramírez-Parietti, I., Contreras-Reyes, J.E., Idrovo-Aguirre, B.J., Cross-sample entropy estimation for time series analysis: a nonparametric approach, *Nonlinear Dynamics*, 105(3), 2021, 2485-2508.

[6] Udhayakumar, R. K., Karmakar, C., Palaniswami, M., Approximate entropy profile: a novel approach to comprehend irregularity of short-term HRV signal, *Nonlinear Dynamics*, 88(2), 2017, 823-837.
[7] Shang, D., Shang, P., Zhang, Z., Efficient synchronization estimation for complex time series using refined cross-sample entropy measure, *Communications in Nonlinear Science and Numerical Simulation*, 94, 2021, 105556.
[8] Mao, X., Shang, P., A new method for tolerance estimation of multivariate multiscale sample entropy and its application for short-term time series, *Nonlinear Dynamics*, 94(3), 2018, 1739-1752.
[9] Freedman, D., Diaconis, P., On the histogram as a density estimator: L2 theory, *Probability Theory and Related Fields*, 57(4), 1981, 453-476.
[10] Contreras-Reyes, J.E., Mutual information matrix based on asymmetric Shannon entropy for nonlinear interactions of time series, *Nonlinear Dynamics*, 104(4), 2021, 3913-3924.
[11] Richman, J.S., Moorman, J.R., Physiological time-series analysis using approximate entropy and sample entropy, *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 278(6), 2000, H2039-H2049.
[12] R Core Team, *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020, Available at http://www.R-project.org.
[13] Arnold, L., *Random dynamical system*, Springer-Varlag, Berlin, Germany, 1998.
[14] Dlamini, P., Simelane, S., An Efficient Spectral Method-based Algorithm for Solving a High-dimensional Chaotic Lorenz System, *Journal of Applied and Computational Mechanics*, 7(1), 2021, 225-234.
[15] Contreras-Reyes, J.E., Chaotic systems with asymmetric and heavy-tailed noise: application to 3D attractors, *Chaos, Solitons & Fractals*, 2021, 145, 110820.
[16] Liu, L.Z., Qian, X.Y., Lu, H.Y., Cross-sample entropy of foreign exchange time series, *Physica A: Statistical Mechanics and its Applications*, 389, 2010, 4785-4792.
[17] Shi, W., Shang, P., Cross-sample entropy statistic as a measure of synchronism and cross-correlation of stock markets, *Nonlinear Dynamics*, 71, 2013, 539-554.
[18] Xia, J., Shang, P., Multiscale entropy analysis of financial time series, *Fluctuation and Noise Letters*, 11(4), 2012, 1250033.
[19] Argyris, J., Andreadis, I., Pavlos, G., Athanasiou, M., The influence of noise on the correlation dimension of chaotic attractors, *Chaos, Solitons & Fractals*, 9, 1998, 343-361.
[20] Abid, S.H., Hasan, H.M., About asymmetric noisy chaotic maps, *International Journal of Basic and Applied Sciences*, 3(2), 2014, 62-73.
[21] Contreras-Reyes, J.E., Fisher information and uncertainty principle for skew-gaussian random variables, *Fluctuation and Noise Letters*, 20, 2021, 21500395.
[22] Li, B., Han, G., Jiang, S., Yu, Z., Composite Multiscale Partial Cross-Sample Entropy Analysis for Quantifying Intrinsic Similarity of Two Time Series Affected by Common External Factors, *Entropy*, 22(9), 2020, 1003.
[23] Contreras-Reyes, J.E., Canales, T.M., Rojas, P.M., Influence of climate variability on anchovy reproductive timing off northern Chile, *Journal of Marine Systems*, 164, 2016, 67-75.
[24] Contreras-Reyes, J.E., Hernández-Santoro, C., Assessing Granger-causality in the southern Humboldt current ecosystem using cross-spectral methods, *Entropy*, 22(10), 2020, 1071.

ORCID iD

Javier E. Contreras-Reyes https://orcid.org/0000-0003-1172-5456

How to cite this article: Contreras-Reyes J.E., Brito A. Refined Cross-sample Entropy based on Freedman-Diaconis Rule: Application to Foreign Exchange Time Series, *J. Appl. Comput. Mech.*, 8(3), 2022, 1005–1013. https://doi.org/10.22055/jacm.2022.39470.3412