Similarity Analysis of JS Divergence Based on Permutation Entropy

0. Abstract

We proposed a new JS (Jenson Shannon) divergence method based on permutation entropy;

Generated different 2D ARFIMA time series for a given coupling strength; using them to test JSD, fractional JSD and generalized JSD methods based on permutation entropy generalization, and generalized fractional JSD method to identify the similarity of two artificial time series;

Established a suitable moving window length, traversing the entire time series, and applied it to the new method, using Python to plot the similarity curve between the simulated data and the real stock data (seven different regional stock KLCI records); plotted genetic tree, and heat effect map by R to visualize the comparison results, and used financial market basic knowledge to verify the accuracy of the results.

1. Introduction

In recent years, the pervasion of information theory and communications technology has sparked enthusiasm towards complex dynamic systems [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. Researchers are increasingly focusing on analyzing the complex system time series, especially for high-frequency financial data. These financial markets are prominent well-defined dynamic complex systems which contain an enormous amount of interacting data points confirming the under- lying economic trends. Therefore, find the adequate measurements or relevant variables to describe their properties are the crucial concern [12]. The majority of empirical literature seems to have interests in stock markets. They applied the statistical mechanics to economic systems in order to investigate financial time series as a reflection of economic trends. Typical time series analysis deal with projects such as clustering similarity measure and prediction. Moreover, one of the specific application of the properties in dynamic complex systems is measure the similarity between different financial time series. Machine learning[13], data mining [14] and clustering analyses [15, 16] are some well-known examples of application domains that all require an explicit method to evaluate and quantify the relationship and similarity among each time series.

Up to now, plenty of methods have been introduced to quantify the similarity in time series. Liu (2010) proposed Cross-ApEn and Cross-SampEn to measure the similarity of two distinct time series[17, 18], and then, in order to analyze the similarity of two series under different timescales, multiscale cross-sample entropy [19] was put forward based on multiscale entropy and the Cross-SampEn. For further investigation, the irreversibility property of multi-scale time [20] was also offered to classify the financial markets. In addition, lots of other methods were introduced to evaluate and quantify the similarity of different stock markets, such as multiscale detrended fluctuation analysis (MSDFA), detrended Cross-Correlation Analysis [21], Time-Varying Copula-GARCH Model [22], and multiscale detrended cross-correlation analysis (MSD- CCA) [23].

Moreover, it has been reported that it is possible to detect the similarity of dynamic structure in the stock markets by using the Kullback-Leibler divergence (KLD) [24]. In statistics analysis and information theory, the KLD is commonly used as a measure of similarity between two probability density distributions [25, 26, 27, 28]. This divergence measurement is primary from Shannon's concept of entropy, which refers to disorder or probabilistic uncertainty [29, 30]. This paper uses Jensen-Shannon divergence (JSD), which is a symmetric version of KLD. The JSD has several practical properties [31]. It arises in information theory and, unlike the KLD, it is symmetric and bounded between zero and one. The KLD and JSD have been applied in many fields of character and image recognition, such as evaluating the similarity between two acoustic models [32, 29, 33, 34], measuring how confusable two articles are [35, 36, 37], determining the best match using histogram image models [38], clustering of models, and model optimization by minimizing or maximizing the KLD between density distributions. Other than the applications in clustering or model optimization, Sato [39] utilized the KLD and JSD between two normalized spectrograms of the tick frequency in the foreign exchange market so that to quantify the similarities among multi-dimensional time series.

Here in this paper, JSD is generalized by permutation entropy (PE) to calculate the similarity between several stock market time series. Entropy is a stable and robust nonlinear measurement used to characterize the complexity of a time series. Among several generalized version of entropy, PE has been recently suggested as the simplest complexity measure which is easily and extremely fast calculation for any type of time series based on the permutation sequences [40, 41, 42, 43, 44, 45, 46]. In this case, PE can be utilized to generalize JSD to get more accurate and robust results. Moreover, the JSD measure can be transformed and rewritten on the basis of the fractional approach [47]. This innovative perspective for fractional entropy is inspired by the properties of Fractional Calculus [48, 49, 50, 51, 52, 53, 54]. In this expression, fractional operators can relax some strict properties and allowing their application in complex dynamical systems [55, 56, 57]. In addition, Agrawal and Rafiei et al [58, 59, 60] have introduced a more intuitive idea, which is two series should be considered similar if they have enough non-overlapping time-ordered pairs of subsequences that are similar. Based on this point of view, we can based on the results of analyzing the similarity of smooth sliding windows over time series to infer the similarity of the entire sequence, or specifically, the similarity between two financial markets. In this paper, We empirically evaluate the similarity performance (effectiveness) based on JSD, Fractional JSD and generalized PE JSD over moving windows algorithm. Moreover, the distance plots and phylogenetic trees according to the dissimilarity results are used to give direct information about different stock markets in order to evaluate the similarity among them.

The remainder of this paper is organized as follows. In Section 2, we describe the methodology of JSD, Fractional JSD, as well as the generalized PE JSD and Fractional PE JSD, as well as the moving window algorithm. Then, Section 3 presents the simulation and empirical data used in this work, including the Autoregressive Fractionally Integrated Moving Average (ARFIMA) simulated series and various stock time series of different areas during the period 2000.1.3-2018.3.28. Section 4 demonstrates the similarity comparison results under moving window algorithm with respect to JSD, Fractional JSD, generalized JSD and generalized Fractional JSD. Moreover, the distance plots and phylogenetic trees based on these three methods are also illustrated in this section. Finally, conclusion is demonstrated in Section 5.

2. Methodology

2.1 Kullback-Leibler Divergence and Jensen-Shannon Divergence

Relative entropy, or Kullback-Leibler divergence, is an asymmetry measure of the difference between two probability distributions. In information theory, the relative entropy is equivalent to the difference of the information entropy of two probability distributions (Shannon entropy). The distance between two random distributions can be measured. When two random distributions are the same, their relative entropy is zero. When the difference between two random distributions increases, their relative entropy also increases. So relative entropy can be used to compare the similarity of text, first calculate the frequency of the word, and then calculate the relative entropy. In addition, in multi-index system evaluation, index weight distribution is a key and difficult point, and can also be handled by relative entropy.

The Kullback-Leibler divergence between the two probability distributions P_1 and P_2 is given by the following equation:

$$D_{KL}(P\|Q) = -\sum_i p_i \ln \frac{p_i}{q_i} \qquad (1)$$

The Jensen–Shannon divergence (D_{JS}) uses a symmetric version of the $D_{KL}$and measures the similarity between two probability distributions. The D_{JS} is given by[14,31]:

$$D_{JS}(P\|Q) = \frac{1}{2}[D_{KL}(P\|M) + D_{KL}(Q\|M)] \qquad (2)$$

Where $M = \frac{1}{2}(P+Q)$is the mixture distribution of P and Q.

The D_{JS} can be successfully applied in complex systems and, in particular, to DNA analysis [32,42].

Based on the D_{JS} index, a symmetric matrix $M_{JS}$of item-to-item distances and the corresponding tree are constructed.

The algorithm of D_{JS} under the sliding window can be expressed as follows:

Step1:

Given a set of time series, $X = (x_1, x_2, \ldots, x_N)$, the window value of the sliding window is l, each window size is n. Then we can divide the time series into K groups. K is the largest integer that satisfies $((k-1)l + n) \leq N$.

The k^{th} group is expressed as $X_k = \{x_{(k-1)l+1}, x_{(k-1)l+2}, \ldots, x_{(k-1)l+n}\}$.

Step2:

For each set of time series that are segmented, we evenly divided the components into h parts according to their size range. Then we calculate the probability of two-dimensional data by calculating the number of points in each region separately.

Step3:

Calculate the value of each set of data that is split by the sliding window into the formula D_{JS}:

$$D_{JS}(P\|Q) = \frac{1}{2}[D_{KL}(P\|M) + D_{KL}(Q\|M)]$$

Where $M = \frac{1}{2}(P + Q)$ is a mixed distribution of P and Q.

Step4:

We take the sliding window parameter as the point on the x-axis, and the corresponding value D_{JS} of each window as the point on the y-axis. Based on this, we can draw an image of sliding window in D_{JS} values.

In the same way, we can also draw the moving window curve of the fractional JSD, generalized fractional JSD.

2.2 Fractional Jensen-Shannon Divergence

2.2.1 Information Entropy and Fractional JSD

Information theory was developed by Claude Shannon in 1948 [26,27] and has been applied in many scientific areas. The fundamental cornerstone is the information content of some event having probability of occurrence p_i:

$$I(p_i) = -\ln p_i \quad (3)$$

The expected value, called Shannon entropy [28,29], becomes:

$$S = E(-\ln p) = \sum_i (-\ln p_i) p_i \qquad (4)$$

where $E(\cdot)$denotes the expected value operator.
Expression (2) obeys the four Khinchin axioms [30,31] and several generalizations of entropy have been proposed, obeying only a sub-set of them.
Recently Ubriaco brought together information theory and FC and proposed [32] the expression:

$$S_q = E[(-\ln p)^q] = \sum_i (-\ln p_i)^q p_i \qquad (5)$$

where $0 \le q \le 1$ denotes the "order" so that $q = 1$ yields Expression (4). This formulation obeys the same properties as the Shannon entropy except additivity and is the expected value of information content given by:

$$I_q(p_i) = (-\ln p_i)^q \qquad (6)$$

It is well known in FC the adoption of a power function for obtaining intermediate values, that is, for "fractionating" classical integer operators. In brief, the Laplace transform of the fractional derivative of order $\alpha \in R$ of a signal $x(t)$with zero initial conditions is given by:

$$\mathscr{L}\{{}_0D^{\alpha} x(t)\} = s^{\alpha} \mathscr{L}\{x(t)\} \qquad (7)$$

where t represents time, and $L\{\cdot\}$ and s denote the Laplace operator and variable, respectively.

This property motivated the approximation of fractional derivatives by expanding the factor s^{α} both with the Fourier and the Z transforms [33,34]. However, the adoption by means of a power function is related with transforms and we can design a distinct fractional approach for information and entropy. In fact, we can give the definition of Shannon information in the perspective of FC, leads to the proposal of information and entropy p_i of order $\alpha \in$ R given by [35]:

$$I_{\alpha}(p_i) = D^{\alpha}(p_i) = -\frac{p_i^{-\alpha}}{\Gamma(\alpha+1)}[\ln p_i + \psi(1) - \psi(1-\alpha)] \qquad (8)$$

$$S_{\alpha} = -\sum_i p_i \left\{ \frac{p_i^{-\alpha}}{\Gamma(\alpha+1)}[\ln p_i + \psi(1) - \psi(1-\alpha)] \right\} \qquad (9)$$

where $\Gamma(\cdot)$ and $\psi(\cdot)$ represent the gamma and digamma functions.
Expression (9) fails to obey some of the Khinchin axioms with exception of the case $\alpha = 0$ that leads to the classical Shannon entropy. This behaviour is in line with what occurs in FC, where fractional derivatives fail to obey some of the properties of integer-order operators. By other words, in both cases, by generalizing operators we loose some classical properties.

The results reveal that tuning the fractional order allow a high sensitivity to the signal evolution, which is useful in describing the dynamics of complex systems. The concepts are also extended to relative distances and tested with several sets of data, confirming the goodness of the generalization.

For convenience, we will denote the fractional JSD as D_{JS}^{α}.

Based on the definition of D_{JS} in 2.1, we extend it to discrete forms:

$$D_{JS}(P\|Q) = -\frac{1}{2}[\sum_i p_i \ln p_i + \sum_i q_i \ln q_i] + \sum_i m_i \ln m_i \quad (10)$$

According to (6), (8), (10), we can further write discrete expressions of D_{JS}:

$$D_{JS}^{q}(P\|Q) = -\frac{1}{2}\sum_i p_i(-\ln p_i)^q - \frac{1}{2}\sum_i q_i(-\ln q_i)^q + \sum_i m_i(-\ln m_i)^q \quad (11)$$

$$\begin{aligned} D_{JS}^{\alpha}(P\|Q) = & \frac{1}{2}\sum_i p_i \left\{ \frac{p_i^{-\alpha}}{\Gamma(\alpha+1)}[\ln p_i + \psi(1) - \psi(1-\alpha)] \right\} \\ & + \frac{1}{2}\sum_i q_i \left\{ \frac{q_i^{-\alpha}}{\Gamma(\alpha+1)}[\ln q_i + \psi(1) - \psi(1-\alpha)] \right\} \quad (12) \\ & - \sum_i m_i \left\{ \frac{m_i^{-\alpha}}{\Gamma(\alpha+1)}[\ln m_i + \psi(1) - \psi(1-\alpha)] \right\} \end{aligned}$$

In both cases, a comparison $n \times n$ symmetrical matrix D of element to element relative distances is constructed, adopting the indices D_{JS}^{q} and D_{JS}^{α}.

These algorithms produces a tree based on matrix D, trying to accommodate the distances into the two dimensional space.

The algorithm of D_{JS}^{α} under the sliding window can be expressed as follows:

Step1:
Given a set of time series, $X = (x_1, x_2, \ldots, x_N)$, the window value of the sliding window is l, each window size is n. Then we can divide the time series into K groups. K is the largest integer that satisfies $((k-1)l + n) \leq N$.

The k^{th} group is expressed as $X_k = \{x_{(k-1)l+1}, x_{(k-1)l+2}, \ldots, x_{(k-1)l+n}\}$.

Step2:
For each set of time series that are segmented, we evenly divided the components into h parts according to their size range. Then we calculate the probability of two-dimensional data by calculating the number of points in each region separately.

Step3:
Calculate the value of each set of data that is split by the sliding window into the formula D_{JS}^{α}:

$$
\begin{aligned}
D_{JS}^{\alpha}(P\|Q) = &\frac{1}{2}\sum_i p_i \left\{ \frac{p_i^{-\alpha}}{\Gamma(\alpha+1)}[\ln p_i + \psi(1) - \psi(1-\alpha)] \right\} \\
&+\frac{1}{2}\sum_i q_i \left\{ \frac{q_i^{-\alpha}}{\Gamma(\alpha+1)}[\ln q_i + \psi(1) - \psi(1-\alpha)] \right\} \\
&-\sum_i m_i \left\{ \frac{m_i^{-\alpha}}{\Gamma(\alpha+1)}[\ln m_i + \psi(1) - \psi(1-\alpha)] \right\}
\end{aligned}
$$

Where $M = \frac{1}{2}(P+Q)$ is a mixed distribution of P and Q.

Step4:
We take the sliding window parameter as the point on the x-axis, and the corresponding value D_{JS}^{α} of each window as the point on the y-axis. Based on this, we can draw an image of sliding window in D_{JS}^{α} values.

In the same way, we can also draw the moving window curve of the fractional JSD, generalized fractional JSD.

2.3 Generalized JSD and Fractional JSD based on entropy

2.3.1 Permutation entropy

We introduce complexity parameters for time series based on comparison of neighboring values. The definition directly applies to arbitrary real-world data. For some well-known chaotic dynamical systems it is shown that our complexity behaves similar to Lyapunov exponents, and is particularly useful in the presence of dynamical or observational noise. The advantages of our method are its simplicity, extremely fast calculation, robustness, and invariance with respect to nonlinear monotonous transformations.

Our entropies are calculated for different embedding dimensions n, but we do not attempt to determine a limit for large n although this is an interesting theoretical problem [16]. For practical purposes, we recommend $n = 3,...,7$.

Consider a time series $\{x_t\}_{t=1,...,T}$. We study all $n!$ permutations π of order n which are considered here as possible order types of n different numbers. For each p we determine the relative frequency (# means number) :

$$p(\pi) = \frac{\#\{t \,|\, t \leq T - n, (x_{t+1}, \ldots, x_{t+n}) \text{ has type } \pi\}}{T - n + 1} \qquad (13)$$

So *The permutation entropy* of order $n \geq 2$ is defined as :

$$H(n) = -\sum p(\pi) \ln p(\pi) \qquad (14)$$

2.3.2 Generalized JSD and Fractional JSD under moving window

We use the definition of the probability density function p_i in permutation entropy for a set of time series and use it in the calculation of JSD and fractional JSD. The calculation steps are as follows:

The algorithm of D_{JS}^{α} under the sliding window can be expressed as follows:

Step1:

Given two sets of time series, $X = (x_1, x_2, \ldots, x_N)$, $Y = (y_1, y_2, \ldots, y_N)$, the window value of the sliding window is l, each window size is n. Then we can divide the time series into K groups. K is the largest integer that satisfies $((k-1)l + n) \leq N$.

The k^{th} group of series X, Y is expressed as $X_k = \{x_{(k-1)l+1}, x_{(k-1)l+2}, \ldots, x_{(k-1)l+n}\}$, $Y_k = \{y_{(k-1)l+1}, y_{(k-1)l+2}, \ldots, y_{(k-1)l+n}\}$ respectively.

Step2:

For each set of time series that are segmented, we evenly divided the components into h parts according to their size range. Then we calculate the probability of two-dimensional data by calculating the number of points in each region separately.

We choose the edge density function of the two-dimensional probability X or Y and record it as a new sequence Z. we calculate the probability density function of the time series Z based on the definition in permutation entropy.

Step3:

Calculate the value of each set of data that is split by the sliding window by the following formulas:

$$D_{JS}(P\|Q) = \frac{1}{2}[D_{KL}(P\|M) + D_{KL}(Q\|M)]$$

$$\begin{aligned} D_{JS}^{\alpha}(P\|Q) = &\frac{1}{2}\sum_i p_i \left\{ \frac{p_i^{-\alpha}}{\Gamma(\alpha+1)}[\ln p_i + \psi(1) - \psi(1-\alpha)] \right\} \\ &+\frac{1}{2}\sum_i q_i \left\{ \frac{q_i^{-\alpha}}{\Gamma(\alpha+1)}[\ln q_i + \psi(1) - \psi(1-\alpha)] \right\} \\ &-\sum_i m_i \left\{ \frac{m_i^{-\alpha}}{\Gamma(\alpha+1)}[\ln m_i + \psi(1) - \psi(1-\alpha)] \right\} \end{aligned}$$

Where $M = \frac{1}{2}(P+Q)$ is a mixed distribution of P and Q.

Step4:

We take the sliding window parameter as the point on the x-axis, and the corresponding value $PE - D_{JS}$ or $PE - D_{JS}^{\alpha}$ of each window as the point on the y-axis. Based on this, we can draw an image of sliding window in $PE - D_{JS}$ or $PE - D_{JS}^{\alpha}$ values.

3. Data

3.1 The generation of Simulated data

According to the Autoregressive Fractionally Integrated Moving Average(ARFIMA) models [50], we construct several series of length N. As we know, the ARFIMA models can model the cross-correlation between two ARFIMA series for any given strength of coupling between them, according to the equations below:

$$\begin{aligned} x_t &= [WX_t + (1-W)Y_t] + \varepsilon_t \\ y_t &= [(1-W)X_t + WY_t] + \varepsilon_t^{'} \\ X_t &= \sum_{n=1}^{\infty} a_n(d_1)x_{t-n} \\ Y_t &= \sum_{n=1}^{\infty} a_n(d_2)y_{t-n} \end{aligned} \qquad (15)$$

$a_n(d) = d\Gamma(n-d)/(\Gamma(1-d)\Gamma(n+1))$

where ε_t and $\varepsilon_t{}'$ are two different i.i.d. Gaussian variables with zero mean and variance = 1. $\Gamma(x)$ is the Gamma function, and the scaling parameters d_1 and d_2 vary with the range of $(-0.5,\ 0.5)$. The parameter W denotes the strength of the coupling and varies from 0.5 to 1, where $W = 0.5$ gives the highest cross- correlation, while $W = 1$ represents the total absence of correlation.

In our simulation, we analyze the similarities between a set of different time series generated by the ARFIMA model and build a phylogenetic tree for these sequences.

3.2 Stock Data

For further discussion on the performance of JSD based on PE entropy, we consider the time series in the stock market. The financial time series obtained from various stock markets consist of the daily adjusted closed price of seven stock indices which are listed in Table 1 during the period 2000.1.3-2018.3.28. These data are attained from Yahoo finance.com. Although the amount of data in all the time series are all above 5000, the lengths of these seven stock time series are all different from each other because they belong to different areas and have different opening dates. Thus, we exclude the asynchronous date and then relink the rest so that we can get all the time series in the same length. After that, in the next section, we will apply the generalized JSD and Fractional JSD based on these seven stock market time series and discuss its effectiveness and practical applications.

Table 1 The list of seven stock indices

Area	Country	Index	Definition
Asia	China	SSE	The SSE Index is a stock market index of all stocks that are traded at the Shanghai Stock Exchange.
	China	HSI	The Hang Seng Index (HSI) is stock-market index of all stocks that are traded at Hong Kong.
	Japan	N225	The Nikkei 225(N225) is a stock market index for the Tokyo Stock Exchange.
America	United State	DJIA	The Dow Jones Industrial Average (DJIA) is a stock market index that indicates the value of 30 large, publicly owned companies based in the United States.
	United State	SPX	The S&P 500 is an American stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE, NASDAQ, or the Cboe BZX Exchange.

Area	Country	Index	Definition
	United State	NDX	The NASDAQ-100 is a stock market index made up of 103 equity securities issued by 100 of the largest non-financial companies listed on the NASDAQ.
	United State	RUT	The Russell 2000 index is an index measuring the performance of approximately 2,000 small-cap companies of the biggest U.S. stocks.

4. Results analysis

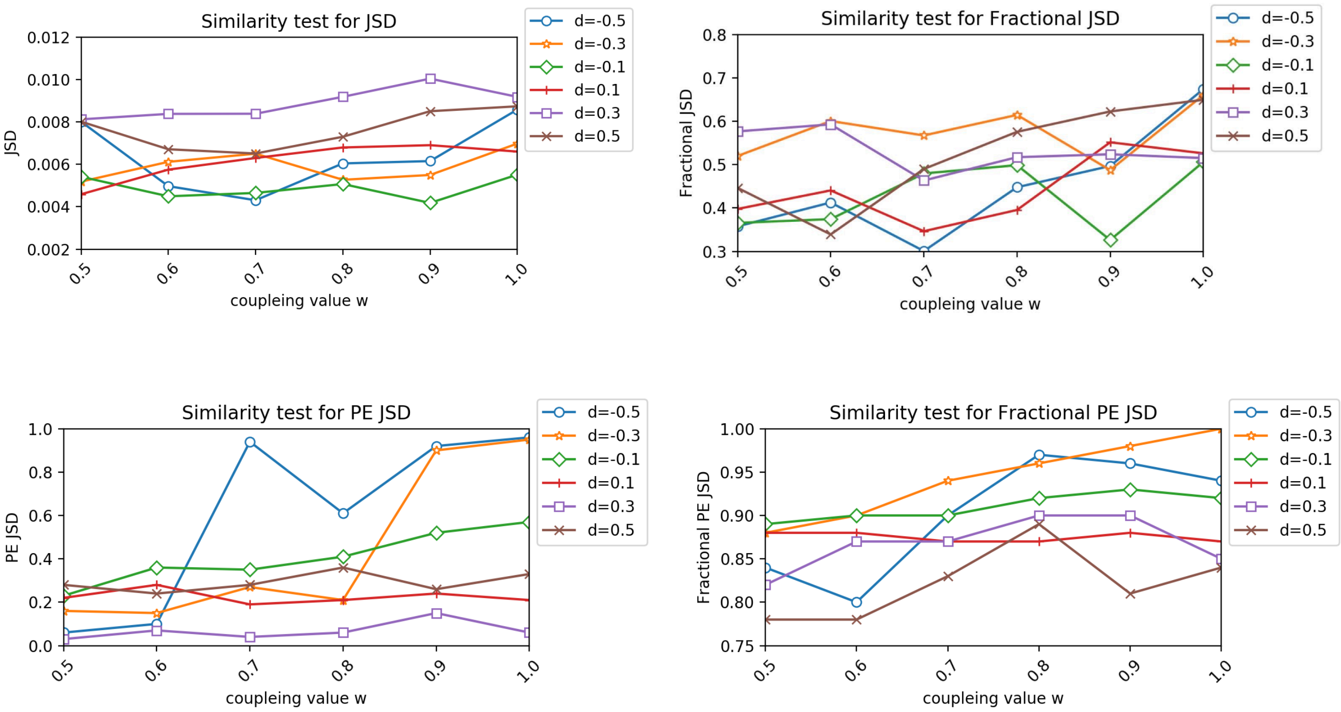
In this part, we performed experiments to analyze the ability of original JSD, Fractional JSD, generalized JSD, and the Fractional generalized JSD to distinguish between different simulated ARFIMA time series, and also for financial time series in the different stock market. Firstly, by controlling the two parameters for the ARFIMA series, we used these four JSD approaches to prove the parameter optimization of similarity based on ARFIMA simulation. Then, we compared the significant performances of similarity coefficient calculated by the original JSD and generalized JSD using the moving window method based on different window-length. Then experiments were conducted on the real datasets to analyze the properties and similarity among different stock market using the JSD methods.

4.1 Artificial data

4.1.1 similarity test of ARFIMA sequence based on different parameters

To better clarify the dependence of distances to both the strength of coupling w and the scaling parameter d , we show how distance varies for different pairs of these two parameters. Specifically, we quantify the similarity distances by using the four JSD methods above with different pairs of parameters to analyze how will the results will depend on the parameters. We confirm the performance and accuracy of these approaches by the cross-correlation properties mentioned in previous research.

In our study, we generate plenty sets of ARFIMA series with N = 5000 data using thirty-six different values of the parameters (d , W) = (-0.5, -0.3, -0.1, 0.1, 0.3, 0.5)x(0.5, 0.6, 0.7, 0.8, 0.9, 1.0). Then, we construct two series for each pair of parameters to compare the similarity between them. Fig. 1 contains the curves of similarity indices obtained by the four JSD methods with changing the scaling parameter d from -0.5 to 0.5, with step 0.2, and the coupling values W from 0.5 to 1, with step 0.1. Previous research proved that when holding d unchanged, the cross-correlation between two simulated ARFIMA sequences is highest when w= 0.5, and it would decrease as w approached 1. In this case, in order to analyze the performance of these four measurements, we expected to observe the upward-trending from these figure. In other words, the similarity is getting lower as W increases which is a confirmation of the cross-correlation properties. We can recognize that the results for PE fractional JSD is the best comparing to other three methods.



4.1.2 Similarity results of artificial data in moving windows

According to the above comparison results of the line chart, it reflects that the similarity of sequences is higher for the case of W = 0.5, which corresponds to the definition of AFRIMA models. Therefore, we generate two sets of simulation sequences based on the same parameters (w = 0.5, d = 0.5), each sequence containing 5000 numbers. In order to analyze the similarity results of JSD, Fractional JSD, PE JSD and Fractional PE JSD, we propose to compare signal similarity under the sliding window algorithm. By changing the length of the window, that is, the amount of data for each similarity calculation, we compare the volatility of the JSD and PE JSD similarity algorithms horizontally. Figure 2 shows the similarity index fluctuation curve of JSD at window lengths of 200 and 50, and Figure 3 shows the fluctuation curve generated by PE JSD. The results show that the larger the window, the closer the similarity index is to zero and the smaller the fluctuation; the smaller the window, the larger the difference in the simulation data, and the greater the fluctuation of the similarity index. So we conclude that our results are better when the data set applied by the method is larger.

Figure 2

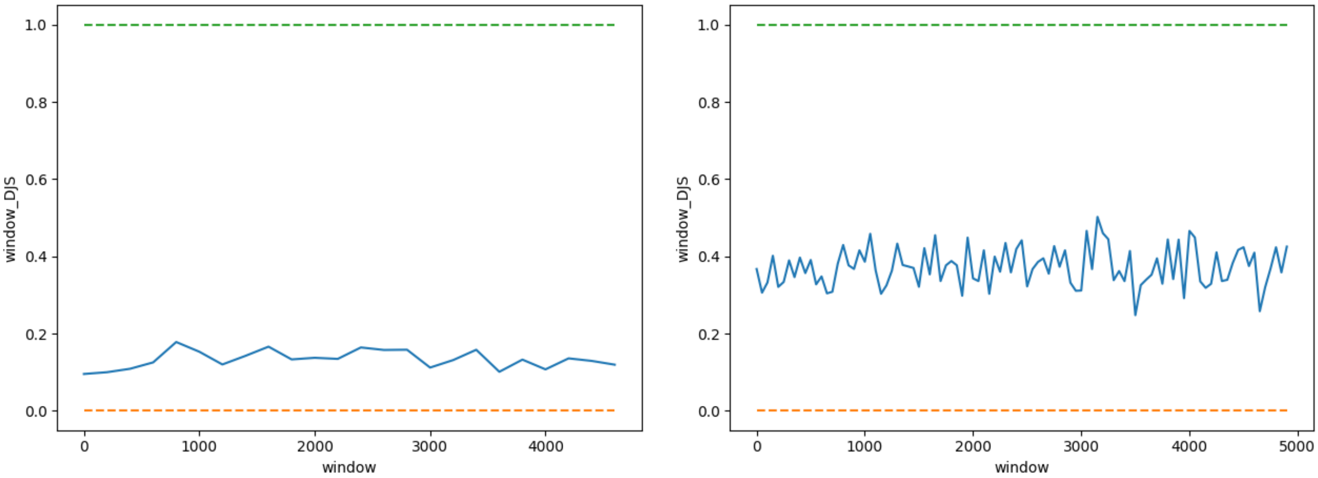


Figure 2 (a): length=200

Figure 2 (b): length=50

Figure 3

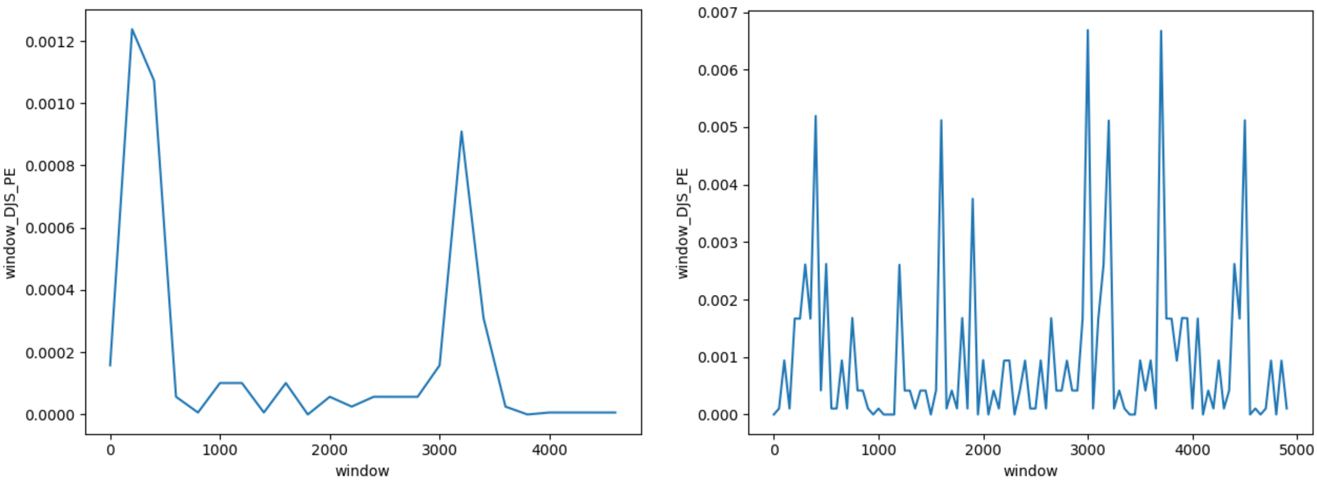


Figure 3 (a): length=200

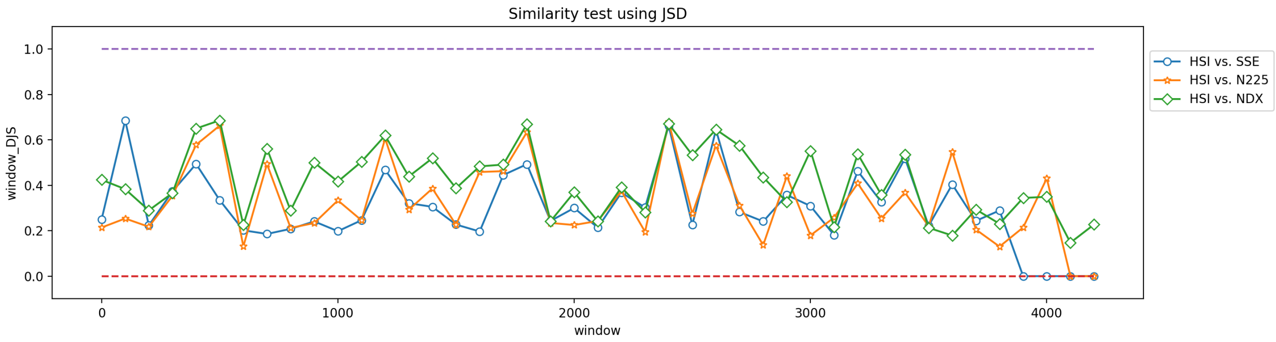
Figure 3 (b): length=50

4.2 Stock Data

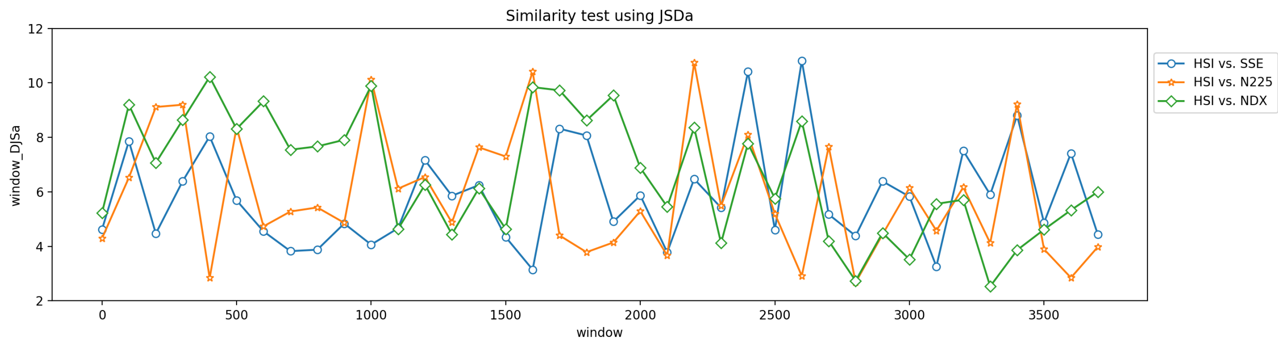
4.2.1 Stock data under living windows

We conducted research on HSI and a Chinese stock index (SSE), a Japanese stock index (Nikkei 225) and a US stock index (NASDAQ100) respectively, and found that under the sliding window, different methods can reflect the characteristics of the stock market differently.

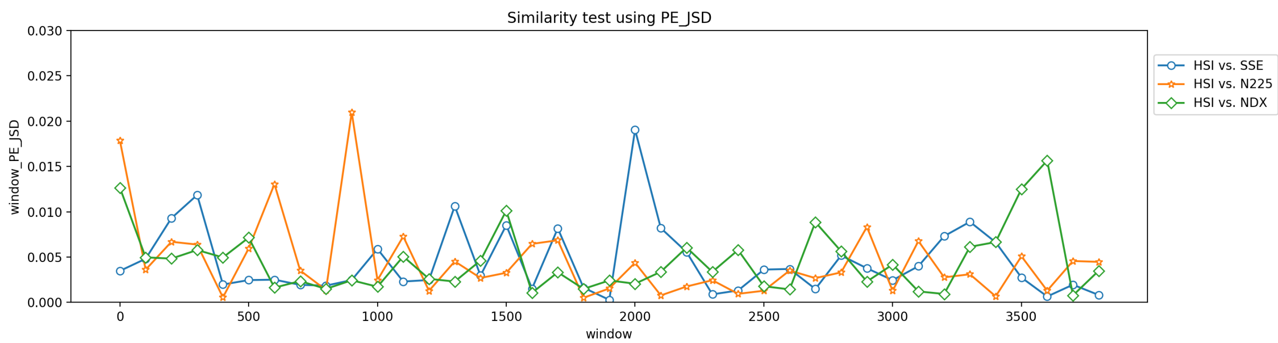
We found that under the sliding window of JSD, the values corresponding to HSI and SSE are basically at a minimum among the three curves. At the same time, from the time point of view, as time goes by, the values of the three curves all show a certain downward trend.



This phenomenon is consistent with the trend of the great integration of the world economy in recent years.



Under the sliding window of the Fractional JSD, the three curves are more volatile. At this time, the values corresponding to HSI and SSE are no longer substantially at the minimum in the three curves. Compared with the sliding window of JSD, the curve trend of the two Chinese stock markets has not changed much. Compared with the Japanese and American stock markets, the trend of the curve has changed greatly.



In the JSD method generalized by permutation entropy, the value of the curve presented by the sliding window is smaller. Moreover, the similarity curves of HSI and Nikkei225 are smaller than the other two. Moreover, the curve trend of the two Chinese stock markets is not much different from that measured by the JSD method, and the trend of the curve is well preserved compared with the non-Chinese market.

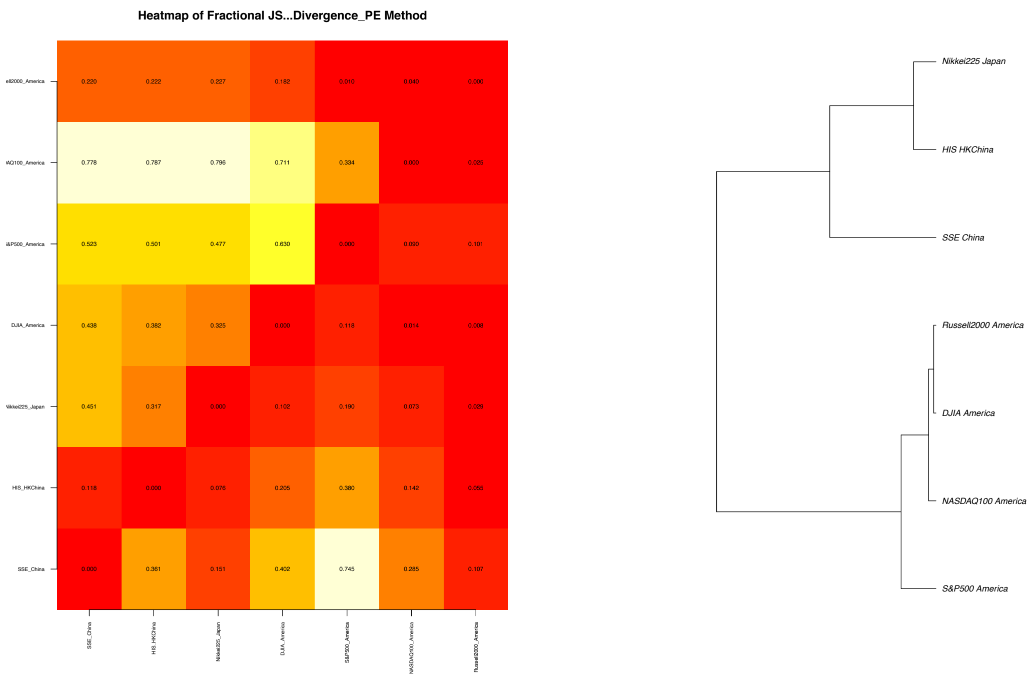
Based on the above, it can be found that under the sliding window of the PE JSD method we promoted, the curve trend corresponding to the same stock market is well preserved and more stable than the fractional JSD method.

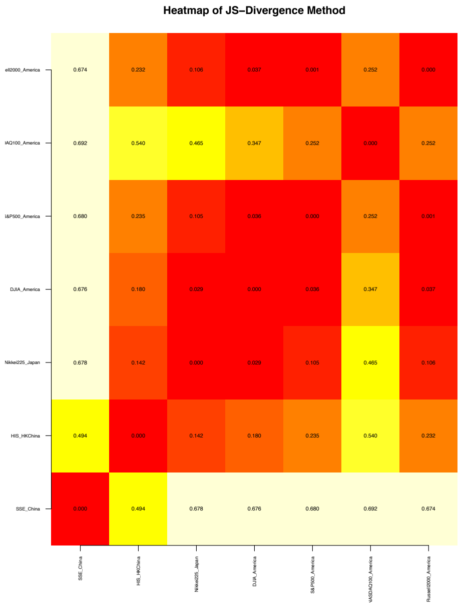
4.2.2 the similarity analysis in financial market based on heatmap and genetic trees

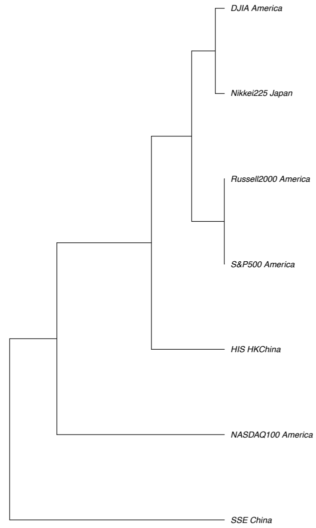
The financial time series is a dynamic system consisting of high complexity data, so it is impossible and inaccurate to evaluate the similarity of them by using the simple Euclidean distance, especially when we intend to analyze several different stock markets together. Thus, these complexity properties require to use these four JSD methods since it based on the information theory and can illustrate the characteristic in the complex system better. In this paper, We investigate seven stock time series in three different areas from January 3, 2000, to March 28, 2018. Different stock time series in different areas show different characteristics among them, and there may be related to others. We calculate the distances between the selected time series in order to confirm these correlations. Fig. 4,5,6,7 show the distance plots and phylogenetic trees of these seven stock indices using JSD， Fractional JSD, PE JSD, and Fractional PE JSD separately.

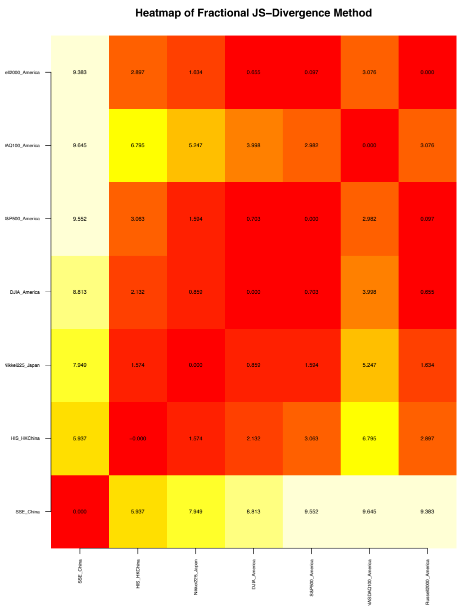
Figure. 7 shows that the modified Fractional PE JSD is better to analyze the similarity between these stock time series. It accurately reflects the high similarity of the time series in America stock markets, also in Asia stock markets. Moreover, the Chinese and Japanese markets is more similar than that with American markets. Besides, the color of distance plots also reflects the same result as the phylogenetic trees. The generalized PE JSD also demonstrates a reasonable classify in Figure 6. It depends on how we classify the Japanese stock market. If we consider that the Japan stock markets are influenced heavier by America markets than by Chinese markets, so it locates closer to US markets which is away from Chinese markets. However, other methods are not showing good classification results. Hence, we conclude that the modified methods for JSD have better performances.

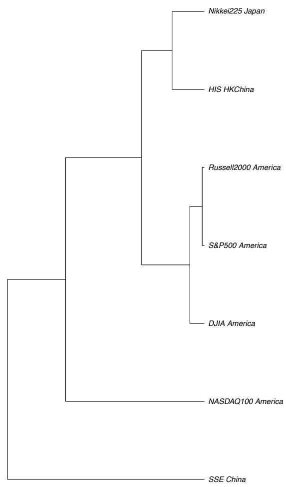
Fig. 4,5,6,7

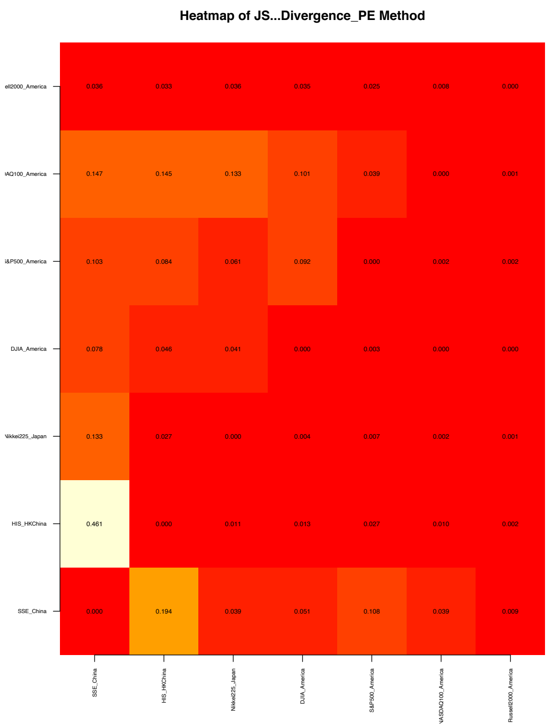


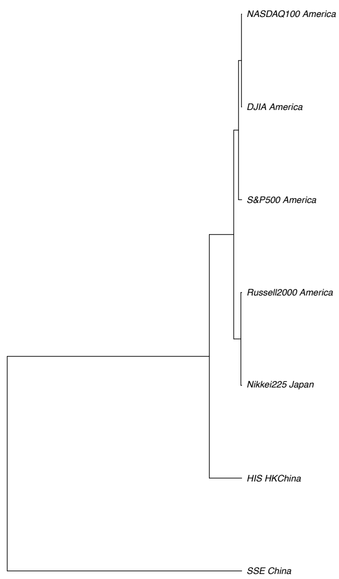

Heatmap of JS−Divergence Method


DJIA America
Nikkei225 Japan
Russell2000 America
S&P500 America
HIS HKChina
NASDAQ100 America
SSE China


Heatmap of Fractional JS−Divergence Method


Nikkei225 Japan
HIS HKChina
Russell2000 America
S&P500 America
DJIA America
NASDAQ100 America
SSE China


Heatmap of JS...Divergence_PE Method


NASDAQ100 America
DJIA America
S&P500 America
Russell2000 America
Nikkei225 Japan
HIS HKChina
SSE China

5. Conclusion

In our paper, we generalize the JS divergence and the fractional JS divergence based on the definition of permutation entropy. By doing so, We obtain a definition that retain the basic mathematical properties of JS divergence and superimposing the advantages of permutation entropy.

We have made several conclusions in the testing of simulated data for different parameters. First, from the absolute size of the numerical values, the recognition effect of the generalized JSD and the generalized fractional JSD on the data generated based on the ARFIMA algorithm is better than that of the JSD and the fractional JSD. At the same time, we find that the data similarity value corresponding to the fractional JSD deviates a lot from zero. At the same time, by comparing the rising trend of the similarity of the simulated data generated by different parameters, we believe that the generalized JS divergence and fractional JS divergence methods that based on the permutation entropy can well identify the similarity of different simulation data.

At the same time, in the application of analog data, we found that under the sliding window, the larger the window, the closer the similarity index is to zero and the smaller the fluctuation; the smaller the window, the larger the difference of the simulation data, and the greater the fluctuation of the similarity index.

Finally, we use the moving window to test the similarity of the data in the two stock markets. We found that under the sliding window of the PE JSD method we promoted, the curve trend corresponding to the same stock market is well preserved and more stable than the fractional JSD method. Moreover, by observing the genetic tree and heat map, we find that the similarity testing effect of JSD and fractional JSD on the stock market is not as good as that of generalized JSD and generalized fractional JSD based on permutation entropy. From this we can see that the method we promoted has a better recognition of the stock market in different countries, which once again verifies the classification effect of the method we promote on financial time series data.

In summary, our stock simulation data based on permutation entropy has a good recognition of the similarity between simulated data and stock market data. After that, we hope to further study the application of our divergence method based on permutation entropy in kernel principal component analysis.

References

- [1] A.-H. Sato, Frequency analysis of tick quotes on the foreign exchange mar- ket and agent-based modeling: A spectral distance approach, Physica A: Statistical Mechanics and its Applications 382 (1) (2007) 258–270.
- [2] R. Mantegna, H. Stanley, An introduction to econophysics: Correlations and complexity in finance, 1999.

- [3] M. Dacorogna, R. Genc¸ay, U. Muller, R. Olsen, O. Pictet, An introduction to high frequency finance academic press, San Diego.
- [4] F. Strozzi, v. n. p. y. p. Zaldı?var, Jos´e-Manuel and Zbilut, Joseph P, journal=Physica A: Statistical Mechanics and its Applications, Applica- tion of nonlinear time series analysis techniques to high-frequency currency exchange data.
- [5] T. Mizuno, S. Kurihara, M. Takayasu, H. Takayasu, Analysis of high- resolution foreign exchange data of usd-jpy for 13 years, Physica A: Statis- tical Mechanics and its Applications 324 (1-2) (2003) 296–302.
- [6] T. Ohnishi, T. Mizuno, K. Aihara, M. Takayasu, H. Takayasu, Statistical properties of the moving average price in dollar–yen exchange rates, Physica A: Statistical Mechanics and its Applications 344 (1-2) (2004) 207–210.
- [7] F. Petroni, M. Serva, Real prices from spot foreign exchange market, Phys- ica A: Statistical Mechanics and its Applications 344 (1-2) (2004) 194–197.
- [8] Y. Aiba, N. Hatano, H. Takayasu, K. Marumo, T. Shimizu, Triangular arbi- trage in the foreign exchange market, in: The Application of Econophysics, Springer, 2004, pp. 18–23. 4
- [9] T. Suzuki, T. Ikeguchi, M. Suzuki, A model of complex behavior of inter- bank exchange markets, Physica A: Statistical Mechanics and its Applica- tions 337 (1-2) (2004) 196–218.
- [10] F. Wang, K. Yamasaki, S. Havlin, H. E. Stanley, Scaling and memory of intraday volatility return intervals in stock markets, Physical Review E 73 (2) (2006) 026117.
- [11] K. Kiyono, Z. R. Struzik, Y. Yamamoto, Criticality and phase transition in stock-price fluctuations, Physical review letters 96 (6) (2006) 068701.
- [12] H. Haken, Information and self-organization: A macroscopic approach to complex systems, Springer Science & Business Media, 2006.
- [13] D. J. MacKay, D. J. Mac Kay, Information theory, inference and learning algorithms, Cambridge university press, 2003.
- [14] S. Klenk, D. Thom, G. Heidemann, The normalized compression distance as a distance measure in entity identification, in: Industrial Conference on Data Mining, Springer, 2009, pp. 325–337.
- [15] R. Ant˜ao, A. Mota, J. T. Machado, Kolmogorov complexity as a data similarity metric: Application in mitochondrial dna, Nonlinear Dynamics 93 (3) (2018) 1059–1071.
- [16] R. Cilibrasi, P. Vitanyi, Clustering by compression ieee transaction on in- formation theory 51.
- [17] Q. Tian, P. Shang, G. Feng, The similarity analysis of financial stocks based on information clustering, Nonlinear Dynamics 85 (4) (2016) 2635–2652.
- [18] L.-Z. Liu, X.-Y. Qian, H.-Y. Lu, Cross-sample entropy of foreign exchange time series, Physica A: Statistical Mechanics and its Applications 389 (21) (2010) 4785–4792.
- [19] J. Xia, P. Shang, Multiscale entropy analysis of financial time series, Fluc- tuation and Noise Letters 11 (04) (2012) 1250033.
- [20] J. Xia, P. Shang, J. Wang, W. Shi, Classifying of financial time series based on multiscale entropy and multiscale time irreversibility, Physica A: Statistical Mechanics and Its Applications 400 (2014) 151–158.
- [21] A. Lin, P. Shang, X. Zhao, The cross-correlations of stock markets based on dcca and time-delay dcca, Nonlinear Dynamics 67 (1) (2012) 425–435.
- [22] A. Czapkiewicz, P. Majdosz, Grouping stock markets with time-varying copula-garch model, Finance a Uver 64 (2) (2014) 144.
- [23] Y. Yin, P. Shang, Modified dfa and dcca approach for quantifying the multiscale correlation structure of financial markets, Physica A: Statistical Mechanics and its

Applications 392 (24) (2013) 6442–6457.

- [24] A.-H. Sato, Characteristic time scales of tick quotes on foreign currency markets: an empirical study and agent-based model, The European Phys- ical Journal B-Condensed Matter and Complex Systems 50 (1-2) (2006) 137–140.
- [25] J. R. Hershey, P. A. Olsen, Approximating the kullback leibler divergence between gaussian mixture models, in: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, Vol. 4, IEEE, 2007, pp. IV–317.
- [26] M.-M. Deza, E. Deza, Dictionary of distances, Elsevier, 2006.
- [27] S. Kullback, R. A. Leibler, On information and sufficiency, The annals of mathematical statistics 22 (1) (1951) 79–86.
- [28] S. Kullback, Information theory and statistics, Courier Corporation, 1997.
- [29] S.-H. Cha, Comprehensive survey on distance/similarity measures between probability density functions, City 1 (2) (2007) 1.
- [30] C. E. Shannon, A mathematical theory of communication, Bell system technical journal 27 (3) (1948) 379–423.
- [31] J. Lin, Divergence measures based on the shannon entropy, IEEE Transac- tions on Information theory 37 (1) (1991) 145–151.
- [32] A. Majtey, P. Lamberti, D. Prato, Jensen-shannon divergence as a measure of distinguishability between mixed quantum states, Physical Review A 72 (5) (2005) 052310.
- [33] P. W. Lamberti, A. P. Majtey, Non-logarithmic jensen–shannon divergence, Physica A: Statistical Mechanics and its Applications 329 (1-2) (2003) 81– 90.
- [34] P. A. Olsen, S. Dharanipragada, An efficient integrated gender detection scheme and time mediated averaging of gender dependent acoustic models, in: Eighth European Conference on Speech Communication and Technol- ogy, 2003.
- [35] H. Printz, P. A. Olsen, Theory and practice of acoustic confusability, Com- puter Speech & Language 16 (1) (2002) 131–164.
- [36] J. Silva, S. Narayanan, Average divergence distance as a statistical discrim- ination measure for hidden markov models, IEEE Transactions on Audio, Speech, and Language Processing 14 (3) (2006) 890–906.
- [37] Q. Huo, W. Li, A dtw-based dissimilarity measure for left-to-right hidden markov models and its application to word confusability analysis, in: Ninth International Conference on Spoken Language Processing, 2006.
- [38] J. Goldberger, S. Gordon, H. Greenspan, An efficient image similarity mea- sure based on approximations of kl-divergence between two gaussian mix- tures, in: null, IEEE, 2003, p. 487.
- [39] A.-H. Sato, Application of spectral methods for high-frequency financial data to quantifying states of market participants, Physica A: Statistical Mechanics and its Applications 387 (15) (2008) 3960–3966.
- [40] S. Xu, M. Shao, W. Qiao, P. Shang, Generalized aic method based on higher-order moments and entropy of financial time series, Physica A: Sta- tistical Mechanics and its Applications 505 (2018) 1127–1138.
- [41] P. K. Sahoo, G. Arora, A thresholding method based on two-dimensional renyi's entropy, Pattern Recognition 37 (6) (2004) 1149–1161.
- [42] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Physical review letters 88 (17) (2002) 174102.

- [43] B. Fadlallah, B. Chen, A. Keil, J. Pr´ıncipe, Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information, Physical Review E 87 (2) (2013) 022911.
- [44] M. Xu, P. Shang, J. Huang, Modified generalized sample entropy and surro- gate data analysis for stock markets, Communications in Nonlinear Science and Numerical Simulation 35 (2016) 17–24.
- [45] T. Schreiber, Measuring information transfer, Physical review letters 85 (2) (2000) 461.
- [46] J. S. Richman, J. R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, American Journal of Physiology- Heart and Circulatory Physiology 278 (6) (2000) H2039–H2049.
- [47] J. Machado, Fractional order generalized information, Entropy 16 (4) (2014) 2350–2361.
- [48] I. Podlubny, Fractional differential equations, volume 198: An introduction to fractional derivatives, fractional differential equations, to methods of their...(mathematics in science and engineering).
- [49] R. Hilfer, Applications of fractional calculus in physics world scientific publ, Co., Singapore.
- [50] G. M. Zaslavsky, G. M. Zaslavskij, Hamiltonian chaos and fractional dy- namics, Oxford University Press on Demand, 2005.
- [51] V. E. Tarasov, Fractional dynamics: applications of fractional calculus to dynamics of particles, fields and media, Springer Science & Business Media, 2011.
- [53] B. Dumitru, D. Kai, S. Enrico, Fractional calculus: models and numerical methods, Vol. 3, World Scientific, 2012.
- [54] C. M. Ionescu, The human respiratory system, in: The Human Respiratory System, Springer, 2013, pp. 13–22.
- [55] J. T. Machado, Entropy analysis of integer and fractional dynamical sys- tems, Nonlinear Dynamics 62 (1-2) (2010) 371–378.
- [56] J. T. Machado, Fractional dynamics of a system with particles subjected to impacts, Communications in Nonlinear Science and Numerical Simulation 16 (12) (2011) 4596–4601.
- [57] J. T. Machado, Entropy analysis of fractional derivatives and their approx- imation, J. Appl. Nonlinear Dyn 1 (1) (2012) 109–112.
- [58] K. Kalpakis, D. Gada, V. Puttagunta, Distance measures for effective clus- tering of arima time-series, in: Proceedings 2001 IEEE international con- ference on data mining, IEEE, 2001, pp. 273–280.
- [59] R. Agrawal, K.-I. Lin, H. S. Sawhney, K. Shim, Fast similarity search in the presence of noise, Scaling, and Translation in Time Series Databases, IBM Almaden Research Center.
- [60] D. Rafiei, A. Mendelzon, Similarity-based queries for time series data, in: ACM SIGMOD Record, Vol. 26, ACM, 1997, pp. 13–25.