Measuring information interactions on the ordinal pattern of stock time series

Xiaojun Zhao,* Pengjian Shang, and Jing Wang

Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing 100044, People's Republic of China

(Received 5 June 2012; published 8 February 2013)

The interactions among time series as individual components of complex systems can be quantified by measuring to what extent they exchange information among each other. In many applications, one focuses not on the original series but on its ordinal pattern. In such cases, trivial noises appear more likely to be filtered and the abrupt influence of extreme values can be weakened. Cross-sample entropy and inner composition alignment have been introduced as prominent methods to estimate the information interactions of complex systems. In this paper, we modify both methods to detect the interactions among the ordinal pattern of stock return and volatility series, and we try to uncover the information exchanges across sectors in Chinese stock markets.

DOI: 10.1103/PhysRevE.87.022805   PACS number(s): 05.45.Tp, 05.10.−a, 89.70.Cf, 89.65.Gh

I. INTRODUCTION

In recent years, considerable techniques have been introduced to measure the interactions among systems or the information exchanges of individual components in single systems. Cross-correlation coefficients and detrended cross-correlation analyses were typically designed to analyze the cross correlations for linear and nonstationary series [1–3]. Both methods are constructed on the covariance estimation, while this may present spurious scaling behavior when the external trend arises [4,5]. Many methods are based on the evaluation of complexity or prediction, e.g., the inner composition alignment and Granger causality [6–8]. Many others use information-theoretic tools, including mutual entropy, conditional entropy, cross-sample entropy, and transfer entropy [9–12]. Each method emphasizes symmetric information exchange or asymmetric information transfer among complex systems. These methods have been applied to such areas of interest as physiology, economics, traffic, hydrology, and geography [13–19].

Nowadays, the interactions among financial markets have been extensively studied [20–25]. Specifically, the investigation on stock time series is an active area of research, which may recommend the selection of an assets portfolio. Many authors have focused on the composite indexes in nations or regions, and a few have considered the performances of individual companies or the sectors in stock markets [26–30]. Among these, they found high degrees of interdependence, indicating the sectors were highly integrated and exchanged information among each other. This paper modifies two prominent methods—the cross-sample entropy [11] and the inner composition alignment [6]—and tries to detect the interactions across sectors in Chinese stock markets. Stock markets in China have expanded rapidly following the establishment of two stock exchanges in Shanghai and Shenzhen in the early 1990s, the activity of which may have real economic effects now.

In this paper, we focus not on the original series but on its ordinal pattern. The underlying ideas behind such applications are as follows:

(i) In stock markets, large trading volumes may not be directly associated with high trading prices, while large trading volume changes appear closely with respect to high trading price changes [31–33]. It is often the case that the increase of one variable accompanies the increase or decrease of another variable. It is also expected to quantitatively detect such information linkage in stock trading price, return, and volatility across stock sectors.

(ii) In detail, a common issue of information-theoretic measures is how to accurately estimate the probability distribution of time series. There exist numerous approaches to estimate the probability distribution of the discrete variable. A straightforward approach is based on frequency statistics by covering the whole range of time series with equally sized sections and counting the number of values in each section, such as the histogram and kernel estimation, e.g., $\hat{p}_r(x_n) = \sum_{n'} \Theta(|x_n - x_{n'}| - r)/N$, where Θ is the Heaviside step function. The artificial parameter r is assigned as the tolerance of the accepting neighborhood. An alternative choice is to coarse grain the time series $\{x_n\}$ into discrete symbolic sequences $\{y_n\}$, e.g., by the step transformation $y_n = \Theta(x_n)$. Inspired from the permutation process of permutation entropy [34], we propose another technique by measuring the ordinal pattern, which divides the whole time series into sections and gathers local ranks in each section. An advantage is that it can deal with small-sized series. Moreover, trivial noises appear more likely to be filtered, and the abrupt influence of extreme values can be weakened.

We organize the paper as follows. In Sec. II, we introduce the cross-sample entropy and the inner composition alignment, both designed for the ordinal pattern, and a method to obtain the ordinal pattern from time series is proposed by the process of phase space reconstruction. Section III presents the empirical analysis of information interactions of return and volatility across Chinese stock sectors. A brief summary is then arranged in Sec. IV.

II. MEASURING INFORMATION INTERACTION

A. Ordinal pattern

In 2002, Bandt and Pompe introduced a reliable method to evaluate the probability distribution considering the time casuality and taking into account the temporal structure of

*Corresponding author: 05271060@bjtu.edu.cn

    ©2013 American Physical Society

time series [34]. The method is constructed on the phase space reconstruction in order to determine a probability distribution of the ordinal pattern. According to Takens's theory, if any single variable has sufficient accuracy for a long period of time, it is possible to reconstruct the underling dynamic structure of the entire system from the behavior of that single variable using delay coordinates and the embedding procedure [35]. The possibility of making a time-delay reconstruction of a phase space to view the dynamics of nonlinear systems has been extensively investigated.

From the time series $\{x_i\}$ of N observations, multivariate state vectors in d-dimensional space,

$$X_i = \{x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(d-1)\tau}\}, \tag{1}$$

are used to trace out the orbit of the system. τ is termed time delay and d is referred to as the embedding dimension. It is important to choose the proper time delay and embedding dimension in the phase space reconstruction procedure. Embedding dimension d is considered as the sufficient dimension for recovering the object without distorting any of its topological properties, and time delay τ is determined by the role that the reconstructed phase space will neither collapse into the main diagonal nor fill the entire phase space.

Through phase space reconstruction, a time series is mapped to a trajectory matrix X by the reconstruction parameters (d,τ):

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{N-(d-1)\tau} \end{bmatrix}. \tag{2}$$

Each state vector $X_i = \{x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(d-1)\tau}\}$ has an ordinal pattern by comparison with neighboring values. For example, suppose $d = 5$, a given value of τ, and the state vector $X_i = \{3.2, 9.7, 4.1, 6.3, 5.5\}$. X_i is sorted in ascending order to make sure that $x_{i+\pi_1\tau} \leqslant x_{i+\pi_2\tau} \leqslant x_{i+\pi_3\tau} \leqslant x_{t+\pi_4\tau} \leqslant x_{t+\pi_5\tau}$. As the relation $x_{i+0\tau} \leqslant x_{i+2\tau} \leqslant x_{i+4\tau} \leqslant x_{i+3\tau} \leqslant x_{i+\tau}$ holds, the ordinal pattern, also called the symbolic state vector, Π_i of X_i is $\{0,2,4,3,1\}$. The values of the ordinal pattern behave more stably and stationary for analysis than those of original irregular series [34]. The trajectory matrix X is then transformed into the ordinal pattern matrix:

$$\Pi = \begin{bmatrix} \Pi_1 \\ \Pi_2 \\ \vdots \\ \Pi_{N-(d-1)\tau} \end{bmatrix}. \tag{3}$$

The transformation from X_i to Π_i is termed permutation. It is clear that for a given embedding dimension d at most $d!$ permutations exist in total. We select all distinguishable permutations and number them Π_j, $j = 1,2,\ldots,d!$. For all the $d!$ possible permutations Π_j of order d, the relative frequency (No. denotes the number of elements) can be computed by

$$p(\Pi_j) = \frac{\text{No. } \{X_i | X_i \text{ has ordinal pattern } \Pi_j\}}{N-(d-1)\tau}, \tag{4}$$

where $1 \leqslant i \leqslant N-(d-1)\tau$.

B. Cross-sample entropy of ordinal pattern

Cross-sample entropy is a symmetric measure to assess the degree of synchrony of time series. It was generalized by the sample entropy designed to measure the irregularity of time series. Here we make a modification on the cross-sample entropy based on the ordinal pattern. Consider two time series $\{x_i\}$ and $\{y_i\}$, $i = 1,2,\ldots,N$. Given a matching (embedding) dimension d and a time delay τ, reconstruct vector series $X_i^d = (x_i, x_{i+\tau}, \ldots, x_{i+(d-1)\tau})$, $1 \leqslant i \leqslant N-(d-1)\tau$ and $Y_j^d = (y_j, y_{j+\tau}, \ldots, y_{j+(d-1)\tau})$, $1 \leqslant j \leqslant N-(d-1)\tau$. Define $B_i^d(y||x)$ as a binary variable that sets one when Y_i^d has the same ordinal pattern with X_i^d and zero otherwise, where i ranges from 1 to $N-(d-1)\tau$. Then define

$$B = \frac{\sum_{i=1}^{N-(d-1)\tau} B_i^d(y||x)}{N-(d-1)\tau}. \tag{5}$$

This is the probability that the ordinal patterns of two series simultaneously match for d points. In the original cross-sample entropy version, the matching process is determined by artificially introducing a tolerance parameter r and counting the number at which the distance of two vectors falls in the range $[0,r]$. Similarly, expand matching dimension d to $d+1$ and define $A_i^{d+1}(y||x)$ as the boolean variable of whether vector Y_i^{d+1} has the same ordinal pattern with X_i^{d+1}. Therefore, the probability that the ordinal patterns of two series match for $d+1$ points is

$$A = \frac{\sum_{i=1}^{N-d\tau} A_i^d(y||x)}{N-d\tau}. \tag{6}$$

Finally, the cross-sample entropy $S_d(x,y)$ is derived by

$$S_d(x,y) = -\log(A/B). \tag{7}$$

S_d measures the synchrony of two time series. We discuss several special cases. It can be inferred that for two completely random series the cross-sample entropy of the ordinal pattern is $S_d = -\log[1/(d+1)!/(1/d!)] = \log(d+1)$, since a random series increases or decreases with equal possibility. For two monotonous series with the same direction, $S_d = -\log(1/1) = 0$, while for two monotonous series with opposite directions both A and B approach zero. In such a case, $S_d = -\log(0/0)$ appears meaningless. However, it still indicates strong asynchrony if A or B equals zero. In real-world applications, this rarely occurs and does not need to be considered. We also propose a new statistic to test the degree of synchrony:

$$\rho_d(x,y) = S_d(x,y)/\log(d+1), \tag{8}$$

where ρ_d generally falls in the range of $[0,1]$. A small value of ρ_d represents a strong synchrony, and vice versa.

C. Inner composition alignment of ordinal pattern

The inner composition alignment was proposed to detect regulatory links from very short time series that facilitated the understanding of emerging structures in complex networks. Here we modify the method on the ordinal pattern. Also consider series $\{x_i\}$ and $\{y_i\}$, $i = 1,2,\ldots,N$. First reconstruct $\{x_i\}$ and $\{y_i\}$ into phase space by the reconstruction parameters

(d,τ), respectively:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{N-(d-1)\tau} \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{N-(d-1)\tau} \end{bmatrix}. \qquad (9)$$

Let $\Pi_i^{(X)}$ be the permutation which arranges each state vector X_i in a nondecreasing order. The new state vector $Z_i = Y_i(\Pi_i^{(X)})$ is the reordering of the state vector Y_i with respect to $\Pi_i^{(X)}$. Then let $Z = [Z_1, Z_2, \ldots, Z_{N-(d-1)\tau}]^\Gamma$. In the inner composition alignment, the monotonicity of reordered series Z_i is quantified by counting the number of intersection points of the series with the horizontal lines which are drawn from each of the time points. In this paper, we use the entropy of the ordinal pattern to estimate the monotonicity of new series. As indicated in Ref. [34], such consideration tends to be more effective. Each vector Z_i in Z presents a unique ordinal pattern forming $\Pi = [\Pi_1, \Pi_2, \ldots, \Pi_{N-(d-1)\tau}]^\Gamma$.

For all possible $d!$ permutations, the occurrence probability of each ordinal pattern is approximated by the relative frequency $P(\Pi_j)$. The information of the ordinal pattern Π_j is measured by $H = -\log[P(\Pi_j)]$. By averaging all the pattern information, the Shannon entropy of ordinal patterns is obtained:

$$H_d(x,y) = -\sum_{j=1}^{d!} P(\Pi_j)\log[P(\Pi_j)]. \qquad (10)$$

H_d measures the coupling strength of two time series. The random series still increases or decreases in uncertain ways through the reordering process by the inner composition alignment. Therefore, for two random series with large enough size, each ordinal pattern Π_i emerges with equal probability for all possible $d!$ permutations, i.e., $P(\Pi_i) = 1/d!$. In such a case, $H_d = \log(d!)$. If x and y are the same series, reordered state vector Z_i appears to increase monotonically. Then it can be easily inferred that $H_d = 0$. We also introduce a new statistic to quantitatively measure the coupling strength:

$$\sigma_d(x,y) = H_d(x,y)/\log(d!), \qquad (11)$$

where theoretically $\sigma_d \in [0,1]$. In brief, a small value of σ_d represents a large coupling strength of two series, while a large value of σ_d represents a small coupling strength. The monotonicity of reordered time series by the order of x is different from that of y. The information transfer from y to x can be obtained from the opposite direction. Even so, it can be inferred that the reordering process is a one-to-one mapping that σ_d keeps constant when information transfer from x to y changes into that from y to x.

III. INFORMATION INTERACTION ACROSS CHINESE STOCK SECTORS

A. Data description

We use daily closing prices of 18 Chinese sector indexes mixed by Shanghai and Shenzhen markets. The data expand from 6 January 2009 to 9 May 2012, with a total of 810 daily observations of each sector [36]. The data cover almost all fields of industries in Chinese stock markets, consisting of communication, construction, extraction, finance, food, forestry, information, machinery, metals, paper, petrochemistry, real estate, service, synthesis, textile, utility, wholesale and retail, and wood industries. Table I presents these sectors and lists them from 1 to 18. We consider the return series r_i as $r_i = \log(x_{i+1}) - \log(x_i)$ and volatility v_i as $v_i = |r_i|$, supposing the daily closing price in a sector at day i is x_i. The traditional statistics of return (columns 2–5) and volatility (columns 6–9) series including mean value, standard deviation,

TABLE I. Eighteen sectors numbered 1–18 in Chinese stock markets. Columns 2–5 and 6–9, respectively, show the traditional statistics including mean value, standard deviation (SD), minimum and maximum of return, and volatility series.

	Return				Volatility			
Sector	Mean	SD	Minimum	Maximum	Mean	SD	Minimum	Maximum
1 Communication	0.0003	0.0211	−0.0816	0.0740	0.0160	0.0138	0.0001	0.0816
2 Construction	0.0000	0.0162	−0.0747	0.0532	0.0121	0.0108	0.0000	0.0747
3 Extraction	0.0007	0.0219	−0.0928	0.0797	0.0162	0.0147	0.0000	0.0928
4 Finance	0.0004	0.0180	−0.0654	0.0773	0.0133	0.0122	0.0000	0.0616
5 Food	0.0010	0.0170	−0.0609	0.0616	0.0129	0.0111	0.0000	0.0616
6 Forestry	0.0005	0.0214	−0.0802	0.0915	0.0163	0.0138	0.0000	0.0915
7 Information	0.0005	0.0185	−0.0670	0.0577	0.0142	0.0119	0.0000	0.0670
8 Machinery	0.0007	0.0186	−0.0737	0.0616	0.0142	0.0120	0.0000	0.0737
9 Metals	0.0006	0.0210	−0.0918	0.0740	0.0157	0.0139	0.0000	0.0918
10 Paper	0.0005	0.0201	−0.0799	0.0596	0.0150	0.0134	0.0000	0.0799
11 Petrochemistry	0.0005	0.0199	−0.0758	0.0660	0.0151	0.0130	0.0000	0.0799
12 Real estate	0.0005	0.0218	−0.0922	0.0752	0.0161	0.0146	0.0000	0.0922
13 Service	0.0006	0.0191	−0.0810	0.0588	0.0143	0.0127	0.0000	0.0810
14 Synthesis	0.0006	0.0200	−0.0905	0.0626	0.0151	0.0131	0.0000	0.0905
15 Textile	0.0007	0.0193	−0.0770	0.0579	0.0145	0.0127	0.0000	0.0770
16 Utility	0.0001	0.0159	−0.0625	0.0504	0.0118	0.0107	0.0001	0.0625
17 Wholesale & retail	0.0006	0.0179	−0.0753	0.0624	0.0135	0.0118	0.0000	0.0753
18 Wood	0.0007	0.0223	−0.0868	0.0811	0.0167	0.0147	0.0000	0.0868

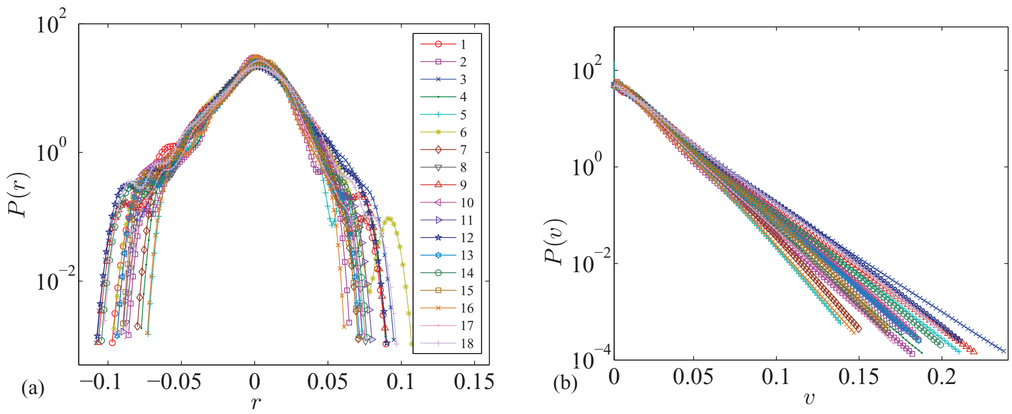

FIG. 1. (Color online) The probability distribution of (a) return and (b) volatility series in 18 sectors.

minimum, and maximum in each sector are presented in the whole time range to allow an overview on the data. We also illustrate the probability distribution of return and volatility series in Fig. 1.

For return series, the maximal mean value occurs in the food sector and the minimal mean value appears in the construction sector. The markets show a more prosperous economy with time, since the mean values of return in all sectors are non-negative. For volatility series, the maximal mean value occurs in the wood sector and the minimal mean value still appears in the construction sector. We first calculate the linear cross-correlation coefficients $r_{x,y} = \sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})/\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}$ of return and volatility and show them in Fig. 2, where $\bar{x}$ represents the mean value of the x series. The coefficients on the diagonal stay at 1 when $x_i \equiv y_i$, $i = 1, \ldots, N$. Two sectors numbered 4 and 12 corresponding to finance and real estate show lower coefficient values. This indicates that these two sectors are more linearly independent and the linear information exchanges with other sectors are less frequent. We also use the dendrograms to illustrate the arrangement of the clusters produced by

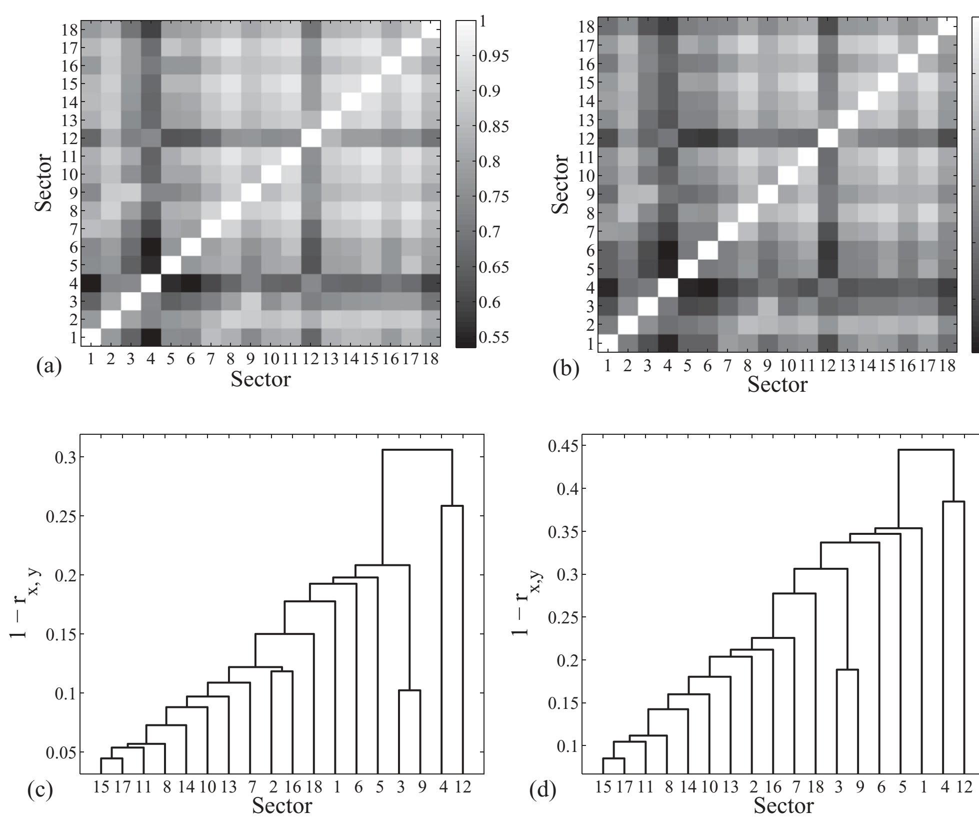

FIG. 2. The linear cross-correlation coefficients $r_{x,y}$ and the dendrogram based on $1 - r_{x,y}$ for (a), (c) return and (b), (d) volatility series across Chinese stock sectors. We consider $1 - r_{x,y}$ rather than $r_{x,y}$ because $r_{x,y}$ increases with the strength of cross correlation, so we use $1 - r_{x,y}$ as the distance to the cluster.

hierarchical clustering. For return and volatility series, sectors 4 and 12 are different from other sectors, as they belong to one cluster while other sectors belong to the other cluster.

B. Cross-sample entropy across Chinese stock sectors

We first use the cross-sample entropy to investigate the interactions in the daily return series across Chinese stock sectors. We mainly consider the case of $\tau = 1$ [see Eq. (1)]. The value of d is closely related to the length of series and can be made large if the series is long enough. To study the effect of finiteness of the series, we generate 50 groups of random series with length 809 in each group and calculate the cross-sample entropy between any two series. We find $S_2 = 1.1028 \pm 0.0745$ (mean $\pm$ standard deviation), $S_3 = 1.4093 \pm 0.1674, S_4 = 1.7190 \pm 0.4202$, and $S_5 = 1.5310 \pm 0.4596$. For $d = 2,3,4,5$, the theoretical cross-sample entropy of random series is $\log(3) = 1.0986, \log(4) = 1.3863, \log(5) = 1.6094, \log(6) = 1.7918$. It is unambiguous that the standard deviation increases with the increase of d. The difference between the empirical cross-sample entropy and theoretical value also increases with d due to the finiteness, but it is rather close to zero for $d = 2,3$. Therefore we set $d = 2,3$ considering the length of daily observations in Chinese stock sectors.

We show the cross-sample entropy of return series in Fig. 3. The figures are symmetric and the values on the diagonal represent the sample entropy of single series [see Eq. (7)]. When $d = 2$, the cross-sample entropy falls in the range of [0,0.6265]. This indicates that, if the return for two days in one series has the same pattern as that in another series, it is uncertain that the return for three days in two series will have the same pattern. The sectors that own the three largest cross-sample entropies with other sectors on average are 4, 12, and 5 (corresponding to finance, real estate, and food). The three smallest cross-sample entropies on average, indicating the largest degree of synchrony, occur in 11, 15, and 17 (corresponding to petrochemistry, textile, and wholesale and retail). This can be more clearly observed from the dendrogram of return in Fig. 3. The results are rather similar for $d = 3$. In general, the finance, real estate, and food sectors are revealed to be more independent from other sectors, while the petrochemistry, textile, and wholesale and retail sectors behave more dependently. Therefore we classify the sectors into three clusters: (i) more independent sectors (4, 5, 12), (ii) more dependent sectors (11, 15, 17), and (iii) other sectors. The linear cross-correlation coefficients and the results reported in Ref. [27], where the finance sector is also found to be more independent, are partly consistent with our findings.

We turn now to estimate the cross-sample entropy of volatility series, and we show them in Fig. 4. For $d = 2$, the cross-sample entropy falls in the range of [0,0.8787]. As a small cross-sample entropy corresponds to a large degree of synchrony, the larger interval compared to that of return series indicates the volatility series show a smaller degree of synchrony. We get the same results from the cross-correlation

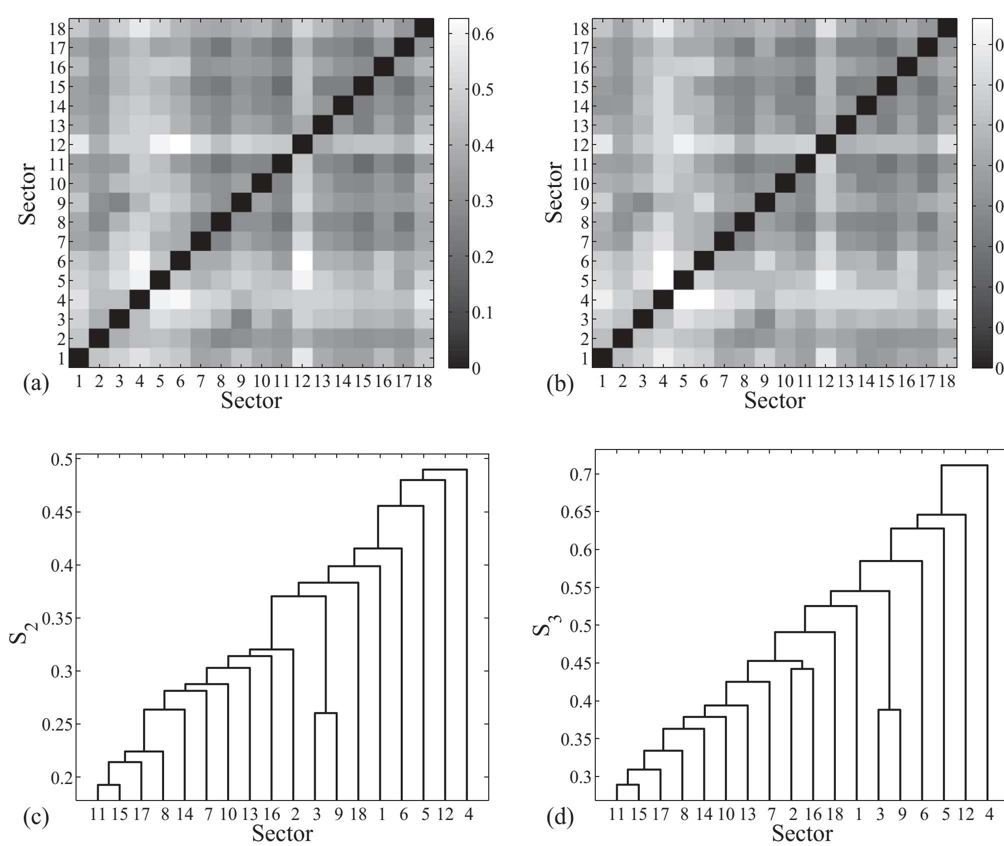

FIG. 3. The cross-sample entropy of the ordinal pattern of daily return series across Chinese stock sectors for (a) $d = 2$ and (b) $d = 3$. (c), (d) Dendrograms based on the values of the cross-sample entropy.

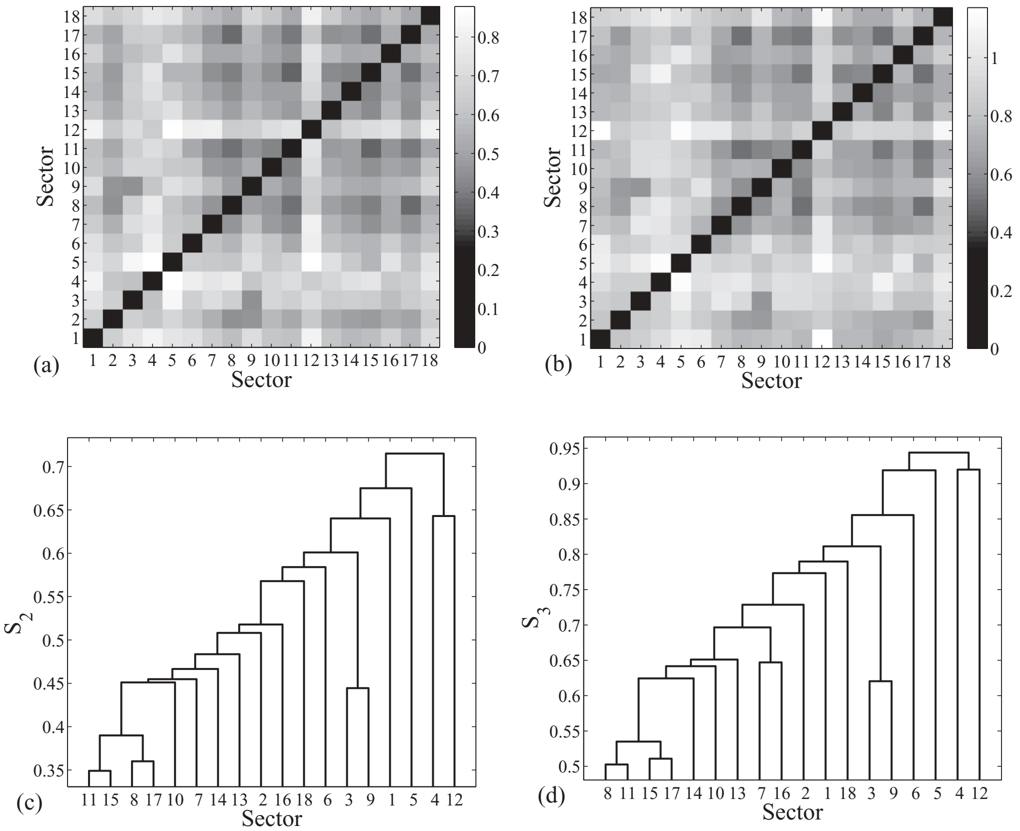

FIG. 4. The cross-sample entropy of the ordinal pattern of daily volatility series across Chinese stock sectors for (a) $d = 2$ and (b) $d = 3$. (c), (d) Dendrograms using the values of the cross-sample entropy as the distance to the cluster.

coefficients since $r_{x,y}$ of return series is always larger than that of volatility series. We also classify the sectors into three clusters: (i) more independent sectors (4, 12), (ii) more dependent sectors (8, 11, 15, 17), and (iii) other sectors. One can observe that the cross-sample entropy generally increases with d.

Statistical inferences based on estimation and hypothesis testing are among the most important aspects of the decision making process in science and business. It is interesting to uncover the statistical significance of the cross-sample entropy difference (i) between the sector and random series and (ii) between any two sectors. For (i), a t test rejects the hypothesis that data in the cross-sample entropy of return and those of random series are with equal means at the 5% significance level, which holds for all sectors. Before that, a Kolmogorov-Smirnov test cannot reject the hypothesis that the normal distribution is the correct description for cross-sample entropy of any two sectors at the 5% significance level. It indicates that for return series the synchrony between sectors behaves significantly different from that of random series. We also got the same results from volatility series. For (ii), we propose a null hypothesis to test whether sectors i and j have the same mean value of cross-sample entropy with other sectors k:

$$h_0 : \bar{S_d}(r_i,r_k) = \bar{S_d}(r_j,r_k), \tag{12}$$

and an alternative hypothesis $h_1 : \bar{S_d}(r_i,r_k) \neq \bar{S_d}(r_j,r_k)$, where $k = 1,2,\ldots,18$ and $\bar{S_d}$ represents the average over k. In Fig. 5, $h_0 = 1$ (black cell) holds if a t test rejects the null hypothesis at the 5% significance level, and $h_0 = 0$ (white cell) otherwise. For some sectors such as 4,5,12, their mean cross-sample entropies appear more likely to be different from others.

We also calculate ρ_d [see Eq. (8)] and get $|\rho_2(r_i,r_j) - \rho_3(r_i,r_j)| < 0.1$ with confidence level 99.4% for return series and $|\rho_2(v_i,v_j) - \rho_3(v_i,v_j)| < 0.1$ with confidence level 95.5% for volatility series, where $i = 1,2,\ldots,18$; $j = 1,2,\ldots,18$; and $i \neq j$. Although cross-sample entropy S_d increases with d, $\partial\rho_d/\partial d = 0$ approximately holds. The measure ρ_d provides a possibility to gain insight into the properties of series even if they are not long enough as one can set $d = 2$. Moreover, the oscillation of ρ may also be observed to uncover the dynamical changes of long series.

C. Inner composition alignment across Chinese stock sectors

In this section, we use the inner composition alignment designed for the ordinal pattern to investigate the stock return and volatility series across Chinese stock sectors. We also set $\tau = 1$. Due to the finiteness of the series, we generate 50 groups of random series and calculate H_d for different d. We find $H_2 = 0.6924 \pm 0.0011$ (mean $\pm$ standard deviation), $H_3 = 1.7873 \pm 0.0028, H_4 = 3.1577 \pm 0.0063, H_5 = 4.6865 \pm 0.0136, H_6 = 5.9782 \pm 0.0245$, and the theoretical values are $\log(2!) = 0.6931, \log(3!) = 1.7918, \log(4!) = 3.1781, \log(5!) = 4.7875, \log(6!) = 6.5793$. One can get the correct results for $d = 2,3,4,5$ of random series while large deviation occurs at $d = 6$. This also indicates the inner

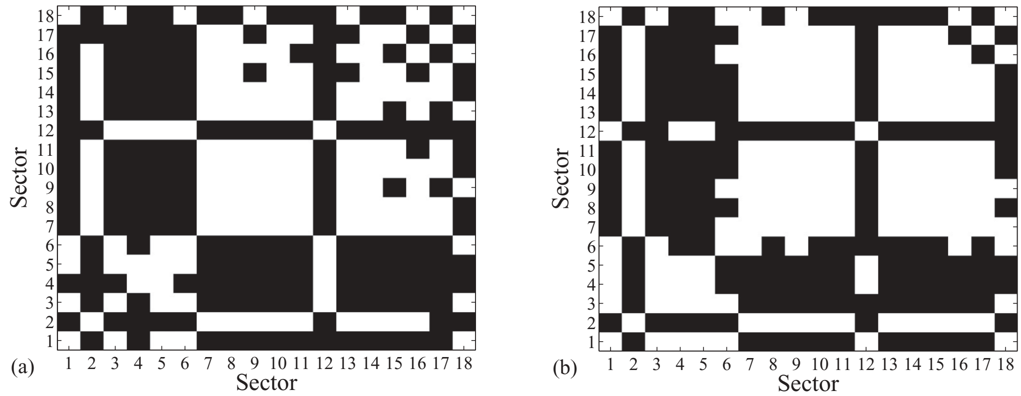

FIG. 5. $h_0 = 1$ (black cell) if one rejects the null hypothesis that the mean cross-sample entropy equals for two sectors and $h_0 = 0$ otherwise (white cell) for (a) return series and (b) volatility series.

composition alignment is more effective than cross-sample entropy (where deviation occurs at $d = 4$) in dealing with small-sized series. We set $d = 2,3,4,5$ and calculate H_d of return series in Fig. 6. According to the algorithm of inner composition alignment, the entropy on the diagonal should be zero. Since for arbitrary i reordered state vector Z_i, that reorders Y_i by the local rank of X_i, appears to increase monotonically if $X_i \equiv Y_i, i = 1, \ldots, N$. Therefore, we replace the values of zero by the theoretical results of two complete random series on the diagonal as $\log(d!)$. It can be observed that the values of H_d of return series behave significantly different from those of random series. A small value of H_d represents a high coupling strength. It indicates that return series across Chinese sectors exchange information with each other and tend to move consistently rather than disorderly. It may be partially driven by the similar policy stimulations, investment environments, and the development of the markets, or it may be partially due to the fact that information flows across sectors. We also observe the cluster phenomena of different sectors from the dendrogram. Sectors 8, 11, 15, and 17 are more dependent with other sectors, and 4, 5, and 12 are more independent. Such a relation does not change with d.

We then estimate the entropy of the ordinal pattern of volatility series. We also take the place of zero on the diagonal by the entropy of two random series for $d = 2,3,4,5$ in Fig. 7. Many similar results show up compared to those of return series. The sectors that perform more independently for return series remain the same for volatility series, such as the finance and real estate sectors. They show larger entropy than other sectors and exchange less information with other sectors, which is consistent with the findings by the cross-sample entropy. The machinery and wholesale and retail sectors appear to be more dependent with other sectors. Moreover, the values of entropy for volatility series are closer to the theoretical results of random series than those for return series. It can thus be concluded that the linkages for volatility series are weaker than those for the return series.

It is interesting to uncover the statistical significance of the difference of H_d(i) between the sector and random series (ii) between any two sectors. A Kolmogorov-Smirnov test

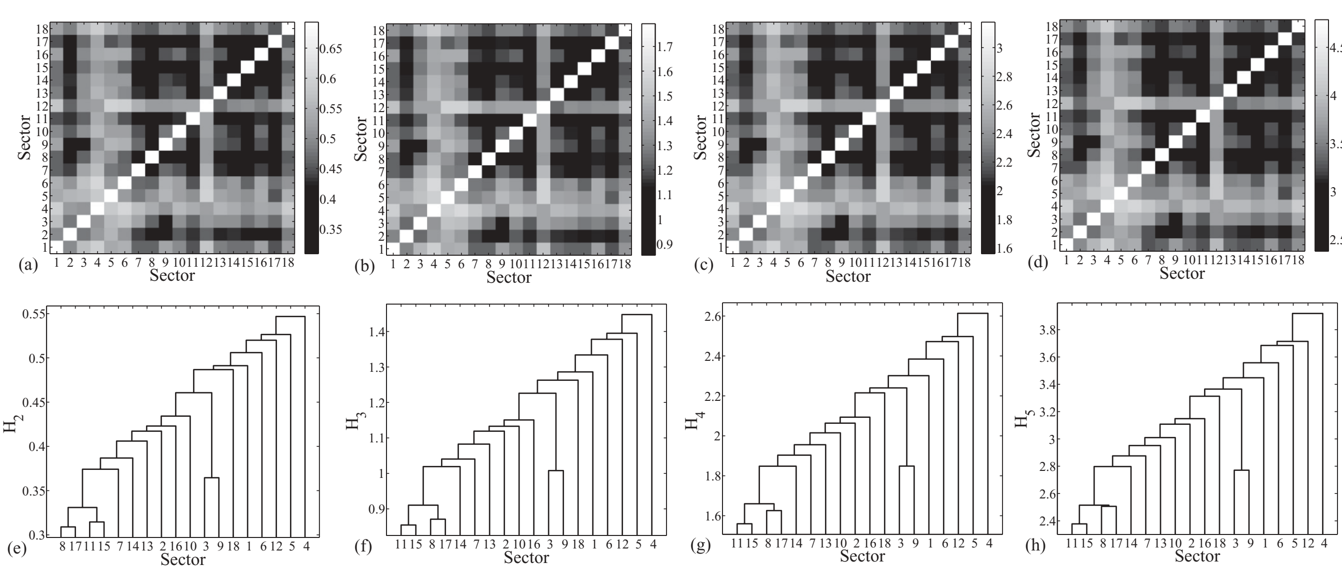

FIG. 6. The entropy derived by inner composition alignment of the ordinal pattern of return series across Chinese stock sectors for (a) $d = 2$, (b) $d = 3$, (c) $d = 4$, and (d) $d = 5$. We use the entropy as the distance to the cluster in the dendrograms.

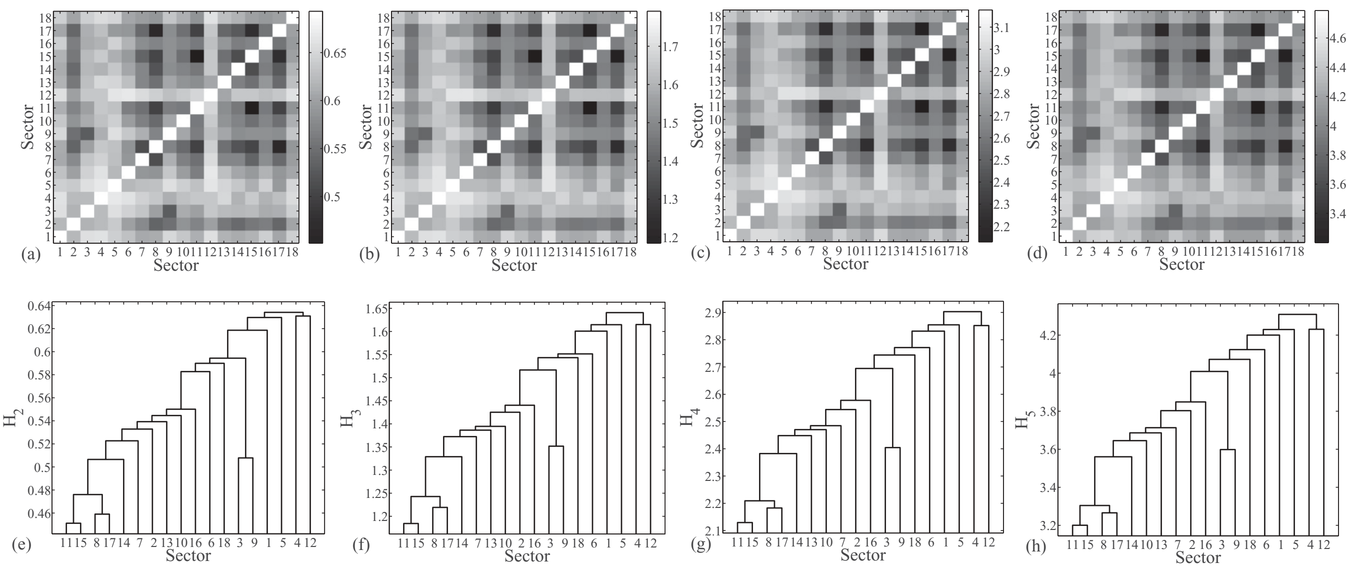

FIG. 7. The entropy derived by inner composition alignment of the ordinal pattern of volatility series across Chinese stock sectors for (a) $d = 2$, (b) $d = 3$, (c) $d = 4$, and (d) $d = 5$. We use the entropy as the distance to the cluster in the dendrograms.

cannot reject the hypothesis that the normal distribution is the correct description for cross-sample entropy of any two sectors at the 5% significance level, and all comparisons are significant by a t test. This indicates that for return series the synchrony between sectors behaves significantly different from that of random series. For (ii), we also propose a null hypothesis to test whether sectors i and j have the same mean value of entropy with other sectors k:

$$h_0 : \bar{H}_d(r_i,r_k) = \bar{H}_d(r_j,r_k), \tag{13}$$

and an alternative hypothesis $h_1 : \bar{S}_d(r_i,r_k) \neq \bar{S}_d(r_j,r_k)$, where $k = 1,2,\ldots,18$; $d = 2$; and $\bar{S}_d$ represents the average over k. In Fig. 8, $h_0 = 1$ (black cell) if a t test rejects the null hypothesis at the 5% significance level, and $h_0 = 0$ (white cell) otherwise.

Although the values of entropy H_d generally increase with d, σ_d almost keeps constant, i.e., $\partial(\sigma_d)/\partial(d) \approx 0$. We also get $\max|\sigma_{d_1}(r_i,r_j) - \sigma_{d_2}(r_i,r_j)| < 0.06$ with confidence level 96.7% for return series and $\max|\sigma_{d_1}(v_i,v_j) - \sigma_{d_2}(v_i,v_j)| < 0.04$ with confidence level 98.0% for volatility series, where $d_1,d_2 = 2,3,4,5; m \neq n; i,j = 1,2,\ldots,18$; and $i \neq j$. So the measure σ_d provides a possibility to gain insight into the properties of series even if they are not long enough.

IV. CONCLUSIONS

In this paper, the information interactions of daily return and volatility series across Chinese sectors are investigated. The ordinal pattern of the series is considered, where the local rank of series is obtained by the technique of phase space reconstruction and permutation process. The cross-sample entropy and the entropy derived by the inner composition alignment method are specially designed for the information detection of the ordinal pattern. Several sectors, such as finance, real estate, and food, show less information exchanges with other sectors; thus, they behave more independently, while some sectors, such as machinery, petrochemistry, textile, and wholesale and retail, exchange information more frequently and rely on other sectors more, since they exhibit more linkages with other sectors. We also compare the cross-sample entropy and the entropy by the inner composition alignment method with

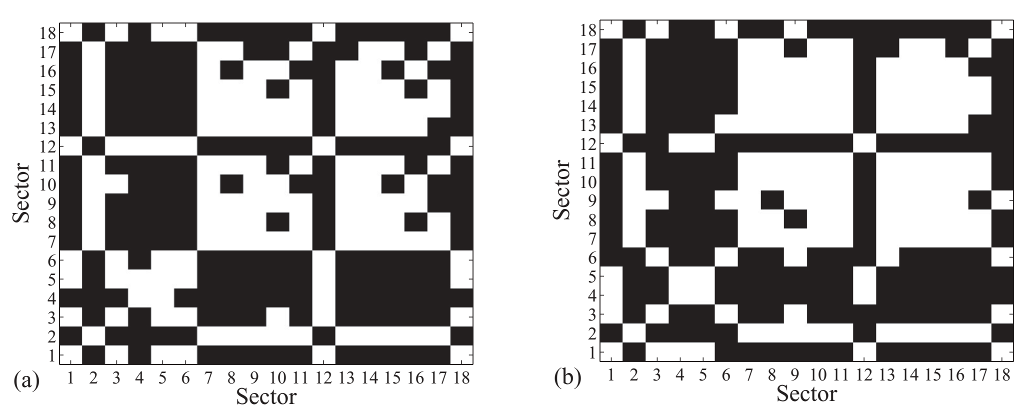

FIG. 8. $h_0 = 1$ (black cell) if one rejects the null hypothesis that the mean H_d equals for two sectors and $h_0 = 0$ otherwise (white cell) for (a) return series and (b) volatility series.

theoretical results of random series. They are lower than the theoretical entropy of random series, which indicates that the return and volatility series tend to move consistently with large possibility. Additionally, the return series show stronger synchrony than volatility series across sectors. We propose two parameters—ρ_d based on cross-sample entropy and σ_d based on inner composition alignment—both of which do not change with d. They actually provide the possibility to study the synchrony and coupling strength for small-sized series with small d. Moreover, σ_d is more robust than ρ_d for small-sized series, according to the comparison between empirical and theoretical results of random series.

As noted in the reference of permutation entropy, the advantages of the permutation process are its simplicity, robustness, and invariance with respect to nonlinear monotonous transformations. We expect that the modified methods for the ordinal pattern will provide a statistically robust basis to assess the information interactions in many fields of science, especially in geophysics.

ACKNOWLEDGMENTS

Financial support from the Fundamental Research Funds for the Central Universities (Grant No. 2012YJS115), the China National Science (Grant No. 61071142), the Beijing National Science (Grant No. 4122059), and the National High Technology Research Development Program of China (863 Program) (Grant No. 2011AA110306) is gratefully acknowledged.

[1] S. Arianos and A. Carbone, J. Stat. Mech. (2009) P03037.
[2] B. Podobnik and H. E. Stanley, Phys. Rev. Lett. **100**, 084102 (2008).
[3] W. Zhou, Phys. Rev. E **77**, 066211 (2008).
[4] X. Zhao, P. Shang, A. Lin, and G. Chen, Physica A **390**, 3670 (2011).
[5] X. Zhao, P. Shang, C. Zhao, J. Wang, and R. Tao, Chaos Solitons Fractals **45**, 166 (2012).
[6] S. Hempel, A. Koseska, J. Kurths, and Z. Nikoloski, Phys. Rev. Lett. **107**, 054101 (2011).
[7] C. W. J. Granger, Econometrica **37**, 424 (1969).
[8] D. Marinazzo, M. Pellicoro, and S. Stramaglia, Phys. Rev. Lett. **100**, 144103 (2008).
[9] S. Frenzel and B. Pompe, Phys. Rev. Lett. **99**, 204101 (2008).
[10] S. Pincus and B. H. Singer, Proc. Natl. Acad. Sci. USA **93**, 2083 (1996).
[11] J. S. Richman and J. R. Moorman, Am. J. Physiol. Heart Circ. Physiol. **278**, 2039 (2000).
[12] T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[13] M. Costa, A. L. Goldberger, and C.-K. Peng, Phys. Rev. Lett. **89**, 068102 (2002).
[14] X. Zhao, P. Shang, and Q. Jin, Fractals **19**, 329 (2011).
[15] O. Kwon and J.-S. Yang, Europhys. Lett. **82**, 68003 (2008).
[16] L. liu, X. Qian, and H. Lu, Physica A **389**, 4785 (2008).
[17] N. Xu, P. Shang, and S. Kamae, Nonlinear Dynamics **61**, 207 (2008).
[18] S. Hajian and M. S. Movahed, Physica A **389**, 4942 (2008).
[19] S. Shadkhoo and G. R. Jafari, Eur. Phys. J. B **72**, 679 (2009).
[20] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, Phys. Rev. E **65**, 066126 (2002).
[21] B. Podobnik, D. Horvatic, A. M. Petersen, and H. E. Stanley, Proc. Natl. Acad. Sci. USA **106**, 22079 (2009).
[22] L. Zunino, M. Zanin, B. M. Tabak, D. G. Pérez, and O. A. Rosso, Physica A **389**, 1891 (2010).
[23] X. Zhao, P. Shang, and Y. Pang, Fluctuation and Noise Letters **9**, 203 (2010).
[24] R. Marschinski and H. Kantz, Eur. Phys. J. B **30**, 275 (2002).
[25] A. Utsugi, K. Ino, and M. Oshikawa, Phys. Rev. E **70**, 026110 (2004).
[26] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, Phys. Rev. Lett. **83**, 1471 (1999).
[27] Z. Wang, A. M. Kutan, and J. Yang, Quarterly Review of Economics and Finance **45**, 767 (2005).
[28] D. H. Kim and H. Jeong, Phys. Rev. E **72**, 046133 (2005).
[29] R. K. Pan and S. Sinha, Phys. Rev. E **76**, 046116 (2007).
[30] J. Shen and B. Zheng, Europhys. Lett. **86**, 48005 (2009).
[31] C. Hiemstra and J. D. Jones, Journal of Finance **49**, 1639 (1994).
[32] M. Smirlock and L. Starks, Journal of Banking and Finance **12**, 31 (1988).
[33] Y. Yuan, X. Zhuang, and Z. Liu, Physica A **391**, 3484 (2012).
[34] C. Bandt and B. Pompe, Phys. Rev. Lett. **88**, 174102 (2002).
[35] F. Takens, *Dynamical Systems and Turbulence* (Springer, Berlin, 1981).
[36] Data source: http://www.51ifind.com.