Statistical tests for power-law cross-correlated processes

Boris Podobnik,[1,2,3] Zhi-Qiang Jiang,[4,5] Wei-Xing Zhou,[4,5] and H. Eugene Stanley[2]
[1] *Faculty of Civil Engineering, University of Rijeka, 51000 Rijeka, Croatia*
[2] *Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA*
[3] *Department of Economics, University of Ljubljana, 1000 Ljubljana, Slovenia*
[4] *School of Business, East China University of Science and Technology, Shanghai 200237, China*
[5] *Research Center for Econophysics, East China University of Science and Technology, Shanghai 200237, China*
(Received 6 September 2011; revised manuscript received 30 November 2011; published 22 December 2011)

For stationary time series, the cross-covariance and the cross-correlation as functions of time lag n serve to quantify the similarity of two time series. The latter measure is also used to assess whether the cross-correlations are statistically significant. For nonstationary time series, the analogous measures are detrended cross-correlations analysis (DCCA) and the recently proposed detrended cross-correlation coefficient, $\rho_{\rm DCCA}(T,n)$, where T is the total length of the time series and n the window size. For $\rho_{\rm DCCA}(T,n)$, we numerically calculated the Cauchy inequality $-1 \leqslant \rho_{\rm DCCA}(T,n) \leqslant 1$. Here we derive $-1 \leqslant \rho_{\rm DCCA}(T,n) \leqslant 1$ for a standard variance-covariance approach and for a detrending approach. For overlapping windows, we find the range of $\rho_{\rm DCCA}$ within which the cross-correlations become statistically significant. For overlapping windows we numerically determine—and for nonoverlapping windows we derive—that the standard deviation of $\rho_{\rm DCCA}(T,n)$ tends with increasing T to $1/T$. Using $\rho_{\rm DCCA}(T,n)$ we show that the Chinese financial market's tendency to follow the U.S. market is extremely weak. We also propose an additional statistical test that can be used to quantify the existence of cross-correlations between two power-law correlated time series.

DOI: 10.1103/PhysRevE.84.066118 PACS number(s): 89.65.Gh, 89.20.−a, 02.50.Ey

I. INTRODUCTION

Many diversified complex systems are composed of constituents that mutually interact in a complex fashion in that the current output value of each constituent depends on the current values of other constituent outputs. The complexity of the mutual interaction can be additionally studied if memory is included, where the current value of each constituent output depends not only on its own past but also on the past values of other constituent outputs. Such complex systems are characterized by both long-range correlations and long-range cross-correlations. Many diverse systems exhibit these properties, ranging from colloidal glasses [1], geophysics [2], and finance [3] to solid-state physics [4]. In colloidal glasses, different cross-correlation methods have revealed and classified a hidden local order within the disorder [1]. In seismology, the degree of cross-correlation among signals registered by different antennas provides an alert that signals earthquakes [2]. In finance, risk is estimated on the basis of cross-correlation matrices for different assets [3]. In nanodevices for quantum information processing, electronic entanglement allows us to compute noise cross-correlations and find whether the sign of the signal is the reverse of what would be expected of standard devices [4].

When cross-correlations are present in stationary time series, at least two measures quantify cross-correlations among the constituents: the cross-correlations as a function of time window n and the cross-covariance. Both measures are related in that the first measure also quantifies whether the cross-correlations are statistically significant.

When nonstationarities are present, Ref. [5] proposes a method, based on detrended covariance, called *detrended cross-correlation analysis* (DCCA), which is a modification of standard covariance analysis in which the global average is replaced by local trends [5–7]. The DCCA method can be easily generalized for multifractal analysis [8–11]. The DCCA method and its multifractal version have been applied to diverse fields including financial data [6,12–16], traffic flows [17–19], seismic data [20], sunspot numbers and river flow fluctuations [21], and meteorological data [22].

For nonstationary time series, in analogy with the cross-correlation coefficient, Ref. [23] recently introduced a DCCA cross-correlation coefficient, but this measure does not quantify the significance of cross-correlations between different time series. Here we examine the statistical significance of the DCCA cross-correlation coefficient with the goal of making it applicable to a wide range of phenomena ranging from physiology to finance. We carry out this study of statistical significance for both nonoverlapping [24] and overlapping [25] windows, and for both polynomial [24] and moving average detrending [8,10,26].

II. METHODS

Following Ref. [5], we consider two long-range cross-correlated time series $\{y_i\}$ and $\{y'_i\}$ of equal length N, and compute two integrated signals $R_k \equiv \sum_{i=1}^{k} y_i$ and $R'_k \equiv \sum_{i=1}^{k} y'_i$, where $k = 1,\ldots,N$. We divide the entire time series into $N-n$ overlapping windows [24], each containing $n+1$ values. For both time series, in each window that starts at i and ends at $i+n$, we define a "local trend," $\widetilde{R_{k,i}}$ [5]. We define the "detrended walk" as the difference between the original walk and the local trend and calculate the covariance of the residuals in each box:

$$f^2_{\rm DCCA}(n,i) \equiv 1/(n-1)\sum_{k=i}^{i+n}(R_k - \widetilde{R_{k,i}})(R'_k - \widetilde{R'_{k,i}}). \quad (1)$$

The detrended covariance is defined as

$$F_{\rm DCCA}(n) \equiv \sqrt{\frac{1}{N-n}\sum_{i=1}^{N-n} f^2_{\rm DCCA}(n,i)}. \quad (2)$$

If $R_k = R'_k$, we deal with one series instead of two, and so for this case the detrended covariance becomes the detrended variance—$F_{\rm DCCA}(n)$ reduces to the average root-mean-square fluctuation function $F_{\rm DFA}(n)$ of the detrended fluctuation analysis (DFA) [24,25,27]

$$F_{\rm DFA}(n) \equiv \sqrt{\frac{1}{N-n}\sum_{i=1}^{N-n} f^2_{\rm DFA}(n,i)}, \quad (3)$$

where $f^2_{\rm DFA}(n,i) \equiv 1/(n-1)\sum_{k=i}^{i+n}(R_k - \widetilde{R_{k,i}})^2$ [24,25,27]. For the $R_k = R'_k$ case, if a power law exists in the correlations, then the detrended variance versus n follows a power law, $F_{\rm DFA}(n) \propto n^{\alpha}$. On average, $F^2_{\rm DCCA}(n)$ versus n should be zero if there are no cross-correlations. However, for finite time series, if the detrended covariance versus n fluctuates around zero, there are no cross-correlations between $\{y_i\}$ and $\{y'_i\}$. If a power law exists in the cross-correlations, then the detrended covariance versus n follows a power law, $F_{\rm DCCA}(n) \propto n^{\lambda}$.

Analogous to the cross-correlations coefficient applied to stationary time series, Ref. [23] recently proposed a DCCA cross-correlation coefficient for nonstationary time series defined as the ratio between the detrended covariance function DCCA of Eq. (2) and two detrended variance functions of Eq. (3), one for each time series,

$$\rho_{\rm DCCA}(\alpha,\alpha',T,n) = \frac{F^2_{\rm DCCA}(n)}{F_{\rm DFA}(n)F'_{\rm DFA}(n)}, \quad (4)$$

which is dependent upon two time series of length T, each characterized by DFA exponents α and α', and window size n. Here $\rho_{\rm DCCA}$ is a dimensionless coefficient ranging between $-1 \leqslant \rho_{\rm DCCA} \leqslant 1$, where the result was not derived in Ref. [23], but was shown to hold using both artificial and empirical time series. Here we derive, first, that this Cauchy inequality, which holds for cross-correlations coefficients $-1 \leqslant \rho \leqslant 1$, also holds for a standard variance-covariance approach, and then for a detrending approach.

We compute two integrated signals $R_k \equiv \sum_{i=1}^{k} X_i$ and $R'_k \equiv \sum_{i=1}^{k} X'_i$, where $k = 1,\ldots,n$. We easily derive [5] $\langle R'_n R_n\rangle = nC_X(0) + 2\sum_{k=1}^{n-1}(n-k)C_X(k)$, where C_X stands for cross-covariance. Similarly we derive $\langle R_n^2\rangle = nC(0) + 2\sum_{k=1}^{n-1}(n-k)C(k)$ and $\langle R_n'^2\rangle = nC'(0) + 2\sum_{k=1}^{n-1}(n-k)C'(k)$, where $C(C')$ is the covariance of $X(X')$. Then we define

$$\frac{\langle R'_n R_n\rangle}{\sqrt{\langle R_n^2\rangle\langle R_n'^2\rangle}} = \frac{nC_X(0) + 2\sum_{k=1}^{n-1}(n-k)C_X(k)}{\sqrt{\left[nC(0) + 2\sum_{k=1}^{n-1}(n-k)C(k)\right]\left[nC'(0) + 2\sum_{k=1}^{n-1}(n-k)C'(k)\right]}}. \quad (5)$$

We assume for simplicity that X_i and X'_i both have the same variance and that each follows the same power law, i.e., $C(k) \propto C'(k) \propto k^{-\gamma}$. Then if power-law cross-correlations exist, cross-correlations follow the same power law $C_X(k) = C_X(1)k^{-\gamma}$:

$$\frac{\langle R'_n R_n\rangle}{\sqrt{\langle R_n^2\rangle\langle R_n'^2\rangle}} = \frac{nC_X(0) + 2\sum_{k=1}^{n-1}(n-k)C_X(k)}{nC(0) + 2\sum_{k=1}^{n-1}(n-k)C(k)} \approx \frac{\sum_{k=1}^{n-1}(n-k)C_X(k)}{\sum_{k=1}^{n-1}(n-k)C(k)} = \frac{C_X(1)}{C(1)}. \quad (6)$$

Since we assume that $\{X_i\}$ and $\{X'_i\}$ have an equal standard deviation, $V = V'$, when we employ that $C(1) < C(0) \equiv V$ (the covariance is a decreasing functional dependence), we obtain

$$\frac{\langle R'_n R_n\rangle}{\sqrt{\langle R_n^2\rangle\langle R_n'^2\rangle}} = \frac{C_X(1)}{C(1)} > \frac{C_X(1)}{C(0)} = \frac{C_X(1)}{\sqrt{VV'}} = \rho(1). \quad (7)$$

Generally, if a power law exists, we may assume that correlations do not change rapidly with lags, and so there is no big difference between correlations for lag zero and lag one. Thus, for $V \equiv C(0) \approx C(1)$, then $-1 \leqslant \rho(1) \leqslant 1$ implies

$$-1 \lesssim \frac{\langle R'_n R_n\rangle}{\sqrt{\langle R_n^2\rangle\langle R_n'^2\rangle}} \lesssim 1.$$

Next we show that similar result holds for the detrending approach of Eq. (4).

For a detrending approach with l nonoverlapping boxes each with n data points, where $T = ln$, we derive for Eq. (4),

$$\frac{F^2_{\rm DCCA}(n)}{F_{\rm DFA1}(n)F_{\rm DFA2}(n)} = \frac{1/(ln)\sum_{j=1}^{l}\sum_{i=1}^{n}\epsilon_{j,i}\epsilon'_{j,i}}{\sqrt{1/(ln)\sum_{j=1}^{l}\sum_{i=1}^{n}\epsilon^2_{j,i}\,1/(ln)\sum_{j=1}^{l}\sum_{i=1}^{n}\epsilon'^2_{j,i}}} = \frac{\sum_{k=1}^{T}\epsilon_k\epsilon'_k}{\sqrt{\sum_{k=1}^{T}\epsilon_k^2\sum_{k=1}^{T}\epsilon_k'^2}}, \quad (8)$$

where ϵ and ϵ' are error terms corresponding to time series X_t and X'_t, precisely [see Eq. (1)] $\epsilon_{k,i} = R_k - \widetilde{R_{k,i}}$ and $\epsilon'_{k,i} = R'_k - \widetilde{R'_{k,i}}$. Using the Cauchy inequality we obtain $-1 \leqslant \rho_{\rm DCCA}(n) \leqslant 1$. If two series are equal, $X_t = X'_t$, then we obtain $\rho_{\rm DCCA}(n) = 1$; if $X'_t = -X_t$, $\rho_{\rm DCCA}(n) = -1$.

A possible problem can arise, however, when one uses $\rho_{\rm DCCA}$ in practice. When there are no cross-correlations, $\rho_{\rm DCCA} = 0$, we can calculate only for an infinitely long time series. For finite time series, due to the size effect,

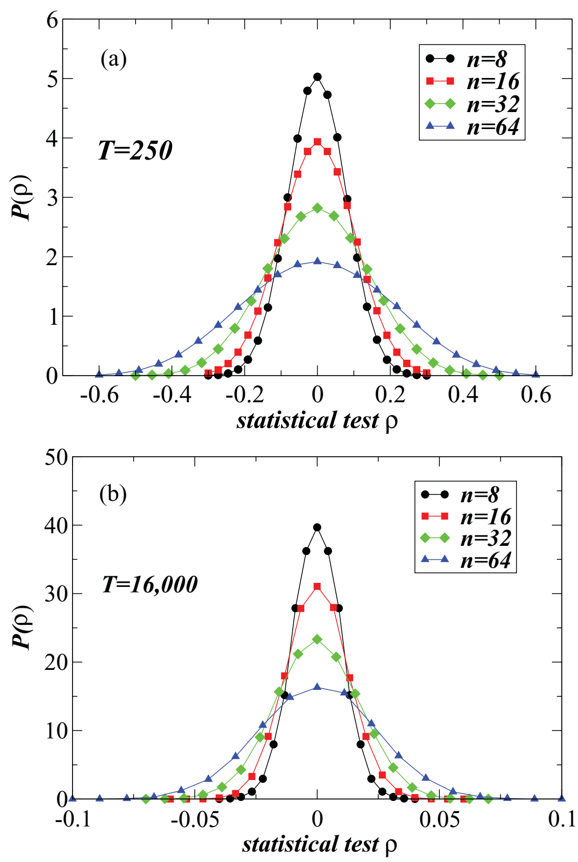

FIG. 1. (Color online) PDF of critical points $\rho_c(\alpha,\alpha',T,n)$ for the statistical test of Eq. (4), where $\alpha = \alpha' = 0.5$.

even if cross-correlations are not present, $\rho_{\rm DCCA}$ is not zero but presumably some small nonzero value. Thus the DCCA cross-correlation coefficient serves only as an indicator of the presence of cross-correlations. Clearly, if $\rho_{\rm DCCA}$ is either -1 or $+1$, cross-correlations can be considered genuine, but what about when $\rho_{\rm DCCA}$ is equal to 0.2 or 0.3? Do those values indicate the presence or absence of cross-correlations?

To test whether the cross-correlations are genuine (significant) or not, we use $\rho_{\rm DCCA}$ of Eq. (4) as the first statistical test [23]. To carry out this test we calculate critical points $\rho_c(\alpha,\alpha',T,n)$ for the 95% confidence level defined such that the integral between $-\rho_c(\alpha,\alpha',T,n)$ and $\rho_c(\alpha,\alpha',T,n)$ is equal to 0.95. We thus determine the range of $\rho_{\rm DCCA}$ within which the cross-correlations can be considered statistically significant. As is usual in statistics, we first determine the null hypothesis. Because this is not a unique choice, we begin by assuming that, under the null hypothesis, the time series are independent and identically distributed random variables (i.i.d.) and calculate the range of $\rho_{\rm DCCA}$ that can be obtained under the assumption that the time series are i.i.d., thus $\alpha = \alpha' = 0.5$.

In Fig. 1 for each of two different choices of time series length—ranging from short $T = 250$ (a) to long $T = 16\,000$ (b)—we calculate the probability distribution function (PDF) $P(\rho_{\rm DCCA})$ of the DCCA cross-correlation statistic (coefficient) $\rho_{\rm DCCA}$ of Eq. (4) for four different values of window size n. Each PDF is obtained by generating 10 000 i.i.d. time series pairs ($\alpha = \alpha' = 0.5$) taken from a Gaussian distribution, where for each time series pair we calculate the detrending variance DFA(n) and the detrending covariance DCCA(n), and then test it using Eq. (4). We first use a trend based on a first-order polynomial fit. We note that $P(\rho_{\rm DCCA})$ is symmetric, as expected, and that it depends on two parameters, the time series length T and the window size n. As expected, for each T, with increasing n, the PDF converges to a Gaussian due to the central limit theorem [28]. Due to an unknown form of PDF for smaller values of n, we calculate the critical values numerically. For each PDF $P(\rho_{\rm DCCA})$ defined by T and n, we calculate the critical point $\rho_c(\alpha = 0.5,\alpha' = 0.5,T,n)$ for the 95% confidence level. We report the critical values in Table I.

Next, using the detrending moving average method [26], for different choice of window size n and series size T, we generate many i.i.d. time series pairs (with no cross-correlations) taken from a Gaussian distribution with zero mean and unit variance. As expected, the critical values obtained based on the detrending moving average method in Table II are similar to those obtained using the polynomial trend in Table I.

In practice, we calculate $\rho_{\rm DCCA}$ from empirical time series and compare it with the critical point $\rho_c(T,n)$ for each T and n. If $\rho_{\rm DCCA} > \rho_c(T,n)$, the cross-correlations are considered statistically significant, and we reject the null hypothesis that $\rho_{\rm DCCA}$ of Eq. (4) comes from two Gaussian i.i.d. time series with no cross-correlations. It means that, for two series of length T with DFA exponents α and α' for every size of window n, $\rho_{\rm DCCA}$ must be larger than the critical value $\rho_c(\alpha,\alpha',T,n)$ calculated for T and n.

Table I shows the critical values obtained using the detrending approach with overlapping (sliding) boxes, a widely used method that allows us to obtain better statistics because the data points are finite [25]. For nonoverlapping boxes, we apply Eq. (8) and calculate the variance of

TABLE I. Polynomial fit of order 1. Critical values $\rho_c(\alpha,\alpha',T,n)$ for the DCCA cross-correlation coefficient of Eq. (4) when for a given couple of time series each series is Gaussian i.i.d. with zero mean and unit variance.

i.i.d.	$n = 4$	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$
$T = 250$	0.137	0.152	0.193	0.271	0.383		
$T = 500$	0.096	0.106	0.138	0.184	0.266	0.384	
$T = 1000$	0.070	0.077	0.097	0.132	0.185	0.261	0.377
$T = 2000$	0.049	0.055	0.068	0.093	0.131	0.186	0.269
$T = 4000$	0.034	0.038	0.049	0.067	0.093	0.132	0.185
$T = 8000$	0.024	0.028	0.035	0.046	0.063	0.091	0.129

TABLE II. Moving average fit. Critical values $\rho_c(\alpha,\alpha',T,n)$ for the DCCA cross-correlation coefficient of Eq. (4) when for a given couple of time series each series is Gaussian i.i.d. with zero mean and unit variance.

i.i.d.	$n=4$	$n=8$	$n=16$	$n=32$	$n=64$	$n=128$	$n=256$
$T=250$	0.129	0.156	0.212	0.305	0.448		
$T=500$	0.091	0.109	0.147	0.210	0.302	0.447	
$T=1000$	0.064	0.077	0.104	0.146	0.208	0.306	0.449
$T=2000$	0.045	0.054	0.073	0.102	0.145	0.209	0.304
$T=4000$	0.032	0.039	0.052	0.072	0.102	0.145	0.209
$T=8000$	0.023	0.027	0.037	0.051	0.072	0.102	0.146

$\rho_{\mathrm{DCCA}}(\alpha,\alpha',T,n)$:

$$E\left[\rho^2_{\mathrm{DCCA}}(\alpha,\alpha',T,n)\right] = \frac{\sum_{j=1}^{l}\sum_{i=1}^{n}\sum_{j'=1}^{l}\sum_{i'=1}^{n}E(\epsilon_{j,i}\epsilon_{j',i'})E(\epsilon'_{j,i}\epsilon'_{j',i'})}{\sum_{j=1}^{l}\sum_{i=1}^{n}E(\epsilon^2_{j,i})\sum_{j=1}^{l}\sum_{i=1}^{n}E(\epsilon'^2_{j,i})} = \frac{\sum_{j=1}^{l}\sum_{i=1}^{n}\sum_{j'=1}^{l}\sum_{i'=1}^{n}\delta_{i,i'}\delta_{j,j'}}{nlnl} = \frac{1}{nl} = \frac{1}{T}. \quad (9)$$

Thus we find that for nonoverlapping boxes, $E[\rho^2_{\mathrm{DCCA}}(\alpha,\alpha',T,n)]$ does not vary with box size n. This implies that even the critical values of $\rho_{\mathrm{DCCA}}(\alpha,\alpha',T,n)$ are not affected by n. When the boxes are overlapping, however, $E(\rho^2_{\mathrm{DCCA}}[\alpha,\alpha',T,n)]$ depends on n, and we calculate the critical values numerically. Each column in Table I shows that $\rho_c(\alpha,\alpha',T,n)$ versus T follows a power law $\propto T^{-0.5}$ with exponent -0.5.

In Fig. 1 we note that, for given T, $P(\rho_{\mathrm{DCCA}})$ becomes broader as window size n increases, implying that, for a given T and with increasing box size n, ρ_{DCCA} also increases so that it is larger than $\rho_c(\alpha,\alpha',T,n)$ when cross-correlations are present (see Table I). Further, for a given n, $P(\rho_{\mathrm{DCCA}})$ broadens with decreasing T.

For overlapping boxes, Fig. 2 shows the plot of critical points $\rho_c(T,n)$ for the statistical test of Eq. (4) versus window size n for varying values of series length T and $\alpha=\alpha'=0.5$.

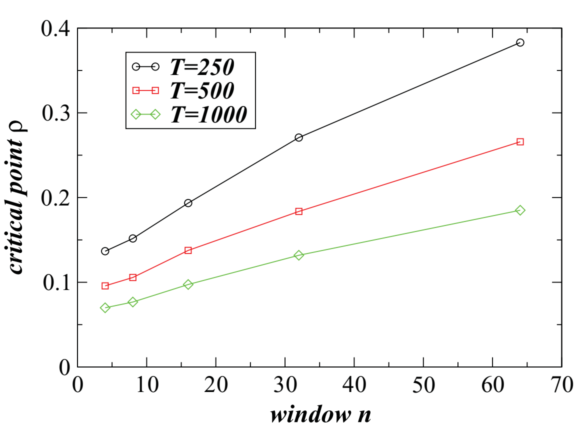

FIG. 2. (Color online) Scaling in critical points. Critical points $\rho_c(T,n)$ for the statistical test of Eq. (4) with the 95% confidence level. For varying values of series size T, we show $\rho_c(\alpha,\alpha',T,n)$ vs window size n.

Each curve represents a row of Table I. In practice, the lengths of the time series differ from those shown in Figs. 1 and 2, and the critical values for any T and n must be calculated by extrapolating from the critical values reported in Table I.

For overlapping boxes, Fig. 3(a) in a linear-linear plot shows the standard deviation of the statistical test $\rho_c(T,n)$ of Eq. (4) versus the length of time series T for different window sizes n. For each n, the standard deviation versus T scales with a power law $A_n T^{-0.5}$ with exponent -0.5,

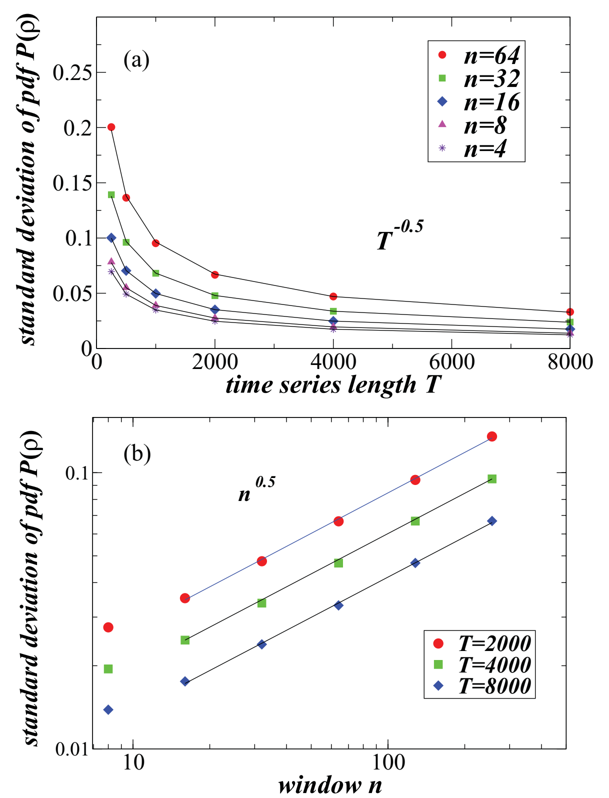

FIG. 3. (Color online) Normality in data. (a) Linear-linear plot. Standard deviation of the statistical test $\rho_c(T,n)$ of Eq. (4) vs series size T for different choice of window size n. For each n, standard deviation vs T scales with a characteristic exponent for a Gaussian PDF. (b) Log-log plot. Standard deviation of the statistical test $\rho_c(T,n)$ of Eq. (4) vs window size n for different choice of series size T. For each T, standard deviation vs n scales with a characteristic exponent for a Gaussian PDF.

characteristic of a Gaussian PDF, which is consistent with the scaling of the standard deviation of critical values versus n. The parameter A_n calculated for $n = 4$, 8, 16, 32, and 64 are 1.098, 1.215, 1.604, 2.269, and 3.456, respectively. Note that, when there are no correlations, the cross-correlation coefficients also follow a Gaussian in distribution with standard deviation $1/T$ [29]. We now show that for the critical value $\rho_c(T,n)$ of Eq. (4), the standard deviation also tends to $1/T$. Note that this is the expression for the standard deviation we have already derived for nonoverlapping boxes in Eq. (9). Figure 3(b) in log-log plot shows the standard deviation of the statistical test $\rho_c(T,n)$ of Eq. (4) versus window size n for different series size T. For each T, the standard deviation versus n scales with a characteristic exponent for a Gaussian PDF.

In order to test for the presence of significant power-law cross-correlations between two time series, we may hypothesize that each series must be a power-law correlated rather than i.i.d. For a time series length $T = 250$, we generate 10 000 time series pairs where each pair is power-law correlated with varying DFA exponent $\alpha = \alpha'$ and varying n. We assume that each of two power-law correlated series X_t and X'_t, with DFA exponent α and α', respectively, follows an autoregressive fractionally integrated moving-average (ARFIMA) process $x_t = \sum_{k=1}^{\infty} a_d(k)\epsilon_t$ with parameters d and d', respectively, where $a_d(k) \equiv -\frac{(k-d-1)!}{(-d-1)!k!}$ [30–33]—parameter d is related to the DFA exponent as $\alpha = 0.5 + d$ [34]. We calculate the critical point $\rho_c(\alpha,\alpha',T,n)$ for the 95% confidence level under the assumption that for a given pair of time series each series is power-law correlated with the same DFA exponent—$\alpha = \alpha'$, but there are no cross-correlations. We study the cases $\alpha = 0.6$, $\alpha = 0.7$, $\alpha = 0.8$, and $\alpha = 0.9$ and report the critical values in Table III. The critical values are shown for $T = 250$ and 1000. Note that, with decreasing correlations (α tending to 0.5, the case implying no correlations), the critical point $\rho_c(\alpha,\alpha,T,n)$ tends, as expected, to the value calculated for i.i.d. series taken from a Gaussian distribution with the same T and n. Generally we find that the critical point $\rho_c(\alpha,\alpha,T,n)$ increases with the increasing of both α and window size n. Again we see that if we have critical point information for each T and n when we calculate $\rho_{\mathrm{DCCA}}(\alpha,\alpha,T,n)$ from an empirical time series, the cross-correlations are statistically significant if $\rho_{\mathrm{DCCA}}(\alpha,\alpha,T,n)$ is larger than the critical value $\rho_c(\alpha,\alpha,T,n)$ calculated for α, T, and n.

TABLE III. Polynomial fit of order 1. Critical values $\rho_c(\alpha,\alpha',T,n)$ for the DCCA cross-correlation coefficient of Eq. (4) when for a given couple of time series each time series is power-law correlated with the same DFA exponent but not cross-correlated.

$T = 250$	$n = 4$	$n = 8$	$n = 16$	$n = 32$	$n = 64$
i.i.d.	0.137	0.152	0.193	0.271	0.383
$\alpha = 0.6$	0.137	0.158	0.203	0.295	0.406
$\alpha = 0.7$	0.137	0.161	0.221	0.313	0.431
$\alpha = 0.8$	0.137	0.172	0.234	0.329	0.443
$\alpha = 0.9$	0.137	0.178	0.244	0.346	0.464
$T = 1000$	$n = 4$	$n = 8$	$n = 16$	$n = 32$	$n = 64$
i.i.d.	0.070	0.077	0.097	0.132	0.182
$\alpha = 0.6$	0.070	0.078	0.105	0.148	0.223
$\alpha = 0.7$	0.070	0.080	0.113	0.154	0.238
$\alpha = 0.8$	0.070	0.082	0.123	0.165	0.247
$\alpha = 0.9$	0.070	0.087	0.128	0.171	0.260

III. EMPIRICAL APPLICATION

Finally in Fig. 4 we apply the $\rho(T,n)$ test in Eq. (4) to the interaction between a U.S. and a Chinese financial index, a U.S. and a German financial index, and between two U.S. financial indices. For the absolute values of a return series, we first calculate$\rho(T,n)$ versus window size n between the NYA (the New York Stock Exchange Composite index) and the SSEC (the Shanghai Stock Exchange Composite index). We also show the critical values $\rho_c(T,n)$ for the 95% confidence level where the area below $\rho_c(T,n)$ versus n implies no cross-correlations between indices. For daily recorded returns over the 11-year period from January 4, 2000, to August 15, 2011, we conclude that the NYA and the SSEC follow each other only very weakly, implying that the U.S. financial market does not have a strong effect on the Chinese financial market [35]. This result is not surprising and confirms that China has become an independent financial eigenvector, since it is only partially open to foreign investors. It is thus also not surprising that China was one of the very few countries that did not experience a severe recession during 2008–2009. In contrast, a U.S. "financial ally" is Germany, and the German index DAX strongly follows the NYA. As expected, the NYA and the NASDAQ Industrial index strongly affect each other since the curve of $\rho(T,n)$ versus n (open symbols) is well above $\rho_c(T,n)$ versus n for each n reported. Note that for each

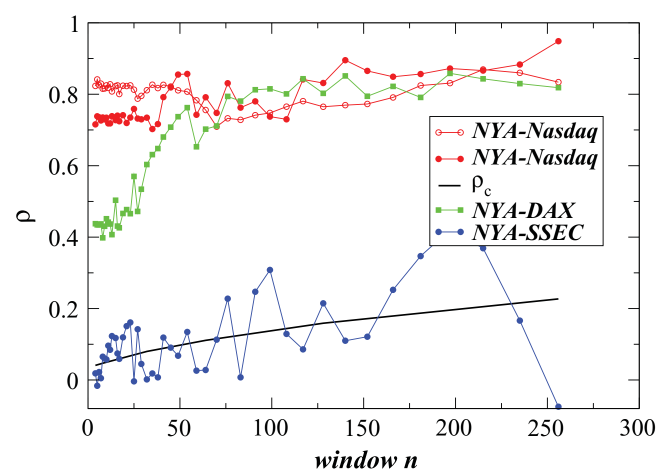

FIG. 4. (Color online) Statistical test $\rho(T,n)$ of Eq. (4) vs window size n calculated for absolute returns series between NYA and a Shanghai index SSEC, NYA, and a German index DAX, and between NYA and NASDAQ Industrial. Shown are also the critical values for the 95% confidence level under assumption of no cross-correlations. The area below $\rho_c(T,n)$ vs n means insignificant cross-correlations. Chinese index and U.S. index follow each other very weakly implying that U.S. financial market does not affect the Chinese financial market very strongly. We also show $\rho(T,n)$ of Eq. (4) vs window size n calculated for returns series between NYA and NASDAQ.

time series pair, $\rho(T,n)$ versus n has an increasing functional dependence, as shown in Ref. [23], that continues to hold when boxes in the detrending approach overlap.

Additionally, besides for the absolute values of return series (open symbols), for the NYA and the NASDAQ Industrial index in Fig. 4 we also show $\rho(T,n)$ versus n of Eq. (4) for the return series (closed symbols). The cross-correlations for return series are weaker than the cross-correlations for the absolute values of return series, but are also significant. It is worthy to note that the return series of financial indices are commonly uncorrelated (or short-range correlated). Here we calculate $F_{\rm DCCA}$ versus n of Eq. (2) and obtain exponent $\lambda = 0.485 \pm 0.02$ implying short-range cross-correlations among time series ($\approx$0.5) in agreement with short-range correlations. To understand this result we obtained for $\rho(T,n)$ versus n for return series, consider two series, X and X', that we define $X_t \equiv \epsilon_t$ and $X'_t \equiv \theta_t + b\epsilon_t$, where b is constant and ϵ_t and θ_t are two Gaussian i.i.d. processes with $\alpha = \alpha' = 0.5$. Then $\mathrm{cov}(X,X') = b\,\mathrm{var}(X)$, and so the square root of $\mathrm{cov}(X,X')$ scales with n as the standard deviation of X with scaling exponent $\lambda \approx 0.5$ since X is uncorrelated. Then, for X and X' previously defined, the $\rho(T,n)$ test in Eq. (4) scales for large n as

$$\lim_{n\to\infty} \frac{F^2_{\rm DCCA}(n)}{F_{\rm DFA}(n)F'_{\rm DFA}(n)} \propto \frac{n^{2\lambda}}{n^{\alpha}n^{\alpha'}} = \frac{n}{n^{0.5}n^{0.5}} = 1. \quad (10)$$

Next we show that the previous result about how the U.S. financial market affecting other financial markets depends on the period analyzed. In Fig. 5 we show the level of cross-correlations quantified by $F_{\rm DCCA}$ versus n of Eq. (2) between the NYA and the SSEC for each of subperiods, 2000–2005 and 2006–2011. As expected, NYA and SSEC have become more related during the second subperiod characterized by a severe recession during 2008–2009.

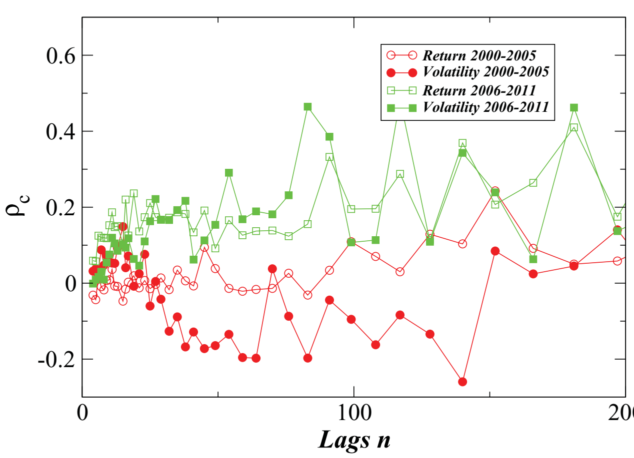

FIG. 5. (Color online) Statistical test $\rho(T,n)$ of Eq. (4) vs window size n between NYA and a Shanghai index SSEC, NYA calculated for two subperiods. We show both returns and absolute values of returns. Chinese index and U.S. index follow each other stronger during the second half of an 11-year period, characterized by a world recession and market crashes. However, the influence between these two indices is still much weaker than that between DAX and NYA (see Fig. 4).

In practice, when two series are short-range correlated or uncorrelated, and also short-range cross-correlated, we expect that, due to size effects, $\rho_{\rm DCCA}(\alpha,\alpha',T,n)$ will depend on scale n with very small scaling exponent. So, if series are short-range correlated or uncorrelated and cross-correlations are strong only for small number of lags, $\rho_{\rm DCCA}(\alpha,\alpha',T,n)$ versus n is virtually constant since $2\lambda - \alpha - \alpha'$ in Eq. (10) is small but nonzero. However, $\rho_{\rm DCCA}(\alpha,\alpha',T,n)$ versus n is also constant for two power-law correlated series with DFA exponents α and α' if *DCCA* cross-correlations exponent $\lambda = (\alpha+\alpha')/2$. Then, using Eq. (10), we obtain $n^{2\lambda}/(n^{\alpha}n^{\alpha'}) = 1$. Thus, the $\rho(T,n)$ test in Eq. (4) behaves equally for two significantly different cases. In practice, we must use Eq. (4) to quantify whether cross-correlations exist, but to find out whether cross-correlations are short-range ($\lambda \approx 0.5$) or long-range one needs to perform the test of Eq. (2).

IV. CONCLUSION

A number of empirical outputs in diverse phenomena ranging from geophysics to finance [2–4,36–38] exhibit cross-correlations with different levels of nonstationarity ranging from multifractality [8,10] to asymmetry. In order to quantify the statistical significance of cross-correlations we define two statistical cross-correlation tests based on the assumption that a series is either uncorrelated or power-law correlated. The tests quantify for which range of statistical tests the cross-correlations can be considered statistically significant. These tests may aid empirical investigations in a variety of phenomena in which cross-correlations exist.

ACKNOWLEDGMENTS

We acknowledge support from the Ministry of Science of Croatia (Grant No. 114-0352827-1370) and the National Science Foundation. ZQJ and WXZ acknowledge financial support from the National Natural Science Foundation of China (Grant No. 11075054) and the Fundamental Research Funds for the Central Universities.

APPENDIX

We also propose a second statistical test that can be used to quantify the existence of cross-correlations between two power-law correlated time series. We assume that each of two power-law correlated series X_t and X'_t, with DFA exponent α and α', respectively, and ARFIMA process $X_t = \sum_{k=1}^{\infty} a_d(k)\epsilon_t$ with parameters d and d', respectively [30–33]. Although there are many other processes that can lead to power-law correlations, we choose ARFIMA for which we can find analytical expressions. Using the expression for covariance of X_t equal to $\gamma_k = \gamma_{-k} = \rho^2(d)!(k-d-1)!/[(-d-1)!(k+d)!] \approx \rho^2(d)!/(-d-1)!)k^{-2d-1} = A\rho^2 k^{-2d-1}$ [30], the variance of

$$C_i = \frac{\sum_{k=i+1}^{T} X_k X'_{k-i}}{\sqrt{\sum_{k=1}^{T} X_k^2 \sum_{k=1}^{T} X'^2_k}}$$

becomes [where we use $E(C_i) = 0$ due to uncorrelations between X and X']

$$V(C_i) \equiv E\left(C_i^2\right) = \frac{\sum_{t=1}^{T-i}\sum_{t'=1}^{T-i} E(X_t X_{t'})E(X'_{t+i}X'_{t'+i})}{\rho^2\rho'^2T^2} = \frac{1}{T^2\rho^2\rho'^2}\left[(T-i)\rho^2\rho'^2 + \frac{\sum_{t=1}^{T-i}\sum_{t'\neq t=1}^{T-i} E(X_t X_{t'})E(X'_{t+i}X'_{t'+i})}{\rho^2\rho'^2T^2}\right], \quad \text{(A1)}$$

$$= \frac{(T-i)\rho^2\rho'^2 + 2[(T-i)\sum_{k=1}^{T-i-1}\gamma_k\gamma'_k - \sum_{k=1}^{T-i-1}k\gamma_k\gamma'_k]}{\rho^2\rho'^2T^2}, \quad \text{(A2)}$$

where by γ_k and γ'_k we denote the covariance of X_t and X'_t previously defined. Finally, replacing sums by integrals, we obtain

$$V(C_i) = \frac{T-i}{T^2}\left\{1 + \frac{2AA'}{2(d+d')+1}\left[1 - \frac{1+2(d+d')-(T-i)^{-2(d+d')}}{2(d+d')(T-i)}\right]\right\} = \frac{T-i}{T^2}[1+F(T,d,d')]. \quad \text{(A3)}$$

To test whether the cross-correlations between a power-law autocorrelated series are genuine (significant) or not, in analogy to the χ^2 distribution we propose the cross-correlation statistic with m degrees of freedom

$$Q_{\mathrm{DXA}}(d,d',T,m) \equiv T^2\sum_{i=1}^{m}\frac{C_i^2}{(T-i)[1+F(T,d,d')]}. \quad \text{(A4)}$$

Here the null hypothesis is that each of two time series is power-law correlated with DFA power-law exponents α and α'. It is assumed there are *no* cross-correlations. An alternative hypothesis is that the time series are not only power-law correlated, but also power-law cross-correlated. Our test is dependent upon the pair (α,α'), the time series length T, and m degrees of freedom. For a given (α,α',T,m) we generate 10 000 ARFIMA time series pairs with α,α' respectively, and calculate the critical point $Q_c(\alpha,\alpha',T,m)$ for the 95% confidence level. For a given empirical time series pair of length T, if $Q_{\mathrm{DXA}}(\alpha,\alpha',T,m) > Q_c(\alpha,\alpha',T,m)$, we conclude that the cross-correlations are genuine (significant). Since the test is based on many parameters, we do not provide the critical values in this paper.

Briefly, two power-law correlated series can be either mutually uncorrelated (statistically independent) or correlated where the latter case assumes either short-range or long-range cross-correlations. The test of Eq. (A4) should be used in order to test existence of power-law cross-correlations in case when each series is power-law correlated.

[1] P. Wochner *et al.*, Proc. Natl. Acad. Sci. USA **106**, 11511 (2009).
[2] M. Campillo and A. Paul, Science **299**, 547 (2003).
[3] L. Laloux, P. Cizeau, J. P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999); V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, *ibid.* **83**, 1471 (1999); Phys. Rev. E **66**, 066126 (2002).
[4] P. Samuelsson, E. V. Sukhorukov, and M. Buttiker, Phys. Rev. Lett. **91**, 157002 (2003); A. Cottet, W. Belzig, and C. Bruder, *ibid.* **92**, 206801 (2004); I. Neder, M. Heiblum, D. Mahalu, and V. Umansky, *ibid.* **98**, 036803 (2007)
[5] B. Podobnik and H. E. Stanley, Phys. Rev. Lett. **100**, 084102 (2008).
[6] B. Podobnik *et al.*, Proc. Natl. Acad. Sci. USA **106**, 22079 (2009).
[7] B. Podobnik *et al.*, Eur. Phys. J. B **71**, 243 (2009).
[8] W.-X. Zhou, Phys. Rev. E **77**, 066211 (2008).
[9] L.-Y. He and S.-P. Chen, Physica A **390**, 3806 (2011).
[10] Z.-Q. Jiang and W.-X. Zhou, Phys. Rev. E **84**, 016106 (2011).
[11] L. Hedayatifar, M. Vahabi, and G. R. Jafari, Phys. Rev. E **84**, 021138 (2011).
[12] E. L. Siqueira Jr., T. Stošić, L. Bejan, and B. Stošić, Physica A **389**, 2739 (2010).
[13] Y.-D. Wang, Y. Wei, and C.-F. Wu, Physica A **389**, 5468 (2010).
[14] Y.-D. Wang, Y. Wei, and C.-F. Wu, Physica A **390**, 864 (2011).
[15] L.-Y. He and S.-P. Chen, Physica A **390**, 297 (2011).
[16] L.-Y. He and S.-P. Chen, Chaos Solitons Fractals **44**, 355 (2011).
[17] N. Xu, P.-J. Shang, and S. Kamae, Nonlin. Dyn. **61**, 207 (2010).
[18] G. F. Zebende, P. A. da Silva, and A. M. Filho, Physica A **390**, 1677 (2011).
[19] X.-J. Zhao, P.-J. Shang, A.-J. Lin, and G. Chen, Physica A **390**, 3670 (2011).
[20] S. Shadkhoo and G. R. Jafari, Eur. Phys. J. B **72**, 679 (2009).
[21] S. Hajian and M. S. Movahed, Physica A **389**, 4942 (2010).
[22] D. Horvatic, H. E. Stanley, and B. Podobnik, Europhys. Lett. **94**, 18007 (2011)
[23] G. F. Zebende, Physica A **390**, 614 (2011).
[24] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Phys. Rev. E **49**, 1685 (1994).
[25] S. V. Buldyrev, A. L. Goldberger, S. Havlin, R. N. Mantegna, M. E. Matsa, C. K. Peng, M. Simons, and H. E. Stanley, Phys. Rev. E **51**, 5084 (1994).
[26] L. Xu, P. C. Ivanov, K. Hu, Z. Chen, A. Carbone, and H. E. Stanley, Phys. Rev. E **71**, 051101 (2005).
[27] Z. Chen, P. C. Ivanov, K. Hu, and H. E. Stanley, Phys. Rev. E **65**, 041107 (2002).
[28] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications* (Springer-Verlag, New York, 2000).
[29] R. L. Anderson, Ann. Math. Stat. **13**, 1 (1942).
[30] J. Hosking, Biometrika **68**, 165 (1981).
[31] C. W. J. Granger, J. Econometrics **14**, 227 (1980).

[32] C. W. J. Granger and R. Joyeux, J. Time Ser. Anal. **1**, 15 (1980).

[33] Power-law autocorrelated and power-law cross-correlated time series were generated using the ARFIMA process [30-32,34], where each variable depends not only on its own past, but also on the past values of the other variable, $y_i = [W \sum_{n=1}^{\infty} a_n(\rho_2) y'_{i-n} + (1-W) \sum_{n=1}^{\infty} a_n(\rho_1) y_{i-n}] + \eta_i$, $y'_i = [(1-W) \sum_{n=1}^{\infty} a_n(\rho_2) y_{i-n} + W \sum_{n=1}^{\infty} a_n(\rho_1) y'_{i-n}] + \eta'_i$. Here η_t and η'_t denote two i.i.d. Gaussian variables with zero mean and unit variance [30-32,34], ρ_m (for $m = 1,2$) are parameters ranging from 0 to 0.5, and W is a free parameter ranging from 0 to 0.5 and controls the link between y_t and y'_t. In the case where $W = 0$, cross-correlations vanish, and the system of two equations decouples on two separate ARFIMA processes.

[34] B. Podobnik *et al.*, Phys. Rev. E **72**, 026121 (2005).

[35] W.-X. Zhou and D. Sornette, Physica A **330**, 543 (2003).

[36] T. Guhr and B. Kalber, J. Phys. A **36**, 3009 (2003).

[37] S. Arianos and A. Carbone, J. Stat. Mech. (2009) P03037.

[38] B. Podobnik *et al.*, Europhys. Lett. **90**, 68001 (2010).

[39] A. Carbone, G. Castelli, and H. E. Stanley, Phys. Rev. E **69**, 026105 (2004); A. Carbone *et al.*, Physica A **344**, 267 (2004).

[40] J. W. Kantelhardt *et al.*, Physica A **295**, 441 (2001).