EPJ B
Condensed Matter and Complex Systems
EPJ.org your physics journal

Eur. Phys. J. B **71**, 243–250 (2009) DOI: 10.1140/epjb/e2009-00310-5

Quantifying cross-correlations using local and global detrending approaches

B. Podobnik, I. Grosse, D. Horvatić, S. Ilic, P.Ch. Ivanov and H.E. Stanley

EDP SCIENCES
Società Italiana di Fisica
Springer

Eur. Phys. J. B **71**, 243–250 (2009)
DOI: 10.1140/epjb/e2009-00310-5

THE EUROPEAN
PHYSICAL JOURNAL B

Regular Article

Quantifying cross-correlations using local and global detrending approaches

B. Podobnik[1,2,3,a], I. Grosse[4], D. Horvatić[5], S. Ilic[6], P.Ch. Ivanov[3,7,8], and H.E. Stanley[3]

[1] Faculty of Civil Engineering, University of Rijeka, Rijeka, Croatia
[2] Zagreb School of Economics and Management, Zagreb, Croatia
[3] Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA
[4] Institute of Computer Science, Martin Luther University, Halle, Germany
[5] Physics Department, Faculty of Science, University of Zagreb, Bijenička c. 32, 10000 Zagreb, Croatia
[6] Lancaster Environment Centre, Lancaster University, Farrer Avenue, Lancaster LA1 4YQ, UK
[7] Institute of Solid State Physics, Bulgarian Academy of Sciences, Sofia 1784, Bulgaria
[8] Harvard Medical School and Division of Sleep Medicine, Brigham and Women's Hospital, Boston, MA 02115 USA

Received 18 May 2009 / Received in final form 27 July 2009
Published online 17 September 2009 –

Abstract. In order to quantify the long-range *cross-correlations* between two time series qualitatively, we introduce a new cross-correlations test $Q_{\rm CC}(m)$, where m is the number of degrees of freedom. If there are no cross-correlations between two time series, the cross-correlation test agrees well with the $\chi^2(m)$ distribution. If the cross-correlations test exceeds the critical value of the $\chi^2(m)$ distribution, then we say that the cross-correlations are significant. We show that if a Fourier phase-randomization procedure is carried out on a power-law cross-correlated time series, the cross-correlations test is substantially reduced compared to the case before Fourier phase randomization. We also study the effect of periodic trends on systems with power-law cross-correlations. We find that periodic trends can severely affect the quantitative analysis of long-range correlations, leading to crossovers and other spurious deviations from power laws, implying both *local* and *global* detrending approaches should be applied to properly uncover long-range power-law auto-correlations and cross-correlations in the random part of the underlying stochastic process.

PACS. 05.45.Tp Time series analysis – 05.40.-a Fluctuation phenomena, random processes, noise, and Brownian motion

There are a number of situations where different signals exhibit cross-correlations, ranging from geophysics [1] to finance [2–14] and solid-state physics [15]. Cross-correlation functions together with auto-correlation functions are commonly used to gain insight into the dynamics of natural systems. By their definitions, these techniques should be employed *only* in the presence of stationarity. However, it is an important fact that many time series of physical, biological, hydrological, and social systems are non-stationary and exhibit long-range power-law correlations [16–22]. In practice, statistical properties of these systems are difficult to study due to these nonstationarities.

For determining the scaling exponent of a long-range power-law auto-correlated time series in the presence of nonstationarities, the *detrended fluctuation analysis* (DFA) method has been developed [23] and its performance has been systematically tested for the effect of different types of trends and nonstationarities [24–27] as encountered in a wide range of different fields, such as cardiac dynamics [28], economics [29], DNA analysis [30], and meteorology [31]. The square root of the detrended variance grows with time scale n as $F_{\rm DFA}(n) \sim n^{\lambda_{DFA}}$, where λ_{DFA} is the DFA scaling exponent [23–26], where $1/2 < \lambda_{DFA} < 1$, indicates the presence of power-law auto-correlations, and $0 < \lambda_{DFA} < 1/2$ indicates the presence of long-range power-law anti-correlations.

There are many realistic situations in which one desires to quantify cross-correlations between two non-stationary time series. Examples include blood pressure and heart rate [32], air temperature and air humidity, and the temporal expression data of different genes. To quantify power-law *cross-correlations* in non-stationary time series, a new method based on detrended covariance, called detrended cross-correlations analysis (DCCA), has been recently proposed [11]. If cross-correlations decay as a power law, the corresponding detrended covariances are either always positive or always negative, and the square root of the detrended covariance grows with time scale n as

$$F_{\rm DCCA}(n) \propto n^{\lambda_{\rm DCCA}}, \quad (1)$$

[a] e-mail: bp@phy.hr

where λ_{DCCA} is the DCCA cross-correlation exponent. If, however, the detrended covariance oscillates around zero as a function of the time scale n, there are no long-range cross-correlations.

In order to investigate power-law auto-correlations and power-law cross-correlations and effects of sinusoidal periodicity on cross-correlations, first we define a periodic two-component fractionally autoregressive integrated moving-average (ARFIMA) process [33–37], where each variable depends not only on its own past, but also on the past values of the other variable,

$$y_i = \left[W\sum_{n=1}^{\infty} a_n(\rho_1)y_{i-n} + (1-W)\sum_{n=1}^{\infty} a_n(\rho_2)y'_{i-n}\right] + A_1\sin\left(\frac{2\pi}{T}i\right) + \eta_i, \tag{2a}$$

$$y'_i = \left[(1-W)\sum_{n=1}^{\infty} a_n(\rho_1)y_{i-n} + W\sum_{n=1}^{\infty} a_n(\rho_2)y'_{i-n}\right] + A_2\sin\left(\frac{2\pi}{T}i\right) + \eta'_i. \tag{2b}$$

Here, η_t and η'_t denote two independent and identically distributed (*i.i.d.*) Gaussian variables with zero mean and unit variance, $a_j(\rho_m)$ are statistical weights defined by $(\rho_m) \equiv \frac{\Gamma(j-\rho_m)}{\Gamma(-\rho_m)\Gamma(1+j)}$, where $\Gamma(x)$ denotes the Gamma function, ρ_m (for $m = 1, 2$) are parameters ranging from 0 to 0.5, T is the sinusoidal period, A_1 and A_2 are two sinusoidal amplitudes, and W is a free parameter ranging from 0.5 to 1 and controlling the strength of power-law cross-correlations between y_t and y'_t. In case of $A_1 = A_2 = 0$, for $W = 1$, cross-correlations vanish, and the system of two equations decouples to two separate ARFIMA processes.

In Appendix A, for a version of the above process $y_i \equiv \sum_{j=1}^{\infty} a_j(\rho_1)y_{i-j}+\eta_i$, $y'_i \equiv \sum_{j=1}^{\infty} a_j(\rho_2)y'_{i-j}+\eta_i$ [11] where both y_i and y'_i share the *same i.i.d.* Gaussian process η_i, we analytically find that the time series $\{y_i\}$ and $\{y'_i\}$ are long-range power-law cross-correlated, where the scaling cross-correlations exponent λ_{DCCA} (Eq. (1)) is equal to the average of the Hurst exponents, $\lambda_{\mathrm{DCCA}} = \frac{H_1+H_2}{2}$, the result found numerically in reference [11], and where $H_m = 0.5 + \rho_m$ [37].

Statistical inferences based on estimation and hypothesis testing are among the most important aspects of the decision making process in science and business. Here we propose a new statistic to test the presence of cross-correlations. Suppose that $\{y_i\}$ and $\{y'_i\}$ are two discrete-time *i.i.d.* stochastic processes, where there are no cross-correlations among the time series. We may define their cross-correlation function

$$X_i = \frac{\sum_{k=i+1}^{N} y_k y'_{k-i}}{\sqrt{\sum_{k=1}^{N} y_k^2 \sum_{k=1}^{N} y_k'^2}}. \tag{3}$$

Under the assumption that $\{y_i\}$ and $\{y'_i\}$ are statistically independent, one can easily show that the X_i are uncorrelated [38]:

$$E(X_iX_{i'}) \propto \sum_{k=i+1}^{N}\sum_{k'=i'+1}^{N} E(y_k y'_{k-i} y_{k'} y'_{k'-i'}) = \sum_{k=i+1}^{N}\sum_{k'=i'+1}^{N} E(y_k y_{k'})E(y'_{k-i} y'_{k'-i'}), \tag{4}$$

which is zero for $i \neq i'$. The expectation value of X_i is equal to zero, $E(X_i) = 0$, because there are no cross-correlations between $\{y_i\}$ and $\{y'_i\}$, and the variance is

$$V(X_i) = E(X_i^2) = \frac{\sum_{k=i+1}^{N}\sum_{k'=i+1}^{N} E(y_k y'_{k-i} y_{k'} y'_{k'-i})}{\sum_{k=1}^{N}\sum_{k'=1}^{N} E(y_k^2)E(y_{k'}'^2)} = \frac{\sum_{k=i+1}^{N}\sum_{k'=i+1}^{N} E(y_k y_{k'})E(y'_{k-i} y'_{k'-i})}{\sigma^2\sigma'^2 N^2}, \tag{5}$$

where we use $E(y_k y_{k'}) = \sigma^2\delta_{k,k'}$ and $E(y'_k y'_{k'}) = \sigma'^2\delta_{k,k'}$. Further, $V(X_i) = \frac{\sum_{k=i+1}^{N}\sum_{k'=i+1}^{N}\delta_{k,k'}\delta_{k-i,k'-i}}{N^2} = \frac{\sum_{k=i+1}^{N}\delta_{k,k}}{N^2} = \frac{N-i}{N^2}$. Thus, we find that $E(X_i^2) = \frac{N-i}{N^2}$, where $E(X_iX_{i'}) = 0$ when $i \neq i'$. The cross-correlation coefficient X_k is normally distributed for asymptotically large values of N [38], as it holds for auto-correlation function r_k [39]. Then $X_i/\sqrt{(N-i)/N^2}$ asymptotically behaves as a Gaussian distribution with zero mean and unit variance, and the sum of squares of these variables approximately follows a χ^2 distribution.

According to definition of the χ^2 distribution, we propose the cross-correlations statistic

$$Q_{\mathrm{CC}}(m) \equiv N^2 \sum_{i=1}^{m} \frac{X_i^2}{N-i}, \tag{6}$$

which is approximately $\chi^2(m)$ distributed with m degrees of freedom. The test can be used to test the null hypothesis that none of the first m cross-correlation coefficients is different from zero. The test of equation (6) is similar to the test statistic [40]

$$Q'(m) \equiv N(N+2) \sum_{i=1}^{m} \frac{X_i^2}{N-i} \tag{7}$$

proposed in analogy to the Ljung-Box (LJB) test [41] that is one of the most widely employed tests for the presence of auto-correlations. The Ljung-Box (LJB) test can be easily obtained if all cross-correlation coefficients X_k in equation (7) are replaced by auto-correlation coefficients. Clearly, for larger samples where $N(N+2) \approx N^2$, the tests of equations (6) and (7) give the same distribution.

Next we show that the test of equation (6) better approximates the $\chi^2(m)$ distribution than the test of equation (7) for small samples. In order to show that the cross-correlation test of equation (6) is applicable to real-world data where time series are commonly of small size, we test the speed of convergence of the distribution of $Q_{\mathrm{CC}}(m)$ to the $\chi^2(m)$ distribution. Thus, we first generate 10^6

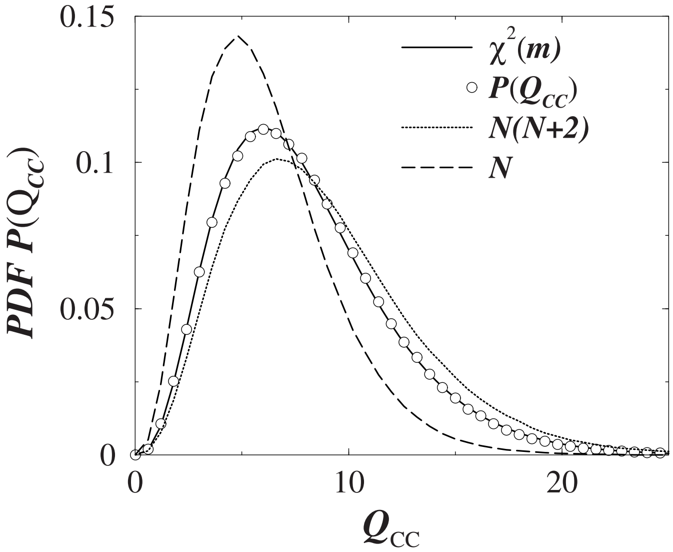

Fig. 1. Probability distribution function (pdf) $P(Q_{\rm CC})$ of $Q_{\rm CC}$ defined in equation (6) together with the $\chi^2(m)$ pdf, where m are the degrees of freedom. We also show $P(Q')$ of Q' defined in equation (7). We generate 10^6 pairs of time series $\{y_i\}$ and $\{y'_i\}$, where each time series is generated by an *i.i.d.* Gaussian process with mean zero and unit variance. Each time series is comprised of $N = 20$ data points (small time series). We choose $m = 8$, and for each pair of time series we calculate the cross-correlations X_i, where $i = 1, ..., 8$. We find a perfect match between $P(Q_{\rm CC}(m))$ and $\chi^2(m)$. In opposite, the pdf of Q' defined in equation (7) deviates from the $\chi^2(m)$ distribution. We also show the pdf of a test defined as $N\sum_i^m X_i^2$, which substantially deviates from $\chi^2(m)$.

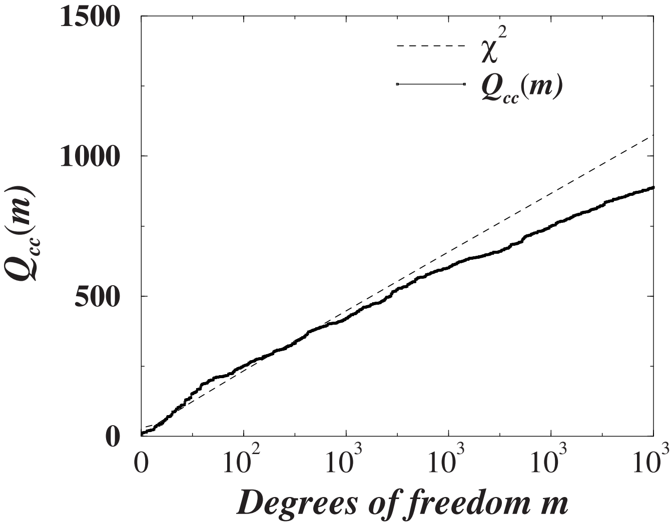

Fig. 2. Cross-correlations between the differences of logarithms of prices for IBM and General Electric (GE). We show $Q_{\rm CC}(m)$ versus the degrees of freedom m. We also show the critical values for the $\chi^2(m)$ distribution at the 5% level of significance. $Q_{\rm CC}(m)$ virtually follows the critical values for the $\chi^2(m)$ distribution that is a signal for no cross-correlations. The test of equation (7) gives practically the same result as the test of equation (6) since $N \gg 1$.

equally-sized *i.i.d.* time series $\{y_i\}$ and $\{y'_i\}$ for a small value $N = 20$ where $m = 8$. For each pair of time series, we calculate the cross-correlations X_k, where $k = 1, ..., 8$, and then the test statistic $Q_{\rm CC}(m)$. In Figure 1 we show the distribution of $P(Q_{\rm CC}(m))$ together with $\chi^2(8)$, and find a perfect agreement between these two probability distributions. We also show the distribution of the test statistic of equation (7), where for a given small sample ($N = 20$), deviation between the given distribution and $\chi^2(8)$ is obvious. Thus, when the cross-correlation test is applied in practice, we can use the critical values of the $\chi^2(m)$ distribution.

Note that the LJB test and hence the cross-correlation test of equation (6) is proposed to be applied for the *residuals* of a given model, not the original time series. However, sometimes the test is applied to the original series, e.g., return time series [42]. Accordingly, the cross-correlation test of equation (6) can be also used to measure the strength of cross-correlations in the original time series. In order to investigate cross-correlation scaling we analyze the daily adjusted closing values of the IBM and General Electric [43]. For each company's price, we calculate the time series of the differences of logarithms for successive days over the period 2 January 1962 till 1 May 2009. Then we calculate the P value of the cross-correlation test of equation (6) for different degrees of freedom m together with the critical values for the $\chi^2(m)$ distribution at the 5% level of significance. In Figure 2 we find that the cross-correlation Q_{CC} test statistic practically follows the critical values for the $\chi^2(m)$, suggesting no cross-correlations in the data.

In opposite to a common practice in statistic when a test statistic is compared with a critical value for a single value of degree of freedom m, here in the paper we plot the statistic test versus the critical value of $\chi^2(m)$ for a broad range of values of m. If for a broad range of m the test statistic of equation (6) exceeds the critical values of $\chi^2(m)$ ($Q_{\rm CC}(m) > \chi^2_{0.95}(m)$), we claim that there are not only cross-correlations, but there are long-range cross-correlations. However, the cross-correlations test of equation (6) should be used to test the presence of cross-correlations only *qualitatively*. In order to test the presence of cross-correlations *quantitatively* – to estimate the cross-correlation exponent – we suggest to employ the DCCA method of equation (1).

Next we show how the cross-correlation test of equation (6) might be useful to estimate the strength and significance of cross-correlations found in data. By using the two-component ARFIMA process of equations (2a, 2b), we generate four different pairs of time series $\{y_i\}$ and $\{y'_i\}$, where each pair is characterized by different values of ρ, while W is constant. We exclude the periodic term for now, so $A_1 = A_2 = 0$. In Figure 3 for each pair of time series we show the cross-correlation test of equation (6) (filled symbols) for different degrees of freedom m. In order to show the strength of cross-correlations, for different values of m, we also show the critical values of the $\chi^2(m)$ distribution at the 5% level of significance. We note that, for a given value of m, the deviation between the test of

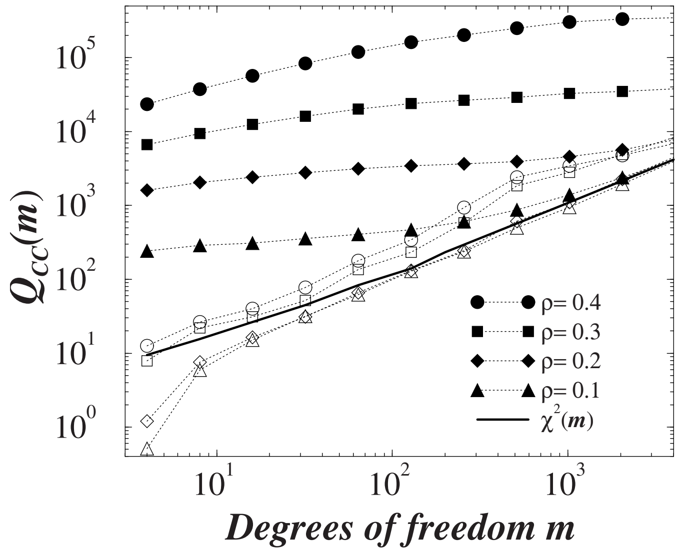

Fig. 3. $Q_{\rm CC}(m)$ for different degrees of freedom m before (filled symbols) and after (open symbols) Fourier phase-randomization of the $Q_{\rm CC}(m)$. For each of four values of ρ, the two-component ARFIMA process of equations (2a, 2b) generates the pair of time series $\{y_i\}$ and $\{y'_i\}$ where $m = 2^i$, and $i = 2, ..., 8$, and fixed $W = 0.5$. For the sinusoidal amplitude we take $A_1 = A_2 = 0$. The time series $\{y_i\}$ and $\{y'_i\}$ are $N = 5 \times 10^4$ data points each. The solid line denotes the critical values for the $\chi^2(m)$ distribution at the 5% level of significance. For each of four pairs $(\{y_i\}, \{y'_i\})$, we calculate the cross-correlations X_i, and the $Q_{\rm CC}(m)$ test statistic. The more positive is the difference between the $Q_{\rm CC}(m)$ test and the critical value of the $\chi^2(m)$ distribution, the stronger are the cross-correlations for a given m. For each value of ρ, we phase randomize the original time series $\{y'_i\}$, and obtain the surrogate time series $\{\tilde{y}'_i\}$. Then for each pair of time series $(\{y_i\}, \{\tilde{y}'_i\})$, we calculate the $Q_{\rm CC}(m)$ test statistic of equation (6). We also show the critical values for $\chi^2(m)$ distribution at the 5% level of significance. Fourier phase-randomization reduces the linear cross-correlations, since after a Fourier phase-randomization procedure (open symbols), we find that for each pair $(\{y_i\}, \{\tilde{y}'_i\})$, the cross-correlations measured by the $Q_{\rm CC}(m)$ test are substantially reduced – in fact, the cross-correlations practically vanish for time series with smaller ρ values.

equation (6) and the critical value of $\chi^2(m)$ increases with ρ, if W is kept fixed. We also note that, for each time series (specified by ρ), the test of equation (6) is larger than the critical value of $\chi^2(m)$ for a broad, but finite range of m. We propose that, if for a broad range of values of m the values of the test of equation (6) between the two time series are larger than the critical values of the $\chi^2(m)$ distribution, the cross-correlations are considered significant.

Often it is unclear to what degree the time series generated by a stochastic process exhibits linear and nonlinear correlations. Linear (nonlinear) auto-correlations are defined as those correlations which are not destroyed (are destroyed) by a Fourier phase-randomization of the original time series [28,44,45]. The Fourier phase-randomization procedure [44] works as follows: (i) perform a Fourier transform of the original time series; (ii) randomize the Fourier phases (thereby eliminating the nonlinearities of the original time series) but keep the Fourier amplitudes unchanged (thereby preserving the power spectrum and the linear properties of the original time series); and (iii) perform an inverse Fourier transform to obtain a surrogate time series.

For the four pairs of time series $\{y_i\}$ and $\{y'_i\}$ of equations (2a, 2b), in Figure 3 we show the cross-correlations test for different degrees of freedom after (open symbols) performing Fourier phase-randomization. To emphasize the impact of a phase randomization on cross-correlations we also show the critical values of $\chi^2(m)$ for different degrees of freedom. We show that after a Fourier phase-randomization (open symbols) cross-correlations are reduced [10] – for each pair of time series and for each m, the test is substantially reduced compared to the case before the Fourier phase randomization. Thus, while the Fourier phase-randomization procedure preserves linear auto-correlations [28,44], the same method substantially reduces the linear cross-correlations.

We next discuss how to quantify the scaling exponent of power-law cross-correlations between two time series, and how it relates to the DFA exponents calculated for each of two cross-correlated time series, which we generate by using the two-component process of equations (2a, 2b). Here, we assume there are no sinusoidal trends, $A_1 = A_2 = 0$. In Figures 4a, 4b, the DFA functions are given for each time series $\{y_i\}$ and $\{y'_i\}$ of 10^5 data points and $\rho_1 = 0.4$ and $\rho_2 = 0.1$. We set the cross-correlation coupling parameter to $W = 0.95$ (Fig. 4a) and $W = 0.05$ (Fig. 4b). In each figure we show that both time series $\{y_i\}$ and $\{y'_i\}$ are power-law auto-correlated, and are also power-law cross-correlated. From the definition of the process of equations (2a, 2b) it is clear that with decreasing value of W (from 1 to 0.05), each of the two processes y_i and y'_i becomes a mixture of two ARFIMA processes. Particularly, for the process y'_i, the DFA correlation exponent λ_{DFA} virtually does not change with varying the parameter $W - \lambda_{DFA} \approx 0.6 = 1/2 + \rho_2$ [46]. In contrast, for the process y_i, the DFA correlation exponent λ_{DFA} gradually decreases from $\lambda_{DFA} \approx 0.9 = 1/2 + \rho_1$ (when $W = 1$, not shown) toward $\lambda_{DFA} \approx 0.6$ (when $W = 1/2$) corresponding to the y'_i process [46].

Next we focus on the DCCA cross-correlation exponent $\lambda_{\rm DCCA}$. We show in Figures 4a, 4b that, by varying the cross-correlation coupling parameter W, $\lambda_{\rm DCCA}$ follows the DFA exponent λ_{DFA} corresponding to the y_i process. By decreasing the value of W from $W = 1$ to $W = 0.5$, both the DFA correlation exponent λ_{DFA} corresponding to the y_i process and the DCCA cross-correlation exponent $\lambda_{\rm DCCA}$ gradually decrease toward $\lambda_{DFA} \approx 0.6$. Generally, for different time series of the process with parameters ρ_1 and ρ_2, where $\rho_1 > \rho_2$, we find that $\lambda_{\rm DCCA}$ is closer to the DFA exponent λ_{DFA} corresponding to the y_i process (larger ρ).

A necessary condition for power-law cross-correlations with a unique power-law exponent is that $F_{\rm DCCA}(n)$

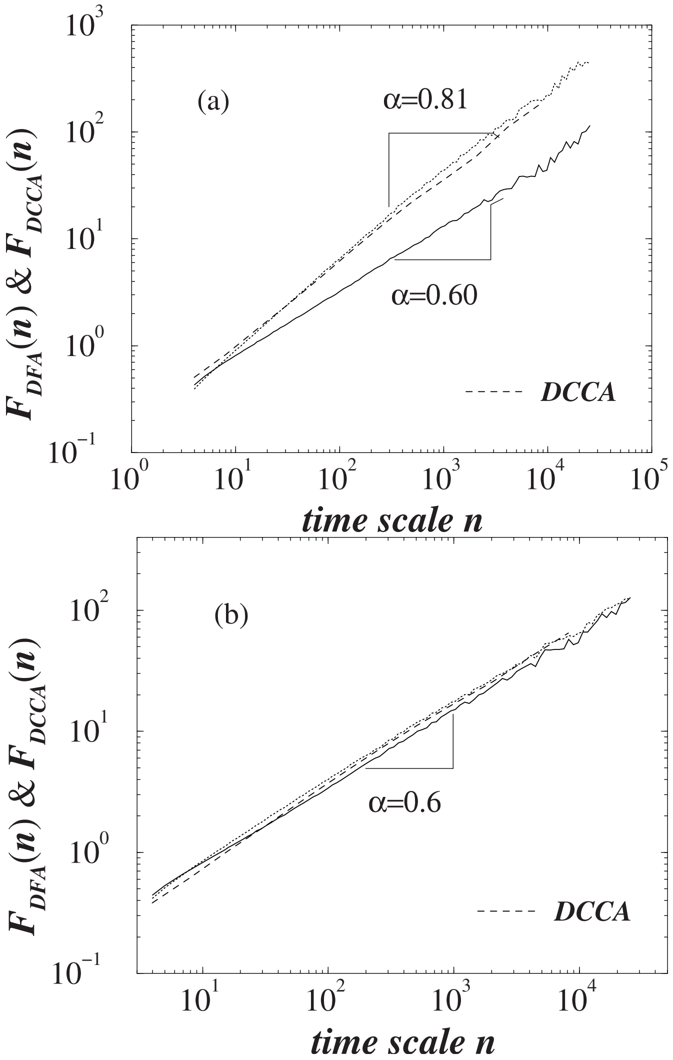

Fig. 4. DFA and DCCA scaling functions $F_{\mathrm{DFA}}(n)$ and $F_{\mathrm{DCCA}}(n)$, respectively, versus time scale n. We generate the time series $\{y_i\}$ and $\{y'_i\}$ defined by the two-component process of equations (2a, 2b) with $\rho_1 = 0.4$ and $\rho_2 = 0.1$. We exclude the sinusoidal amplitude, so $A_1 = A_2 = 0$. We show the two DFA functions, $F_{\mathrm{DFA}}(n) \propto n^{\lambda_{DFA}}$, and the DCCA function, $F_{DCCA}(n) \propto n^{\lambda_{\mathrm{DCCA}}}$, for (a) $W = 0.95$ and (b) $W = 0.5$. The closer W is to 0.5, the more the two processes y_i and y'_i become alike. λ_{DCCA} gradually decreases toward $\lambda_{DFA} \approx 0.6$. Generally, by varying W, λ_{DCCA} becomes closer to λ_{DFA} corresponding to y_i, but eventually the λ_{DFA} value corresponding to y_i tends to the λ_{DFA} value corresponding to y'_i.

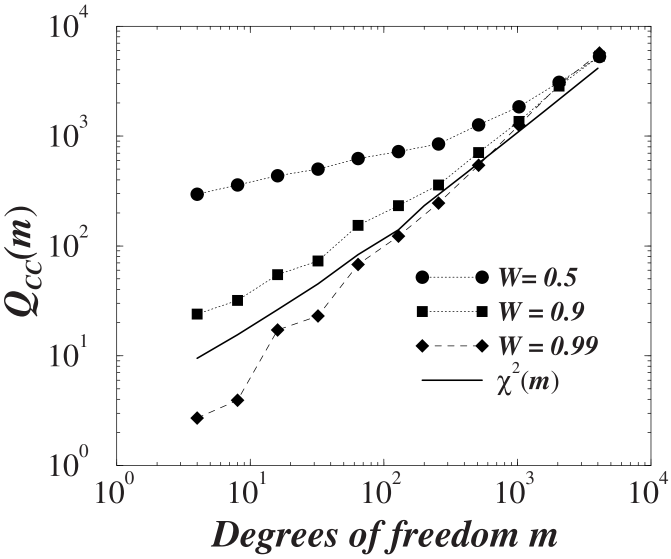

Fig. 5. $Q_{\mathrm{CC}}(m)$ versus the number of degrees of freedom m for different values of the cross-correlation coupling W. For each W, we generate one pair of time series $\{y_i\}$ and $\{y'_i\}$ defined by equations (2a, 2b). Each time series is comprised of $N = 10^4$ data points. There is no sinusoidal trend, so $A_1 = A_2 = 0$. We also show the curve of the critical values of $\chi^2(m)$ distribution at the 5% level of significance. Parameters are $\rho_1 = 0.2$ and $\rho_2 = 0.4$. For $W = 0.5$, the cross-correlations between $\{y_i\}$ and $\{y'_i\}$ are strongest and, for a wide range of m values, the curve of the test statistic of equation (6) for $W = 0.5$ is above all other curves including the curve of the critical values of $\chi^2(m)$ distribution. For values of W very close to 1, the cross-correlations between $\{y_i\}$ and $\{y'_i\}$ become very weak, below the curve of the critical values of $\chi^2(m)$ distribution.

does not change sign with increasing n, i.e. $F_{\mathrm{DCCA}}(n) = An^{\lambda_{\mathrm{DCCA}}}$ where A is constant. To this end, we find that the process of equations (2a, 2b) with $W = 0.99$ generates two particular time series $\{y_i\}$ and $\{y'_i\}$ where $F_{\mathrm{DCCA}}(n)$ versus n starts to oscillate, indicating the loss of a unique power-law dependence. Thus, even though there are cross-correlations between $\{y_i\}$ and $\{y'_i\}$, the cross-correlations are weak because the cross-correlations coupling parameter W is close to 1. For the limiting case $W = 1$, the processes y_i and y'_i are decoupled, and thus not cross-correlated, and each of two processes y_i and y'_i becomes a separate ARFIMA process controlled by parameters ρ_1 and ρ_2, respectively.

Next we analyze the cross-correlation tests between time series $\{y_i\}$ and $\{y'_i\}$, with varying W and fixed ρ parameters. We generate three pairs of time series $\{y_i\}$ and $\{y'_i\}$ with parameters $\rho_1 = 0.2$ and $\rho_2 = 0.4$, and varying W. For each pair of time series (10^4 data points each), we perform the test given in equation (6) for different degrees of freedom, m. In Figure 5, the results of the test are plotted versus m, for each pair of time series. We also show the critical values of the $\chi^2(m)$ distribution versus m at the 5% level of significance. Note that for the pairs of time series investigated with W equal to 0.5 and 0.95, the curves of the test statistic are above the curve of the critical values of the $\chi^2(m)$ distribution. Now we find that for a very small cross-correlation coupling parameter ($W = 0.99$), the values of the test of equation (6) are very close to the critical values of the $\chi^2(m)$ distribution. Generally, except for values of W very close to 1, for a broad range of m values, the difference between the value of the test of equation (6) and the corresponding critical value of the $\chi^2(m)$ distribution is *positive* ($Q_{\mathrm{CC}}(m) > \chi^2_{0.95}(m)$). If the values of the test of equation (6) calculated between two time series are smaller than the critical values of the $\chi^2(m)$ distribution, $Q_{\mathrm{CC}}(m) < \chi^2_{0.95}(m)$,

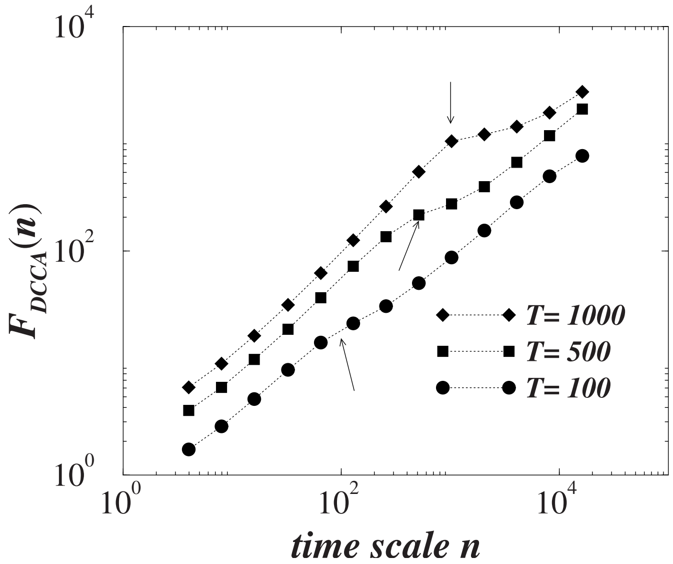

Fig. 6. Crossovers in the detrended cross-correlation analysis function $F_{\rm DCCA}(n)$, calculated for cross-correlated noise with a sinusoidal trend. The cross-correlated time series $\{y_i\}$ and $\{y'_i\}$ are generated by the periodic two-component ARFIMA process defined by equations (2a, 2b). For $\rho = 0.4$, $W = 0.5$, $A_1 = A_2 = 0.3$, and varying period T, we find a crossover in the $F_{\rm DCCA}(n)$ function that increases with T approximately as $n_{\rm 2CC} \propto T$.

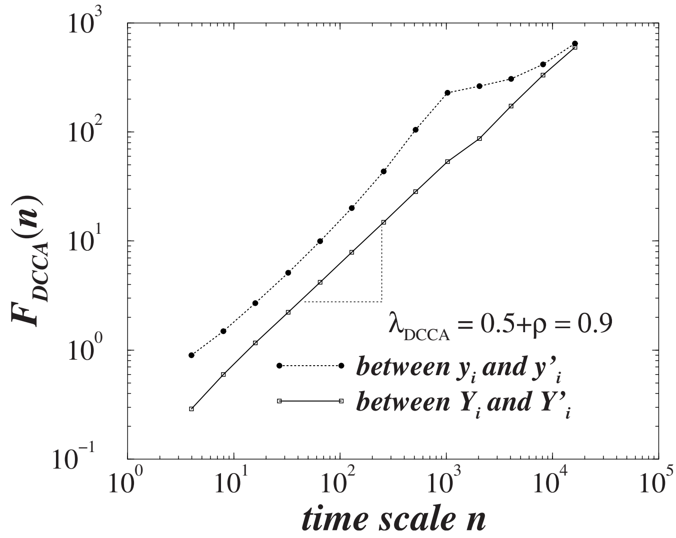

Fig. 7. DCCA function $F_{\rm DCCA}(n)$ after global and local sinusoidal detrending approaches. We generate two time series $\{y_i\}$ and $\{y'_i\}$ of the periodic two-component ARFIMA process of equations (2a, 2b) with $\rho = \rho_1 = \rho_2 = 0.4$, $W = 0.5$, $A_1 = A_2 = 0.3$, and $T_1 = T_2 = T = 10^3$. We see that $F_{\rm DCCA}(n)$ of y_i and y'_i exhibit a bump similar to that characteristic for DFA functions obtained for time series with periodic trends. After performing a global minimization of $\sum_{i=1}^{N}(y_i - A_1\sin(2\pi/Ti + \phi_1))^2$ and $\sum_{i=1}^{N}(y'_i - A_2\sin(2\pi/Ti + \phi_2))^2$, we find the parameters of the first harmonic (A_i, ϕ_i, T_i) in both time series $\{y_i\}$ and $\{y'_i\}$. Then, we define two new time series, $Y_i = y_i - A_1\sin(2\pi/Ti + \phi_1)$ and $Y'_i = y'_i - A_2\sin(2\pi/Ti + \phi_2)$. $\{Y_i\}$ and $\{Y'_i\}$ exhibit a "pure" power-law cross-correlation, with expected power-law exponent $\lambda_{\rm DCCA} = \lambda_{DFA} = 0.5 + \rho = 0.9$, since both time series are defined by the same ρ parameter.

the cross-correlations between two time series are insignificant.

Next we apply the detrended cross-correlations analysis (DCCA) to investigate cross-correlations between time series where each time series is auto-correlated and sinusoidal. In Figure 6 we present three DCCA functions $F_{\rm DCCA}(n)$ obtained for three pairs of cross-correlated time series $\{y_i\}$ and $\{y'_i\}$ with sinusoidal trends generated by the process of equations (2a, 2b), with $\rho = 0.4$, $W = 0.5$, $A_1 = A_2 = 0.3$, and varying T. We find from Figure 6 that each DCCA function $F_{\rm DCCA}(n)$ shows a crossover at time scale $n_{\rm 2CC} \approx T$ similar to the findings for DFA functions [24]. We next study numerically if the relation $n_{\rm 2CC} \approx T$ holds independently of the values of A, W, and ρ. We also find that the crossover bump becomes more pronounced with increasing A, but the crossover time scale $n_{\rm 2CC}$ does not depend on A. We also find that the crossover time scale $n_{\rm 2CC}$ virtually does not depend on ρ.

The correct interpretation of the scaling results is crucial for understanding the system that is analyzed. If the real-world time series is both correlated and periodic, periodicities should be eliminated before analyzing the correlations of the time series. First we generate two cross-correlated time series generated by the process of equations (2a, 2b) with $\rho = 0.4$ when periodicity is present in the time series. For the sake of simplicity, we set $T_1 = T_2 = T = 10^3$. We show the DCCA function $F_{\rm DCCA}(n)$ in Figure 7, and we find a crossover at scale $n_{\rm 2CC} \propto T$.

Next, we investigate the influence of global detrending on DCCA results. By global fit we assume one fit for the entire time series in contrast to a local detrending approach where local fits are accomplished for windows of different sizes. To eliminate periodicities in the original time series, we globally detrend the periodicity – $\sum_{i=1}^{N}[y_i - A_1\sin(2\pi/Ti + \phi_1)]^2$ and $\sum_{i=1}^{N}[y'_i - A_2\sin(2\pi/Ti + \phi_2)]^2$ – and find the parameters $(A_i, \phi_i,$ and $T)$ in both time series $\{y_i\}$ and $\{y'_i\}$. Then, we define the globally detrended time series $Y_i \equiv y_i - A_1\sin(2\pi/Ti + \phi_1)$, and $Y'_i \equiv y'_i - A_2\sin(2\pi/Ti + \phi_2)$. We find in Figure 7 that $F_{\rm DCCA}(n)$ practically loses the bump characteristic at the crossover scale, allowing us to calculate the cross-correlations exponent $\lambda_{\rm DCCA}$.

In this paper, we propose a new test to quantify the presence of cross-correlations. We propose that both the cross-correlation test and the detrended cross-correlations analysis (DCCA) should be used together to measure the degree of cross-correlations between different time series. We demonstrate that a good indication for the presence of cross-correlations is if the results of the statistical test of equation (6) exceeds the critical value of the $\chi^2(m)$ distribution at the given level of significance. We study long-range power-law cross-correlations between two time series, each power-law auto-correlated, in the presence of a periodic sinusoidal trend. We show that due to the sinusoidal trend, a spurious crossover exists in the DCCA cross-correlations plots. We study the impact of a Fourier phase-randomization on the cross-correlation test and show that the cross-correlations between two cross-correlated time series practically vanish by a Fourier phase randomization.

We thank the Ministry of Science of Croatia, NIH, and NSF for financial support.

Appendix A: Analytical cross-correlations derivation

Consider two stationary time series $\{y_j\}$ and $\{y'_j\}$, denote the covariance by $\{X_j\}_{j=-\infty}^{\infty}$, and denote the cross power spectrum by $s_{YY'}(\omega)$. Due to the cross-correlation theorem, sometimes called the Wiener-Khintchine theorem, the cross covariance function and the cross power spectrum are one-to-one related by

$$s_{YY'}(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} X_j \exp(-i\omega j). \tag{A.1}$$

A similar relationship exists for the auto-covariance function and the power spectrum for a single time series.

As an example, let us define two cross-correlated moving average MA(1) processes

$$y_i \equiv (1+\theta_1 L)\eta_i \equiv \Psi(L)\eta_i, \tag{A.2}$$

$$y'_i \equiv (1+\theta_2 L)\eta_i \equiv \tilde{\Psi}(L)\eta_i, \tag{A.3}$$

where η_i is an *(i.i.d.)* process with expectation value $E(\eta) = 0$ and variance $E(\eta^2) - E^2(\eta) = \sigma^2$, and L denotes the backward (lag) operator defined by $L\eta_i = \eta_{i-1}$, i.e., it simply relates two adjacent discrete-time coordinates i and $i-1$. Clearly, $\Psi(L)$ and $\tilde{\Psi}(L)$ are two linear polynomials in L.

For this example, one can easily calculate the only non-vanishing cross-covariances $X_0 \equiv E(y_i y'_i) = E(\eta_i^2 + \theta_1\theta_2\eta_{i-1}^2) = 1+\theta_1\theta_2$; $X_1 \equiv E(y_i y'_{i+1}) = \theta_2\sigma^2$, and $X_{-1} \equiv E(y_i y'_{i-1}) = \theta_1\sigma^2$. By using equation (A.1) for the power spectrum of two MA(1) processes we obtain $s_{YY'}(\omega) = \frac{\sigma^2}{2\pi}[X_0 + X_1\exp(i\omega) + X_{-1}\exp(-i\omega)]$. If $\exp(i\omega)$ is replaced by the complex number z, we obtain

$$\begin{aligned} s_{YY'}(\omega) &= \frac{1}{2\pi}[X_0 + X_1 z + X_{-1} z^{-1}] \\ &= \frac{1}{2\pi}[1+\theta_2 z][1+\theta_1 z^{-1}]. \end{aligned} \tag{A.4}$$

Using equations (A.2) and (A.3) the previous equation can be expressed as [47]: $s_{YY'}(\omega) = \frac{1}{2\pi}\tilde{\Psi}(z)\Psi(z^{-1})$. This relation for finding $s_{YY'}$ generally extends to the $MA(\infty)$ processes y_i and y'_i, where e.g. $y_i = \Psi(L)\eta_i$ and $\Psi(L) = a_0 + a_1 L + a_2 L^2 + ...$

The ARFIMA process y_i can not only be represented in the AR representation, but also in the $MA(\infty)$ representation:

$$y_i = (1-L)^{-d}\eta_i = \sum_{j=0}^{\infty} \frac{\Gamma(j+\rho)}{\Gamma(\rho)\Gamma(j+1)}\eta_{i-j}, \tag{A.5}$$

where ρ need not be an integer, provided $\rho < 1/2$, where the last expression is obtained after binomial expansion, $E(\eta) = 0$, and $E(\eta^2) - E^2(\eta) = \sigma^2$.

Consider a two-component ARFIMA process $\{y_i\}$ and $\{y'_i\}$ defined $y_i \equiv \sum_{j=1}^{\infty} a_j(\rho_1) y_{i-j} + \eta_i$, $y'_i \equiv \sum_{j=1}^{\infty} a_j(\rho_2) y'_{i-j} + \eta_i$ with parameters ρ_1 and ρ_2. For the cross power spectrum we obtain:

$$\begin{aligned} s_{YY'}(\omega) &= (1-\exp(i\omega))^{-\rho_1}(1-\exp(-i\omega))^{-\rho_2} \\ &= \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} a_k(\rho_1) a_{k'}(\rho_2)\exp(i(k-k')\omega), \end{aligned} \tag{A.6}$$

where $a_k(\rho) = \Gamma(k+\rho)/[\Gamma(\rho)\Gamma(k+1)]$ as defined in equation (A.5). Taking the inverse Fourier transform of the cross power spectrum, we obtain

$$\begin{aligned} X_n &= \frac{1}{2\pi}\int_{-\pi}^{\pi} s_{XY}(\omega)\exp(i\omega n) d\omega \\ &= \frac{\sigma^2}{2\pi}\frac{1}{2\pi}\sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} a_k(\rho_1) a_{k'}(\rho_2)\int_{-\pi}^{\pi}\exp(i(n+k-k')\omega) d\omega \\ &= \frac{\sigma^2}{2\pi}\sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} a_k(\rho_1) a_{k'}(\rho_2)\delta(n+k-k') \\ &= \frac{\sigma^2}{2\pi}\sum_{k=0}^{\infty} a_k(\rho_1) a_{n+k}(\rho_2). \end{aligned} \tag{A.7}$$

By using Stirling's expansion we obtain $a_k(\rho) \propto k^{\rho-1}$ and thus $X_n \propto \sum_{k=0}^{\infty} k^{\rho_1-1}(n+k)^{\rho_2-1}$, which can be approximated by $X_n \propto \int_0^{\infty} dk\ k^{\rho_1-1}(n+k)^{\rho_2-1}$. By defining a new variable $k/n = t$ we obtain

$$X_n \propto n^{\rho_1+\rho_2-1}\int_0^{\infty} t^{\rho_1-1}(t+1)^{\rho_2-1} dt. \tag{A.8}$$

If long-range power-law cross-correlations exist, we obtain $X_n \propto n^{-\gamma_{CC}}$ for the asymptotic regime $n \gg 1$, i.e., by using equation (A.8) we obtain $\gamma_{CC} = 1 - \rho_1 - \rho_2$. The parameter γ_{CC} and parameter λ_{DCCA} of the covariance growth (see Eq. (1)) are related as $\lambda_{DCCA} = 1 - 0.5\ \gamma_{CC}$ [11]. Hence, we obtain the result of equation (8), $\lambda_{DCCA} = \frac{H_1+H_2}{2}$, where H_1 and H_2 are the Hurst exponents related to the processes $\{y_i\}$ and $\{y'_i\}$, respectively, and where $H_1 = 0.5 + \rho_1$ and $H_2 = 0.5 + \rho_2$ [37]. Hence, we find analytically that the time series y_i and y'_i are long-range power-law cross-correlated (besides being long-range power-law auto-correlated), where the exponent λ_{DCCA} is equal to the arithmetic mean of the two Hurst exponents H and H'.

References

1. M. Campillo, A. Paul, Science **299**, 547 (2003)
2. B. LeBaron, W.B. Arthur, R. Palmer, J. Econ. Dyn. Control **23**, 1487 (1999)
3. R.N. Mantegna, Eur. Phys. J. B **11**, 193 (1999)
4. L. Laloux et al., Phys. Rev. Lett. **83**, 1467 (1999)
5. V. Plerou et al., Phys. Rev. Lett. **83**, 1471 (1999); V. Plerou et al., Phys. Rev. E **66**, 066126 (2002)
6. L. Kullmann, J. Kertesz, K. Kaski, Phys. Rev. E **66**, 026125 (2002)
7. T. Mizunoa, H. Takayasu, M. Takayasu, Physica A **364** 336 (2006)
8. M. Tumminello, T. Aste, T. Di Matteo, R.N. Mantegna, Proceedings of the National Academy of Sciences of the United States of America (PNAS) **102**, 10421 (2005)
9. R. Coelho, S. Hutzler, P. Repetowicz, P. Richmond, Physica A **373**, 615 (2007)
10. B. Podobnik et al., Eur. Phys. J. B **56**, 47 (2007)
11. B. Podobnik, H.E. Stanley, Phys. Rev. Lett. **100**, 084102 (2008)
12. T. Conlon, H.J. Ruskin, M. Crane, Physica A **388** 705 (2009)
13. S. Arianos, A. Carbone, J. Stat. Mech. P03037 (2009)
14. P. Sieczka, J.A. Holyst, Physica A **388**, 1621 (2009)
15. P. Samuelsson et al., Phys. Rev. Lett. **91**, 157002 (2003); A. Cottet et al., Phys. Rev. Lett. **92**, 206801 (2004)
16. H.E. Hurst, Proc. Inst. of Civ. Eng. **1**, 519 (1951)
17. F. Caserta et al., Phys. Rev. Lett. **64**, 95 (1990); D.C. Hong et al., Phys. Rev. B **30**, 4083 (1984)
18. T. Vicsek, *Fractal Growth Phenomenon*, 2nd edn. (World Scientific, Singapore, 1993)
19. *Fractals in Science*, edited by A. Bunde, S. Havlin (Springer, Berlin, 1994)
20. J.B. Bassingthwaighte, L.S. Liebovitch, B.J. West, *Fractal Physiology* (Oxford U. Press, New York, 1994)
21. H. Takayasu, *Fractals in the Physical Sciences* (Manchester U. Press, Manchester, 1997)
22. D. Markovic, M. Koch, Water Resour. Res. **41**, 09420 (2005); D. Markovic, M. Koch, Geophys. Res. Lett. **32**, 17401 (2005)
23. C.-K. Peng et al., Phys. Rev. E **49**, 1685 (1994)
24. K. Hu et al., Phys. Rev. E **64**, 011114 (2001)
25. Z. Chen et al., Phys. Rev. E **65**, 041107 (2002)
26. Z. Chen et al., Phys. Rev. E **71**, 011104 (2005)
27. L. Xu et al., Phys. Rev. E **71**, 051101 (2005)
28. P.Ch. Ivanov et al., Chaos **11**, 641 (2001)
29. Y. Liu et al., Physica A **245**, 437 (1997); Y. Liu et al., Phys. Rev. E **60**, 1390 (1999); P. Cizeau, Phys. Rev. E **245**, 441 (1997); P.Ch. Ivanov et al., Phys. Rev. E **69**, 056107 (2004)
30. S. Buldyrev et al., Biophys. J. **65**, 2673 (1993); S. Buldyrev et al., Phys. Rev. E **47**, 4514 (1993)
31. K. Ivanova, M. Ausloos, Physica A **274**, 349 (1999); K. Ivanova, J. Geophys. Res. **108**, 4268 (2003)
32. Z. Chen et al., Phys. Rev. E **73**, 031915 (2006)
33. C.W.J. Granger, J. Econometrics **14**, 227 (1980)
34. C.W.J. Granger, R. Joyeux, J. Time Series Analysis **1**, 15 (1980)
35. J. Hosking, Biometrika **68**, 165 (1981)
36. B. Podobnik et al., Phys. Rev. E **71**, 025104(R) (2005)
37. B. Podobnik et al., Phys. Rev. E **72**, 026121 (2005)
38. R.H. Shumway, D.S. Stoffer, *Time Series Analysis anf Its Applications*, Springer Texts in Statistics (Springer-Verlag, New York, 2000)
39. R.L. Anderson, Annals Math. Statistics **13**, 1 (1942)
40. G.E.P. Box, G.M. Jenkins, G.C. Reinsel, *Time Series Analysis: Forecasting and Control* (Prentice Hall, New Jersey, 1994)
41. G.M. Ljung, G.E.P. Box, Biometrika **65**, 297 (1978)
42. H. Tastan, Physica A **360**, 445 (2006); K.M. Wang, T.B.N. Thi, Physica A **376**, 422 (2007); M.C. Lee, C.L. Chiu, Y.H. Lee, Physica A **377**, 199 (2007); A. Kasman, S. Kasman, Physica A **387**, 2837 (2008); C.H. Tseng, S.T. Cheng, Y.H. Wang, J.T. Peng, Physica A **387**, 3192 (2008)
43. `www.yahoo.finance`
44. J. Theiler et al., Physica D **58**, 77 (1992)
45. P.Ch. Ivanov et al., Nature **383**, 323 (1996)
46. B. Podobnik et al., Physica A **387**, 3954 (2008)
47. W.W.S. Wei, *Time Series Analysis Univariate and Multivariate Methods* (Addison-Wesley, Prentice Hall, 2006)