Cross-Correlations Methods

B. Podobnik[a] and **H.E. Stanley**[b]

[a]Faculty of Civil Engineering, University of Rijeka, Rijeka, Croatia
V. Cara Emina 3, 51000 Rijeka, Croatia
bp@phy.hr

[b] Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215
hes@buphy.bu.edu

Keywords detrending methods, cross-correlations, time series analysis, nonlinearity

Auto-correlation analysis and cross-correlation analysis are two commonly-used techniques for providing insight into the dynamics of natural systems in presence of stationarity. However, many real-world data are nonstationary and exhibit long-range correlations [1, 2]. Recently the detrended cross-correlation analysis (DCC) was proposed to quantify cross-correlations in presence of nonstationarity [3]. Here, we introduce a new cross-correlations test $Q_{\rm CC}$, to quantify presence of cross-correlations in data. Motivated by the Ljung-Box test [4], we consider two *i.i.d.* time series, $\{y_i\}$ and $\{y'_i\}$, and calculate their cross-correlation coefficients $X_k \equiv \sum_{i=1}^{N-k} y_i y'_{i+k} / \sqrt{\sum_{i=1}^{N} y_i^2} \sqrt{\sum_{i=1}^{N} y_i'^2}$, where $E(X_k, X_{k'}) = 0$ $(k \neq k')$, and $E(X_k^2) = (N-k)/N^2$. The absence of cross-correlations between $\{y_i\}$ and $\{y'_i\}$ implies $E(X_k) = 0$ [5]. The X_k is normally distributed for $N >> 1$ [5]. Then $X_k/\sqrt{(N-k)/N^2}$ approximately follows a Gaussian distribution with zero mean and unit variance, and the sum of squares of these variables approximately follows the $\chi^2(m)$ distribution with m degrees of freedom. We propose the cross-correlation test which is approximately $\chi^2(m)$ distributed,

$$Q_{\rm CC}(m) \equiv N^2 \sum_{k=1}^{m} \frac{X_k^2}{N-k}. \tag{1}$$

As the LJB test, the test of Eq. (1) should be applied for the *residuals* of a given model. However, the test of Eq. (1) can be also used to measure the strength of cross-correlations in the original time series. If the test exceeds the critical value of the $\chi^2(m)$ distribution, then we say that the cross-correlations are significant. If for a broad range of m the test of Eq. (1) exceeds the critical values of $\chi^2(m)$ $(Q_{\rm CC}(m) > \chi^2_{0.95}(m))$, we claim that there are long-range cross-correlations. We propose that both the DCC function $F_{\rm DCC}(n) \propto n^{\lambda_{\rm CC}}$ [3], where $\lambda_{\rm CC}$ is the detrended cross-correlation scaling exponent, and the cross-correlations test of Eq. (1) need to be used in order to confirm existence of long-range cross-correlations. In order to investigate the applicability of the proposed test above, we propose the two-component ARFIMA model (based on ARFIMA process [6]), yielding two long-range auto-correlated and long-range cross-correlated time series:

$$y_i = \left[W \sum_{n=1}^{\infty} a_n(\rho_1) y_{i-n} + (1-W) \sum_{n=1}^{\infty} a_n(\rho_2) y'_{i-n} \right] + \eta_i, \tag{2}$$

$$y'_i = \left[(1-W) \sum_{n=1}^{\infty} a_n(\rho_1) y_{i-n} + W \sum_{n=1}^{\infty} a_n(\rho_2) y'_{i-n} \right] + \eta'_i. \tag{3}$$

Here, η_t and η'_t are two *i.i.d* Gaussian variables with zero mean and unit variance, and $0.5 < W < 1$ is a free parameter controlling the strength of power-law cross-correlations. $a_j(\rho_k) \equiv \frac{\Gamma(j-\rho_m)}{\Gamma(-\rho_m)\Gamma(1+j)}$ are statistical weights, where $0 < \rho < 0.5$.

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 [7]. In Fig. 1 we show the test (filled symbols) for the pairs of time series $\{y_i\}$ and $\{y'_i\}$ of Eqs. (2)-(3) together with the critical values of $\chi^2(m)$ for different m values. We also show the test after (open symbols) performing Fourier phase-randomization, for which case the cross-correlations are reduced compared to the case before Fourier phase-randomization. Thus, while Fourier phase-randomization procedure preserves linear auto-correlations [7], the same method substantially reduces the linear cross-correlations.
Next we investigate how the cross-correlation exponent $\lambda_{\rm CC}$ is estimated and how it relates to the DFA exponents α [8] calculated for each of two cross-correlated time series of Eqs. (2)-(3). In Figs. 2(a)-2(b), we show the DFA plots for each time series $\{y_i\}$ and $\{y'_i\}$ and $\rho_1 = 0.4$ and $\rho_2 = 0.1$, where $W = 0.95$ [Fig. 2(a)] and $W = 0.5$ [Fig. 2(b)]. We show that $\{y_i\}$ and $\{y'_i\}$ are power-law auto-correlated and cross-correlated. For the y'_i, the DFA exponent virtually does not change with W—$\alpha \approx 0.6 = 0.5 + \rho_2$. In contrast, for the y_i, the DFA α exponent gradually decreases from $\alpha \approx 0.9 = 0.5 + \rho_1$ (when $W = 1$, not shown) toward $\alpha \approx 0.6$ (when $W = 0.5$) corresponding to the y'_i process.
We show in Figs. 2(a)-2(b) that, by varying the parameter W, $\lambda_{\rm CC}$ follows the DFA α corresponding to the y_i. By decreasing the value of W from $W = 1$ to $W = 0.5$, $\lambda_{\rm CC}$ gradually decreases toward $\alpha \approx 0.6$. Generally, for different time series of the process of Eqs. (2)-(3), where $\rho_1 > \rho_2$, we find that $\lambda_{\rm CC}$ is closer to the α exponent of the y_i process.
We also study the effect of periodic trends on systems with long-range auto-correlations and with long-range cross-correlations. We find that periodic trends severely affect the quantitative analysis of long-range cross-correlations, leading to crossovers and other spurious deviations from power laws. We find that both local and global detrending approaches together with phase-randomization should be applied to properly uncover long-range auto-correlations and cross-correlations in the random part of the underlying stochastic process. Precisely, one needs first to detrend sinusoidal trends in the time series globally. Then one needs to accomplish both the cross-correlation test and the DCC method. Finally we employ our methods on real-world financial data.

References

[1] A. Bunde and S. Havlin, eds., *Fractals in Science* (Springer, Berlin, 1994).

[2] H. Takayasu, *Fractals in the Physical Sciences* (Manchester U. Press, Manchester, 1997).

[3] B. Podobnik and H.E. Stanley, Phys. Rev. Lett. **100**, 084102, 2008.

[4] G. M. Ljung and G. E. P. Box, Biometrika **65**, 297 , 1978.

[5] R. H. Shumway and D. S. Stoffer, *Time Series Analysis anf Its Applications* (Springer Texts in Statistics. Springer-Verlag, New York, 2000).

[6] C. W. J. Granger, J. Econometrics **14**, 227, 1980.

[7] P. Ch. Ivanov, *et al.*, Chaos **11**, 641, 2001.

[8] C.-K. Peng *et al.*, Phys. Rev. E **49**, 1685, 1994.

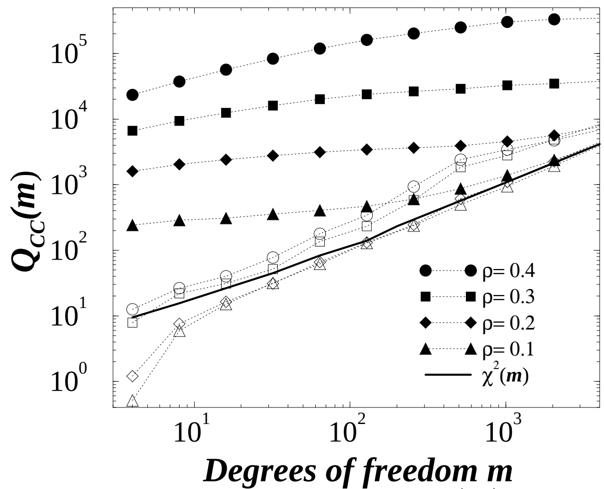

Figure 1: Effect of Fourier phase-randomization on $Q_{\mathrm{CC}}(m)$ of Eq. (1) for different m. For each ρ, the process generates $\{y_i\}$ and $\{y'_i\}$, where $W = 0.5$. For each ρ, we phase-randomize $\{y'_i\}$, and obtain series $\{\tilde{y}'_i\}$. For each $(\{y_i\}, \{\tilde{y}'_i\})$, we calculate the $Q_{\mathrm{CC}}(m)$ test. After a phase-randomization procedure (open symbols), the cross-correlations are substantially reduced.

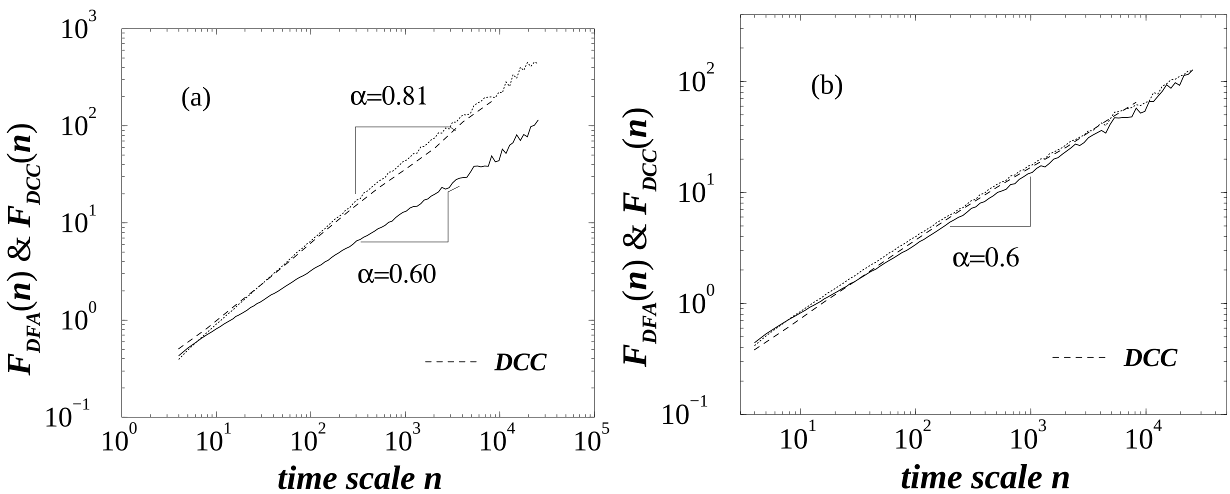

Figure 2: Detrended fluctuation analysis (DFA) and detrended cross-correlations analysis (DCC) functions $F_{\mathrm{DFA}}(n)$ and $F_{\mathrm{DCC}}(n)$, respectively, versus scale n. We generate the time series $\{y_i\}$ and $\{y'_i\}$ of Eqs. (2)-(3) with $\rho_1 = 0.4$ and $\rho_2 = 0.1$, respectively. We show the two DFA functions, $F_{\mathrm{DFA}}(n) \propto n^{\alpha}$, and the DCC function, $F_{DCC}(n) \propto n^{\lambda_{\mathrm{CC}}}$, for coupling (a) $W = 0.95$ and (b) for $W = 0.5$. Generally, by varying W, λ_{CC} becomes closer to α corresponding to y_i, but eventually the α value corresponding to y_i tends to the α value corresponding to y'_i.