ELSEVIER

Physica A 302 (2001) 234–243

PHYSICA A

www.elsevier.com/locate/physa

Scaling analysis of trends using DFA

Dmitry Vjushin[a,b], R.B. Govindan[a,b], Roberto A. Monetti[a,*], Shlomo Havlin[a], Armin Bunde[b]

[a] *Minerva Center and Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel*

[b] *Institut für Theoretische Physik III, Justus-Liebig-Universität Giessen, Heinrich-Buff-Ring 16, 35392 Giessen, Germany*

Abstract

We study two aspects of the detrended fluctuation analysis (DFA) method, namely the scaling behavior of the leading terms of the best-fit polynomials and the detection of trends. We show analytically and numerically that the standard deviation of the leading terms of the best-fit polynomials used in DFA displays scaling behavior. Furthermore, we demonstrate that the distribution of these terms can be used to reveal the presence of trends in the data. We also argue that this distribution can be used as a sensitive tool for identifying weak trends. © 2001 Elsevier Science B.V. All rights reserved.

PACS: 02.70.Hm; 05.45.Tp; 89.75.Da

Keywords: Correlations; Trends; DFA; Scaling

1. Introduction

Long-range correlations in time series are frequently studied by means of the detrended fluctuation analysis (DFA). The basic method was introduced by Peng et al. [1] to study long-range correlations in DNA sequences and has been successfully applied to a variety of systems ranging from DNA [2–4], heart rate dynamics [5–8], human gait [9,10], neuron spikes [11] to atmospheric temperature [12–15] and stock market variability [16,17].

Recently, attempts have been made to understand the effect of trends present in the data through the DFA method [18,19]. Here, we present an alternative method. Instead of analyzing the fluctuations around the best polynomial fits for a certain time window of length s, we analyze the best fits themselves. We show that both correlations and

* Corresponding author.

E-mail address: roberto@shoshi.ph.biu.ac.il (R.A. Monetti).

PII: S0378-4371(01)00467-8

trends can be identified through the histogram of the coefficients of the polynomial terms used in the DFA method. In fact, while the mean value of the histogram of the slopes indicates the presence of trends, the standard deviation of the histogram exhibits scaling behavior and provides information about long-range correlations. We also propose an analytical approach for these results.

This paper is organized as follows. In Section 2 we present the methodology used for trend analysis. Section 3 is devoted to results and discussions and finally we conclude in Section 4.

2. Methodology

In order to explain our approach, we briefly review some of the steps comprising the DFA algorithm. Consider a typical time window s such as $s \ll N$, where N is the length of the time series. As in conventional DFA, instead of analyzing the raw time series $\{u_i\}$ the profile $Y(l) = \sum_{i=1}^{l} u_i$ is studied. We use a DFA algorithm where the profile is divided into $r = N - s + 1$ overlapping intervals of size s and the local trend for each interval v is calculated via a least-square fit of the data. Then, the detrended profile is defined for intervals of duration s, as the difference between the profile and the fits

$$Z_v(l) = Y_v(l) - p_v^{(k)}(l)\,, \tag{1}$$

where $p_v^{(k)}(l)$ is the best-fit polynomial of order k to $Y_v(l)$ in the vth interval. Different orders k of DFA (DFA1, DFA2, etc.) only differ in the order of the polynomial used in the fitting procedure. Since the detrending of the time series is done by the subtraction of the fits from the profile, different orders differ in their capability of eliminating trends in the data. In fact, DFAk removes polynomial trends of order k in the profile and hence $(k-1)$th order in the time series.

The DFA method has been designed to reveal the scaling behavior of the detrended profile, Z_v. For a given window size s, we compute the root mean square fluctuation

$$F(s) = \sqrt{\frac{1}{rs}\sum_{v=1}^{r}\sum_{l=1}^{s} Z_v^2(l)}\,.$$

The above computation should be repeated for windows of different lengths to establish a relationship between $F(s)$ and s. In the presence of long-range correlations the fluctuation function increases according to the following power law

$$F(s) \sim s^{\alpha}\,,$$

where α is the fluctuation exponent. In the alternative approach developed here, we focus on the analysis of the coefficients of the polynomial terms and study two aspects: (i) the histogram of the values of the coefficients (direct method) since they contain information about the trends, and (ii) the scaling behavior of the standard deviation of the coefficients (scaling approach). We describe the approach for DFA1 and DFA2.

In the conventional DFA1 introduced by Peng et al. [1], a first-order polynomial is used for each segment, i.e.,

$$p_v^{(1)}(l) = a_{1,v}^{(1)} l + a_{0,v}^{(1)} , \tag{2}$$

where $a_{1,v}^{(1)}$ and $a_{0,v}^{(1)}$ are the fitting coefficients corresponding to the vth interval. Since there are r time windows of length s, DFA1 will provide r pairs $(a_{1,v}^{(1)}, a_{0,v}^{(1)})$ of values, each of them corresponding to a particular time window. Then, it is straightforward to generate histograms using $a_{1,v}^{(1)}$ or $a_{0,v}^{(1)}$. For DFA2, we use a second-order polynomial in the fitting procedure

$$p_v^{(2)}(l) = a_{2,v}^{(2)} l^2 + a_{1,v}^{(2)} l + a_{0,v}^{(2)} , \tag{3}$$

where $a_{2,v}^{(2)}$, $a_{1,v}^{(2)}$, and $a_{0,v}^{(2)}$ are the fitting coefficients. So, histograms can be generated in the same way using the values of $a_{2,v}^{(2)}$, $a_{1,v}^{(2)}$, and $a_{0,v}^{(2)}$ corresponding to each segment v. The fitting coefficients $a_{1,v}^{(1)}$ and $a_{0,v}^{(1)}$ for DFA1 and $a_{2,v}^{(2)}$, $a_{1,v}^{(2)}$, and $a_{0,v}^{(2)}$ for DFA2, etc., can be best analyzed by forming histograms. It can be analytically proved (not given here) that the histogram curves generated from correlated random Gaussian noise are also Gaussian.

Next, we show how these histograms for a particular time window s can be useful to detect trends. Also, we will investigate how the standard deviation σ of the histogram varies for different time windows, i.e., the scaling behavior of histograms.

3. Results and discussion

3.1. Direct method

To demonstrate our approach we generated a series of Gaussian distributed correlated data x_i (of length $N = 100$ K) with fluctuation exponent $\alpha = 0.7$, zero mean and unit standard deviation by means of the Prakash et al. algorithm [20,21]. In addition, using this time series, we generated a new one by adding a trend as follows:

$$u_i = x_i + Ai , \tag{4}$$

where A is a constant.

Since the aim of DFA method is to remove trends (see Eqs. (2) and (3)), histograms obtained from time series where trends are absent should have a mean value equal to zero. However, for finite time series the histograms will usually exhibit a mean value different from zero. Then, a criterion is necessary to decide whether or not a trend is present in the data, i.e., we need to estimate the lower detection limit for our method. This can be accomplished by calculating the mean m of the slopes over $R \gg 1$ different realizations (of length N of a pure noise with the same fluctuation exponent α) and the standard deviation δ among the mean values. The number of realizations R should be

Table 1
Values of the standard deviation δ among the means of the histograms calculated for several values of N and α[a]

N/α	0.3	0.5	0.7
10 K	2.9×10^{-4}	5×10^{-4}	8×10^{-4}
100 K	2.7×10^{-5}	4.4×10^{-5}	7.8×10^{-5}
1000 K	3.5×10^{-6}	6.2×10^{-6}	1.1×10^{-5}

[a]The values correspond to $s = 32$.

increased until convergence is obtained in the value of δ. Now, we propose a criterion for the existence of a trend just by checking whether or not the mean value of the slopes lies within the interval $[-\delta, +\delta]$. Consequently, if the mean value lies well outside this interval, we can confidently say that the data contain a trend. For this purpose, we evaluated for three different values of N and three different values of α the testing intervals to apply the criterion for the existence of trends. The results are shown in Table 1. We have checked that δ does not change significantly when increasing R, so the values reported can be considered as asymptotic values.

Figs. 1(a) and (b) show histograms obtained using DFA1 for $A = 0$ and time windows of length $s = 32$ and 512, respectively. We observe that the mean values of the histograms (see Figs. 1(a) and (b)) are within the range defined in Table 1 indicating the absence of trends. Figs. 1(c) and (d) show similar plots obtained for $A = 0.25 \times 10^{-5}$. It is now possible to see that the mean values (see Figs. 1(c) and (d)) of the histograms are well outside the range defined in the same table. This is the signature of the presence of a trend. It can be shown that for a linear trend (Ai) the mean value of the slopes is given by

$$\langle a_{1,v}^{(1)} \rangle = AN/2 , \tag{5}$$

which is independent of the analyzing time window s. The values shown in Figs. 1(c) and (d) are in accordance to Eq. (5) within the numerical error.

Earlier works on the detection of trends using DFA [18,19] have focused on the analysis of the position of the crossover of the DFA curves. Hu et al. [19] have shown that for correlated noise with exponent α containing a linear trend, the position s_z of the crossover in the curve generated by DFA1 is given by

$$s_z \sim A^{1/(\alpha-2)} . \tag{6}$$

Requesting in Eq. (6) that the position of the crossover s_z lies beyond the longest analyzing time window in DFA (normally $N/4$), it is possible to estimate the weakest linear trend (lower bound) that DFA1 can identify

$$A > \left(\frac{N}{4}\right)^{\alpha-2} \equiv A_{\min} . \tag{7}$$

To illustrate the sensitivity of our method, we compare $A_{\min}$ to the minimum trend A_{low} that the present method can detect. In fact, A_{low} can be calculated by requesting

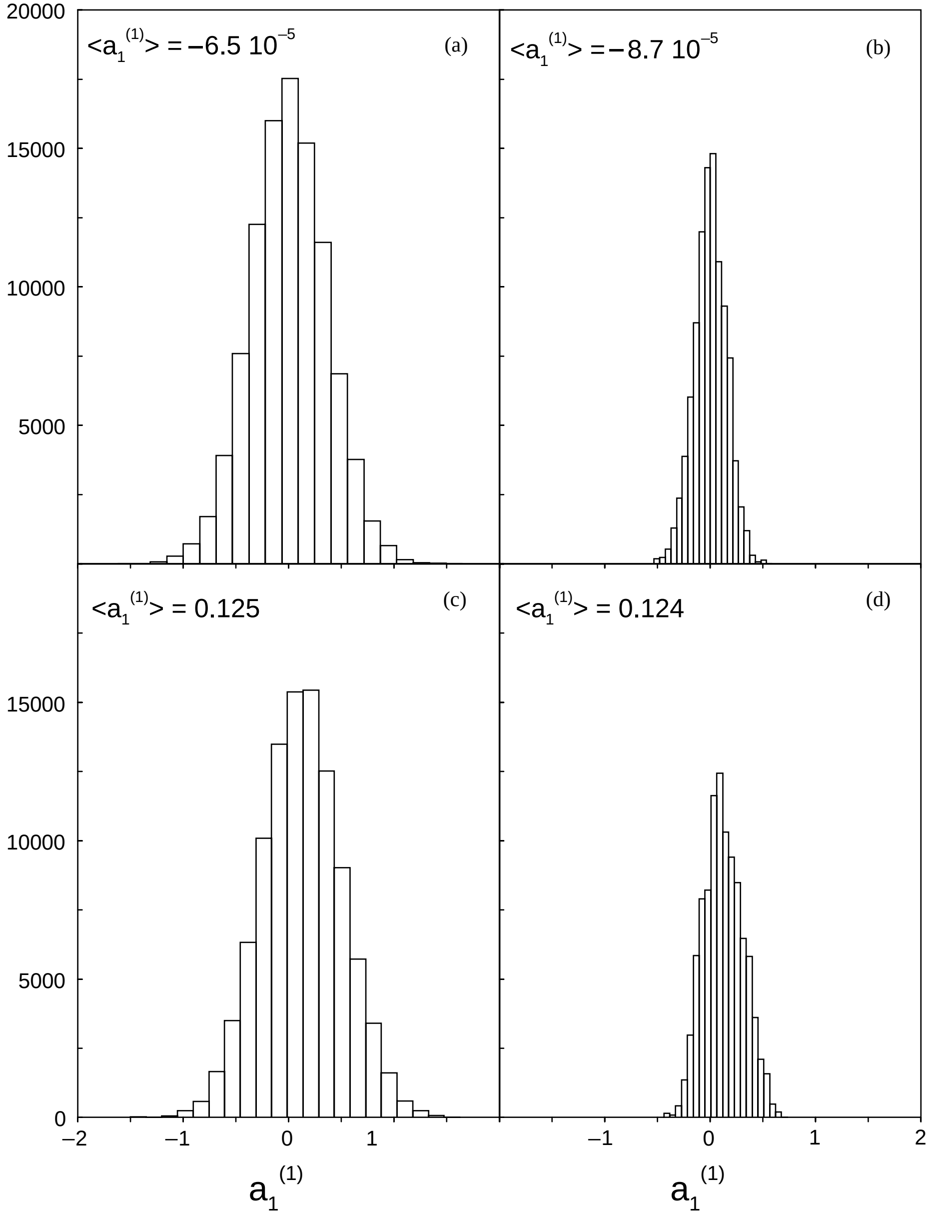

Fig. 1. Histograms of the leading slopes obtained from DFA1. Panels (a) and (c) represent results for a time window 32, and (b) and (d) are for a time window 512. While (a) and (b) consider the data (correlated noise) without any trend, (c) and (d) consider the data with a linear trend of $A = 0.25 \times 10^{-5}$.

that $\langle a_{1,v}^{(1)} \rangle = \delta$, so we obtain

$$A_{\text{low}} = \frac{2\delta}{N} . \tag{8}$$

It should be noticed that for the cases considered in Table 1, the values of A_{low} are at least 2 orders of magnitude smaller than the values of A_{min}.

In order to provide a criterion for the existence of trends in a real situation, the simulation suggested above should be performed using artificial data of the same length, distribution, and fluctuation exponent as those of the real data.

3.2. Scaling analysis

It is clear that the histogram width (i.e., the standard deviation of the slopes) decreases when increasing s (see Fig. 1). We will restrict ourselves, without loss of generality, to the study of histograms generated using the leading slopes. Fig. 2 shows log–log plots of σ of the leading slopes versus s. In the case of pure noise (Fig. 2 (a)), σ displays a power law behavior for DFA1-5. The values of the exponents obtained are also shown in the figure. It is possible to obtain by means of simple scaling arguments the behavior of σ. Consider the standard deviation of the leading slopes $\sigma_s^{(k)}$ corresponding to order k of DFA and window size s. As discussed above, the DFA method involves a fitting of the profile Y using a polynomial of order k. The leading slope can be estimated through the following relation:

$$\frac{\mathrm{d}^k Y(l)}{\mathrm{d}l^k} \sim a_{k,v}^{(k)} . \tag{9}$$

For pure noise in the absence of trends, the scaling behavior of the fluctuations of the profile is given by the following expression [22]:

$$\sigma_s(Y) \sim s^{\alpha} . \tag{10}$$

Then, from Eqs. (9) and (10) we obtain

$$\sigma_s^{(k)} \sim s^{-(k-\alpha)} = s^{-\eta_k} , \tag{11a}$$

where

$$\eta_k = (k - \alpha) . \tag{11b}$$

A more rigorous proof of this scaling relation for DFA1 is given in the appendix. The values of the exponents corresponding to DFA1-5 (see Fig. 2(a)) are in a good agreement with the scaling behavior predicted by Eq. (11a). Fig. 2(b) shows the scaling analysis performed using noise containing a weak linear trend. Since DFA of order k removes polynomial trends up to the order $(k-1)$, deviations from the scaling form predicted by Eq. (11a) could only be observed in $\sigma_s^{(1)}$. However, such a weak trend is not strong enough to cause a significant change in $\sigma_s^{(1)}$. Noise containing a stronger trend is analyzed in Fig. 2(c). In this case, the scaling behavior of $\sigma_s^{(1)}$ displays a crossover at s_z revealing the presence of the trend. Due to the scale used in Fig. 2 the

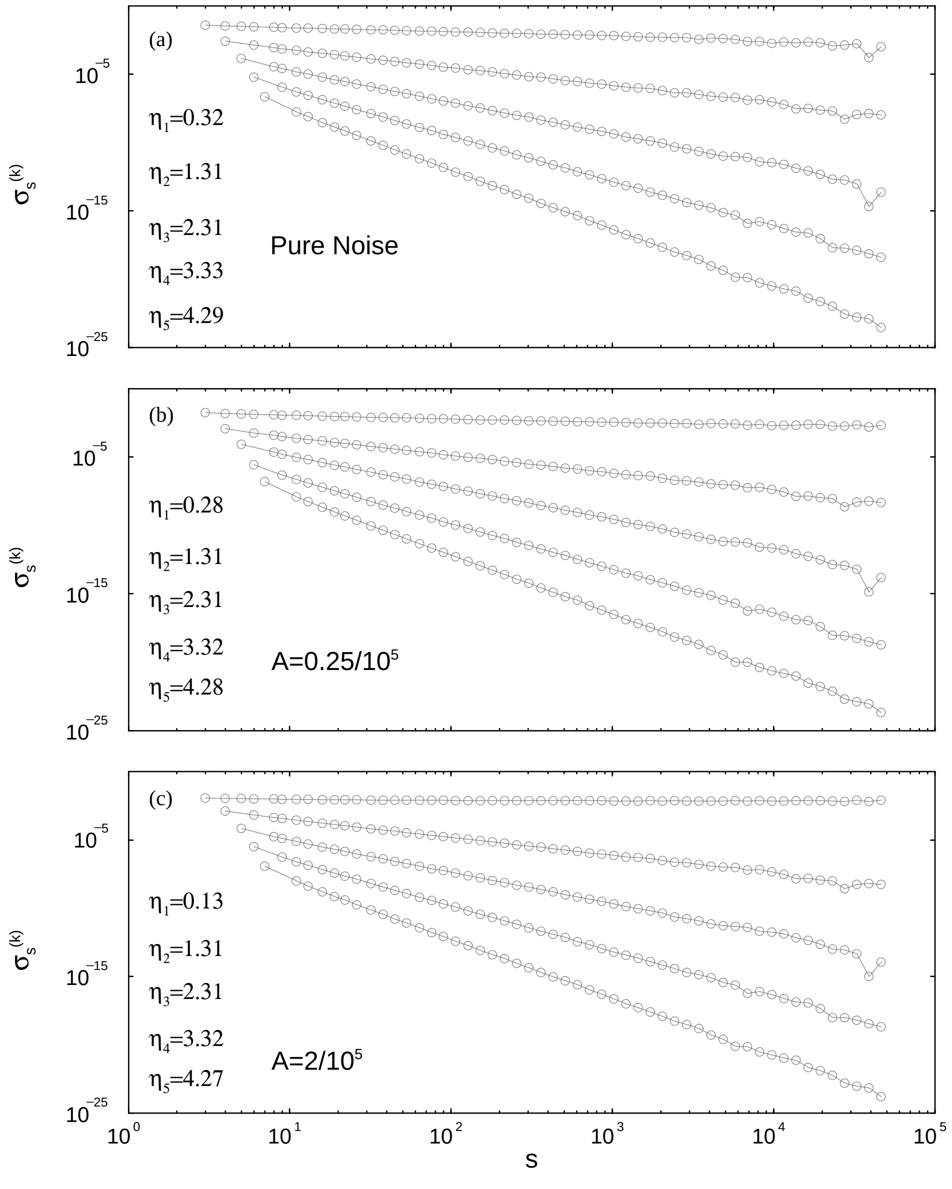

Fig. 2. Plots of the standard deviation of the leading slopes σ versus the time window s. From top to bottom, results correspond to DFA1-5. (a) Correlated noise with $\alpha = 0.7$ without any trend, (b) noise with linear trend of $A = 0.25 \times 10^{-5}$ and (c) $A = 2 \times 10^{-5}$.

position of the crossover s_z in the $\sigma_s^{(1)}$ curve is not clearly observed. In fact, within the short-time regime $s < s_z$, the scaling behavior predicted by Eq. (11a) holds because the noise dominates over the trend [18]. However, since the trend controls the system for $s > s_z$, the value of the asymptotic slope ($\eta_1 = 0.13$) differs from the value given by

Eq. (11a) ($\eta_1 = 0.30$). We expect that for the σ curves, the behavior of the position of the crossover s_z displays similar features to that found in the analysis of DFA method [19].

4. Conclusions

In this paper, we have proposed a new method for the detection of trends based on DFA. In addition, we have found that the analysis of the slopes of the best-fit polynomials provides a new way of estimation of the fluctuation exponent. These results might be useful for forecasting some aspects of the dynamics of systems such as trends in the atmosphere, tendencies in stock markets, patchiness in DNA sequences, etc.

Acknowledgements

We wish to acknowledge Dr. Webman for useful discussions. RAM would like to acknowledge CONICET (Argentina) for financial support.

Appendix

For DFA1, the slope $a_{1,v}^{(1)}$ for vth window of size s is given by

$$a_{1,v}^{(1)} = 6\frac{2\sum_{i=v}^{v+s-1} iY_i - (2v+s-1)\sum_{i=v}^{v+s-1} Y_i}{s(s^2-1)} . \tag{A.1}$$

Since the mean of $a_{1,v}^{(1)}$ is equal to zero for a pure noise, $\sigma_s^2[a_{1,v}^{(1)}] = M[(a_{1,v}^{(1)})^2]$. Then,

$$\begin{aligned} M[(a_{1,v}^{(1)})^2] &= \frac{36}{s^2(s^2-1)^2} M\left\{\left[2\sum_{i=v}^{v+s-1} iY_i - (2v+s-1)\sum_{i=v}^{v+s-1} Y_i\right]^2\right\} \\ &= \frac{36}{s^2(s^2-1)^2}\left\{4M\left[\left(\sum_{i=v}^{v+s-1} iY_i\right)^2\right]\right. \\ &\quad - 4M[2v+s-1]M\left[\sum_{i=v}^{v+s-1} iY_i \sum_{i=v}^{v+s-1} Y_i\right] \\ &\quad \left. + M[(2v+s-1)^2]M\left[\left(\sum_{i=v}^{v+s-1} Y_i\right)^2\right]\right\} . \end{aligned} \tag{A.2}$$

Assuming that Y_i is a fractional Brownian motion with zero mean, variance $M[Y_i^2] = i^{2\alpha}$ and covariance $M[Y_iY_j] = \frac{1}{2}(i^{2\alpha} + j^{2\alpha} - |i-j|^{2\alpha})$ [22], we evaluate separately the following expressions:

$$M\left[\left(\sum_{i=v}^{v+s-1} iY_i\right)^2\right] \simeq \frac{(2\alpha^2+5\alpha+1)s^{2\alpha+4}}{4(\alpha+1)(\alpha+2)(2\alpha+1)},$$

$$M\left[\sum_{i=v}^{v+s-1} Y_i \sum_{i=v}^{v+s-1} iY_i\right] \simeq \frac{(3\alpha+1)s^{2\alpha+3}}{4(2\alpha+1)(2\alpha+2)},$$

$$M\left[\sum_{i=v}^{v+s-1} Y_i\right] \simeq \frac{s^{2\alpha+2}}{2\alpha+2},$$

$$M[2v+s-1] = N+2s-1,$$

$$M[(2v+s-1)^2] = \tfrac{4}{3}N^2 + \tfrac{14}{3}Ns - \tfrac{8}{3}N + \tfrac{13}{3}s^2 - \tfrac{4}{3}s + 1.$$

Substituting all this expressions in Eq. (A.2) we obtain

$$M[(a_{1,v}^{(1)})^2] \simeq 6\frac{20\alpha^2+53\alpha+20}{(\alpha+1)(\alpha+2)(2\alpha+1)}s^{2(\alpha-1)}. \tag{A.3}$$

Therefore

$$\sigma_s[a_{1,v}^{(1)}] \sim s^{-(1-\alpha)}. \tag{A.4}$$

References

[1] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Phys. Rev. E 49 (1994) 1685.

[2] S.V. Buldyrev, A.L. Goldberger, S. Havlin, R.N. Mantegna, M.E. Matsa, C.-K. Peng, M. Simons, H.E. Stanley, Phys. Rev. E 51 (1995) 5084.

[3] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, R.N. Mantegna, M. Simons, H.E. Stanley, Physica A 221 (1995) 180.

[4] S.V. Buldyrev, N.V. Dokholyan, A.L. Goldberger, S. Havlin, C.-K. Peng, H.E. Stanley, G.M. Viswanathan, Physica A 249 (1998) 430.

[5] C.-K. Peng, S. Havlin, H.E. Stanley, A.L. Goldberger, Chaos 5 (1995) 82.

[6] P.Ch. Ivanov, A. Bunde, L.A. Nunes Amaral, S. Havlin, J. Fritsch-Yelle, R.M. Baevsky, H.E. Stanley, A.L. Goldberger, Europhys. Lett. 48 (1999) 594.

[7] Y. Ashkenasy, P.Ch. Ivanov, S. Havlin, C.-K. Peng, A.L. Goldberger, H.E. Stanley, Phys. Rev. Lett. 86 (2001) 1900.

[8] A. Bunde, S. Havlin, J.W. Kantelhardt, T. Penzel, J.-H. Peter, K. Voigt, Phys. Rev. Lett. 85 (2000) 3736.

[9] J.M. Hausdorff, S.L. Mitchell, R. Firtion, C.-K. Peng, M.E. Cudkowicz, J.Y. Wei, A.L. Goldberger, J. Appl. Physiol. 82 (1997) 262.

[10] C.-K. Peng, J.M. Hausdorff, S. Havlin, J.E. Mietus, H.E. Stanley, A.L. Goldberger, Phys. Rev. E 55 (1997) 845.

[11] S. Blesic, S. Milosevic, D. Stratimirovic, M. Ljubisavljevic, Physica A 268 (1999) 275.

[12] E. Koscielny-Bunde, A. Bunde, S. Havlin, H.E. Roman, Y. Goldreich, H.-J. Schellnhuber, Phys. Rev. Lett. 81 (1998) 729.

[13] E. Koscielny-Bunde, A. Bunde, S. Havlin, Y. Goldreich, Physica A 231 (1996) 393.

[14] P. Talkner, R.O. Webber, Phys. Rev. E 62 (2000) 150.

[15] R.B. Govindan, D. Vjushin, S. Brenner, A. Bunde, S. Havlin, H.-J. Schellnhuber, Physica A 294 (2001) 239.
[16] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Phys. Rev. Lett. 83 (1999) 1471.
[17] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Physica A 279 (2000) 443.
[18] J.W. Kantelhardt, E. Koscielny-Bunde, H.H.A. Rego, S. Havlin, A. Bunde, Physica A 295 (2001) 441.
[19] K. Hu, P.Ch. Ivanov, Z. Chen, P. Carpena, H.E. Stanley, e-print physics/0103018.
[20] S. Prakash, S. Havlin, M. Schwartz, H.E. Stanley, Phys. Rev. A 46 (1992) R1724.
[21] H.A. Makse, S. Havlin, M. Schwartz, H.E. Stanley, Phys. Rev. E 53 (1996) 5445.
[22] M.S. Taqqu, V. Teverovsky, W. Willinger, Fractals 3 (1995) 785.