STATISTICAL PROPERTIES OF OLD AND NEW TECHNIQUES IN DETRENDED ANALYSIS OF TIME SERIES*

Dariusz Grech

Institute of Theoretical Physics, Wroclaw University
pl. M. Borna 9, 50-405 Wrocław, Poland
e-mail: dgrech@ift.uni.wroc.pl

and Zygmunt Mazur

Institute of Experimental Physics, Wroclaw University
pl. M. Borna 9, 50-405 Wrocław, Poland

(Received April 11, 2005)

Recently introduced Detrended Moving Average (DMA) method is examined and compared with Detrended Fluctuation Analysis (DFA) technique for artificial stochastic Brownian time series of various length $L \sim 10^3 \div 10^5$. Our analysis reveals some statistical properties of the Hurst exponent values measured with the use of DFA and DMA methods. Good agreement between DFA and DMA techniques is found for long time series $L \sim 10^5$, however for shorter series two methods are clearly distinguishable. No clear systematic relation previously postulated in literature between DFA and DMA results is found. However, it is shown that on the average, DMA method gives overestimation of the Hurst exponent compared with DFA technique.

PACS numbers: 05.40.Jc, 05.45.Tp, 89.65.Gh, 89.75.Da

1. Introduction

Investigation of stochastic time series is crucial for better understanding of various physical, biological, financial and economical processes. The main problem discussed in this context is the presence or absence of the autocorrelations in data. There are various techniques to do so. One of them is based on the measurement of the fractal structure of the given time series

* Presented at the First Polish Symposium on Econo- and Sociophysics, Warsaw, Poland, November 19–20, 2004.

and is related to the so called scaling exponent H sometimes denoted also as α [1–3]. This exponent plays a significant role as the main concept upon which fluctuations of a time series around its local trend (drift) are formed. Therefore it may be considered as the one of the crucial points responsible for 'genetic code' of time series of various origin.

The scaling exponent α is variously defined in literature. For the purpose of fractal analysis it can be introduced as follows.

Let $x(t)$ $(t = 1, \ldots, L)$ is the time series defined for discrete time points t. By rescaling time axis γ times (*e.g.* enlarging it $\times 10^n$), one reveals the tiny structure of time series not visible for smaller resolution ($\gamma \sim 1$). The fractal structure of the series comes from the relation:

$$x'(t') = \Gamma x(\gamma^{-1} t) \sim x(t) \,. \tag{1}$$

The above formula indicates that the magnitude of a time series should be simultaneously rescaled Γ times in order to get (local) self similarity correspondence between $x(t)$ and $x'(t')$ series. It turns out that the scaling factor Γ can be expressed in terms of time rescaling factor γ as:

$$\Gamma = \gamma^{\alpha} \,, \tag{2}$$

where $\alpha > 0$ is the real parameter known in literature as Holder exponent [1] called sometimes also Hurst or Hausdorff exponent [2,3].

The best known method to measure α exponent for both stationary and nonstationary series is Detrended Fluctuation Analysis (DFA) [4]. Recently, new method called Detrended Moving Average (DMA) has also been proposed [5,6]. Searches for better understanding how the results of these two methods relate to each other are in progress [6–8].

A DFA method first developed for biological purposes [4] and then applied also to finances [9–11] is a detrendisation technique basically measuring fluctuations of a given time series around its local trend as a function of its length. It is based on the following steps:

1. A given signal $x(t)$ $(t = 1, \ldots, L)$ of time series is divided into L/τ not overlapping boxes of length τ each.

2. A polynomial fit $x_{\tau,k}$ is constructed in each box representing the local trend in that box, where k is the order of polynomial fit.

3. A detrended signal $X_{\tau,k}(t)$ is found:

$$X_{\tau,k}(t) = x(t) - x_{\tau,k}(t) \tag{3}$$

and then its fluctuation (standard deviation)$F_{\mathrm{DFA}}(\tau,k)$ is calculated

$$F_{\mathrm{DFA}}(\tau,k)=\sqrt{\frac{1}{L}\sum_{t=1}^{L}X^2_{\tau,k}(t)}\,. \tag{4}$$

4. From the basic differential stochastic equation of the time series $x(t)$ with a local drift $\mu(t)$ and a local dispersion $\sigma(t)$

$$dx(t)=\mu(t)dt+\sigma(t)dX(t) \tag{5}$$

one expects the power law behavior:

$$F_{\mathrm{DFA}}(\tau,k)\sim\tau^{\alpha(k)}\,, \tag{6}$$

where $\alpha(k)$ is the searched Hurst exponent.

The last equation enables us to calculate α exponent directly from log–log linear fit

$$\log F_{\mathrm{DFA}}(\tau,k)\sim\alpha(k)\log\tau\,. \tag{7}$$

It can be proved that $\alpha(k)$ depends very weakly on k [11, 12] so in most applications one takes linear function ($k=1$) as a good candidate for $x_{\tau,k}$. This approach will be used further in our paper.

The examples of artificial time series constructed for different values of α exponent are shown in Fig. 1. It is seen that the bigger α the more

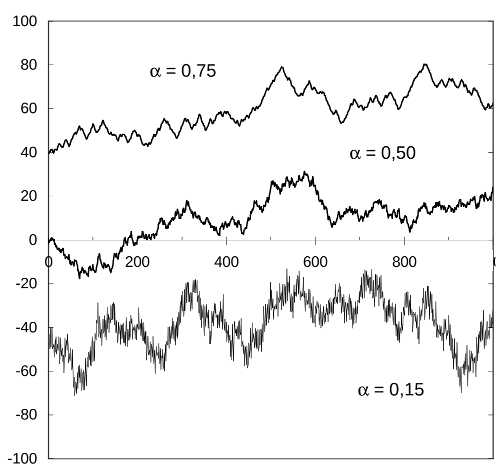

Fig. 1. Examples of time series with different α exponents.

'quiet' time series is, *i.e.* a signal fluctuates in a more correlated way. In fact, for $0 < \alpha < 1/2$ we have negative autocorrelations (antipersistence) in time series. On the other hand, if $1/2 < \alpha \leq 1$, there are positive autocorrelations (persistence) in signal. The case $\alpha = 1/2$ corresponds to completely uncorrelated signal, so called integer Brownian walk. An existing link between α exponent and the probability that a given trend will last in the immediate future, if it did so in the immediate past, gives an additional hint about trend changes forecast possibility [13].

A Detrended Moving Average (DMA) technique looks very similar to DFA. The main difference one meets here is that instead of linear or polynomial detrendisation procedure in equally sized boxes, one uses moving average of a given length λ. Thus basic steps of DMA analysis are as follows:

1. A simple moving average of length λ ($\lambda = 1, \ldots, L$) is constructed for $x(t)$ series ($t \geq \lambda$):

$$\langle x(t)\rangle_\lambda = \frac{1}{\lambda}\sum_{k=0}^{\lambda-1} x(t-k)\,. \tag{8}$$

2. A detrended signal is found similarly to Eq. (3):

$$X_\lambda(t) = x(t) - \langle x(t)\rangle_\lambda \tag{9}$$

   and its fluctuation within a window of size λ reads now:

$$F_{\mathrm{DMA}}(\lambda) = \sqrt{\frac{1}{L-\lambda+1}\sum_{t=\lambda}^{L} X_\lambda^2(t)}\,. \tag{10}$$

3. Similarly to DFA a power law should be observed

$$\log F_{\mathrm{DMA}}(\lambda) \sim \alpha \log \lambda\,, \tag{11}$$

   where α is the searched Hausdorff–Holder–Hurst exponent.

The DMA technique is less complicated and seems to be faster in practical application than DFA algorithm. But does it give the same results as DFA? So far no final conclusion has been reached. This article contributes to the above area of interest.

2. DMA–DFA comparison study

Preliminary results obtained for real financial series [6] suggest that $\alpha_{\rm DMA}$ values are lower than corresponding $\alpha_{\rm DFA}$ results. Also for the set of artificial time series of length $L \sim 2^{18}$ constructed with the use of Random Midpoint Displacement (RMD) algorithm one finds $\alpha_{\rm DFA} \sim \alpha_{\rm DMA} + 0.05$ [5] what supports the existence of systematic displacement between DFA and DMA results, at least for series of length mentioned above. On the other hand, in many practical applications the length of time series we deal with is shorter (*e.g.* finance, biology, genetics, medicine), especially if one looks at the *local α exponent* value rather than the global one [9].

To attack the problem of mutual dependence between DMA and DFA results we looked first at the set of artificial arithmetic integer Brownian time series of length $L = 3 \times 10^4$ with discrete time interval $\Delta t = 1$, *i.e.*:

$$x(L\Delta t) = x_0 + \sum_{k=1}^{L} \Delta x_k \,, \tag{12}$$

where Δx_k $(k = 1, \ldots, L)$ are centered and normalized random variables taken from normal distribution and generated by random number generator.

The obtained results for $\alpha_{\rm DMA}$ and $\alpha_{\rm DFA}$ exponents for such series were distributed randomly and no systematic difference between α's measured with two techniques for the same series was found. Two cases with opposite relation $\alpha_{\rm DMA}$ *versus* $\alpha_{\rm DFA}$ are shown in Fig. 2(a), (b). In one case $\alpha_{\rm DFA} > \alpha_{\rm DMA}$ and $\alpha_{\rm DFA} - \alpha_{\rm DMA} = 0.02$, in the other one $\alpha_{\rm DFA} < \alpha_{\rm DMA}$ and $\alpha_{\rm DMA} - \alpha_{\rm DFA} = 0.04$.

This forced us to treat the problem statistically, *i.e.* we decided to find statistical distributions of Hurst exponents measured within two methods for artificial series of various length. It seemed to us interesting to compare two statistics together and to work out correlations between scaling exponents measured within DMA and DFA techniques for the same sample of time series.

We took for this purpose samples of arithmetic Brownian time series of length L in the range $10^2 - 10^5$. Each sample contained $N \sim 65000$ series of fixed length. We covered uniformly the whole range of L in log-scale keeping $L \sim L_0 q^n$. The approximate log step $q \sim 3/2$ has been used to create variety of lengths.

For any sample of fixed length series the averaged scaling range $\langle\tau\rangle$ or $\langle\lambda\rangle$ has been calculated for defined number of candidates (~ 30) with its standard deviation σ_τ (σ_λ). The scaling range was taken as the range of τ or λ variables strictly obeying scaling laws of Eqs. (7), (11) and assumed to terminate respectively at $\langle\tau\rangle - \sigma_\tau$ for DFA and $\langle\lambda\rangle - \sigma_\lambda$ for DMA. Only series with fitting parameter $R^2 > 0.98$ were taken into account for α exponent extraction.

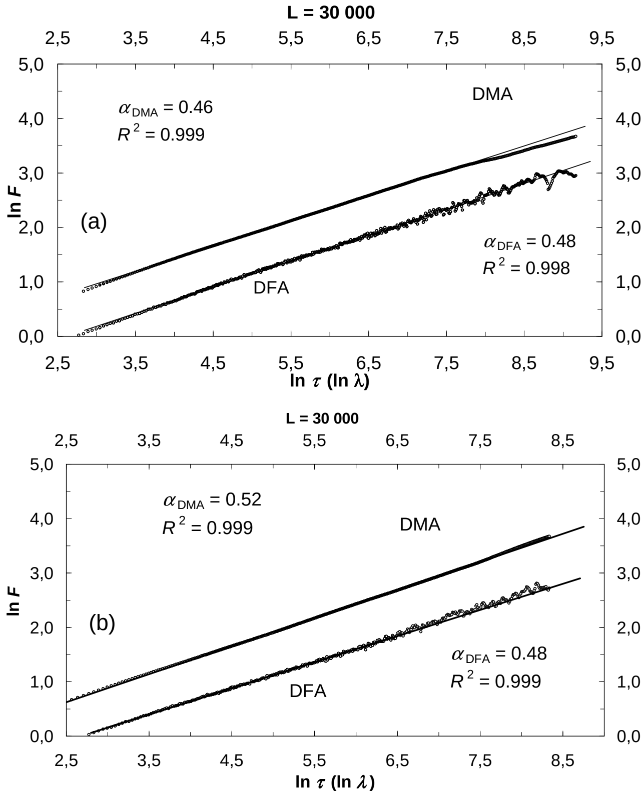

Fig. 2. Example of DFA and DMA α exponent fit for artificial Brownian time series of length $L = 30000$, where $\alpha_{\mathrm{DFA}} > \alpha_{\mathrm{DMA}}$ — (a), $\alpha_{\mathrm{DFA}} < \alpha_{\mathrm{DMA}}$ — (b).

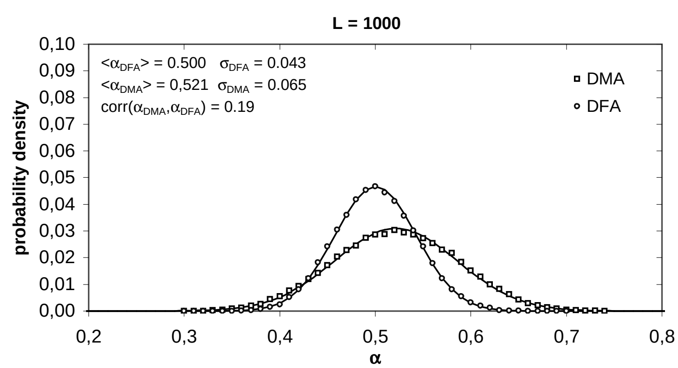

Fig. 3. Probability density of scaling α exponents obtained with the use of DFA (circles) and DMA (squares) techniques for the sample of 65000 time series of length $L = 1000$. The normal distribution fit with its parameters is also shown and marked as a solid line.

For any sample of time series a statistical distribution of α_{DFA} and α_{DMA} has been built. Examples of such distributions for short, medium and long time series are shown in Figs. 3, 4, 5.

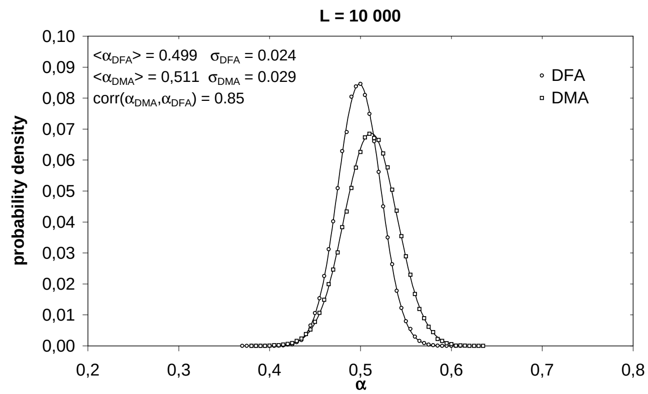

Fig. 4. Probability density as in Fig. 3, but for a sample of series with length $L = 10000$.

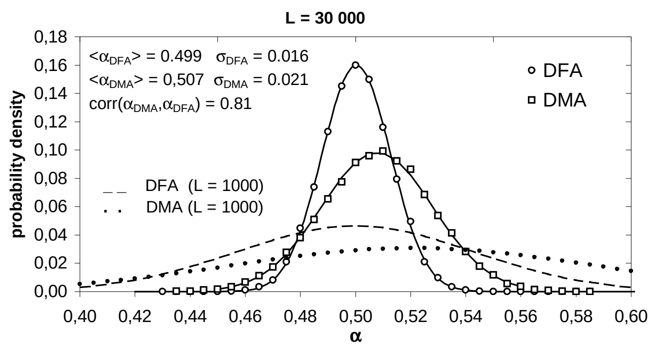

Fig. 5. The same as in Fig. 4, but for the sample of series with length $L = 30000$. Additional lines represent $L = 1000$ normal fit drawn for comparison in the same scale.

The first observation is that for any length L both distributions fit very well normal distributions, but with different parameters. In particular, the standard deviation σ_{DFA} of α_{DFA} scaling parameters is always smaller than the corresponding standard deviation σ_{DMA} of α_{DMA} exponents. Moreover, both standard deviations decrease when L grows. This may be explained in terms of different sensitivity of DFA and DMA techniques to the presence of random autocorrelations in time series. Such autocorrelations are naturally randomly distributed in any sample of generated time series and hence a distribution of α exponent is normal. The probability of random autocorrelations is bigger for short time series, where all statistical fluctuations manifest in a more vivid way. When L increases, their influence on the presumed global autocorrelation in series can be neglected. Therefore, both standard deviations $\sigma_{\rm DFA}$ and $\sigma_{\rm DMA}$ drop with increasing L. However, we always observe $\sigma_{\rm DFA} < \sigma_{\rm DMA}$, what indicates that DMA technique is more sensitive to such "autocorrelation noise" than DFA one.

One may look at this problem also from another side — like in Fig. 6. Here we have drawn several plots of DFA and DMA analysis, *i.e.* $\ln F$ *versus* $\ln\tau$ or $\ln\lambda$ plots for several corresponding artificial Brownian series of length $L = 1000$. It is seen that deviations from the strict power law behavior, if occur, are more drastic for DMA than for DFA case and the dispersion of produced slopes is also larger for DMA than for DFA, despite the fact that DMA plots are more smooth in comparison with DFA ones.

The next observation concerns the mean values. We have got $\langle\alpha_{\rm DFA}\rangle_N < \langle\alpha_{\rm DMA}\rangle_N$ for all L, where $\langle.\rangle_N$ is taken over a sample of N time series. A clear shift of the central DMA values to the right with respect to DFA ones (see Figs. 3–5) does not suggest however that any systematic relation between $\alpha_{\rm DMA}$ and $\alpha_{\rm DFA}$ exists. Indeed, evaluating the correlation factor

$$\mathrm{corr}\,(\alpha_{\rm DFA}, \alpha_{\rm DMA}) = \frac{\langle\alpha_{\rm DFA}\alpha_{\rm DMA}\rangle_N - \langle\alpha_{\rm DFA}\rangle_N\langle\alpha_{\rm DMA}\rangle_N}{\sigma_{\rm DFA}\sigma_{\rm DMA}} \tag{13}$$

we found it increasing with L, nevertheless it never indicates full or almost full correlation. It is maximal for large L, where $\mathrm{corr}\,(\alpha_{\rm DFA}, \alpha_{\rm DMA}) \sim 0.8$ for $L \sim 10^4$–10^5.

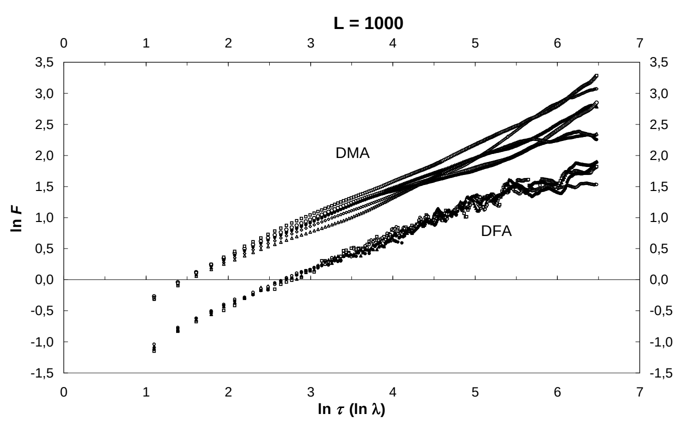

Fig. 6. Examples of DMA and corresponding DFA plots $\ln F$ *versus* $\ln\tau(\ln\lambda)$ for several randomly chosen Brownian integer time series of length $L = 1000$.

This can be graphically illustrated in Figs. 7, 8, where a correlation plot $\alpha_{\rm DFA}$ *versus* $\alpha_{\rm DMA}$ is shown for Hurst exponent values obtained for $L = 10000$ and $L = 30000$ series. We notice that DMA gives higher values than DFA method in most series. This result is independent on the length of time series. In fact the percentage excess (n_+) of cases where $\alpha_{\rm DMA} > \alpha_{\rm DFA}$ over the cases where $\alpha_{\rm DMA} < \alpha_{\rm DFA}$ (n_-), *i.e.*:

$$\delta_\pm = \frac{n_+ - n_-}{n_+ + n_-} \tag{14}$$

changes from $\sim 20\%$–25% for series with $L < 10000$ up to $\sim 50\%$ for longer series.

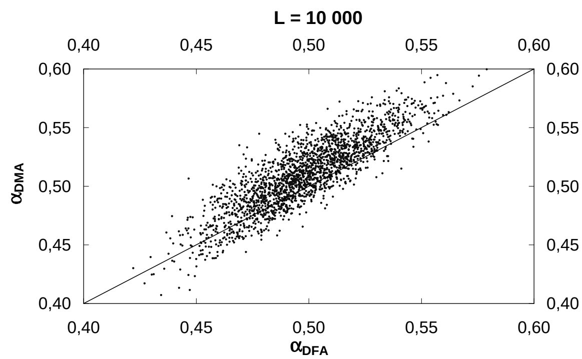

Fig. 7. Correlation plot $\alpha_{\rm DFA}$ *versus* $\alpha_{\rm DMA}$ for the sample of 65000 Brownian time series of length $L = 10000$.

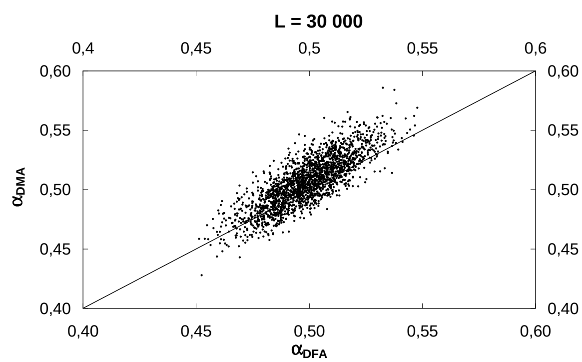

Fig. 8. Correlation plot $\alpha_{\rm DFA}$ *versus* $\alpha_{\rm DMA}$ for the sample of 65000 Brownian time series of length $L = 30000$.

It implies that the mean of difference $\delta_{\rm DFA-DMA}$, where

$$\delta_{\rm DFA-DMA} = \langle \alpha_{\rm DFA} - \alpha_{\rm DMA} \rangle_N \tag{15}$$

is not a good measure of 'distance' between two investigated methods. It is more convenient therefore to define this distance in a standard way, *i.e.*:

$$\Delta_{\mathrm{DFA-DMA}} = \sqrt{\langle(\alpha_{\mathrm{DFA}} - \alpha_{\mathrm{DMA}})^2\rangle_N}\,. \tag{16}$$

We worked out the sufficient number of time series samples of various length to find a relationship $\Delta_{\mathrm{DFA-DMA}}(L)$. The line of the best fit for collected data is drawn in Fig. 9. This plot indicates that the average displacement between α_{DFA} and α_{DMA} exponents for a given time series ranges from 15% for series with $L \leq 10^3$, down to 2% for long series ($L \sim 10^5$). The latter value is much smaller than one reported in [5]. The fastest drop in DFA–DMA distance is observed for medium length series, *i.e.* when $L \sim 10^3 - 10^4$. For such series $\Delta_{\mathrm{DFA-DMA}}$ makes on the average $\sim 10\%$ of α_{DFA} value.

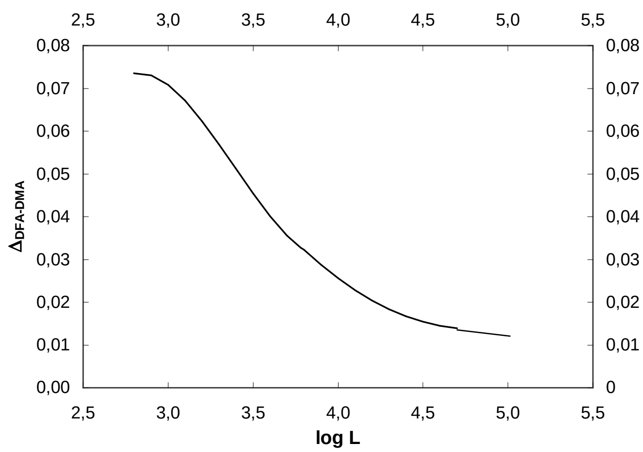

Fig. 9. A mean difference $\Delta_{\mathrm{DFA-DMA}}$ between α_{DFA} and α_{DMA} exponents calculated for the same series as a function of the series length L.

This difference might be of interest if more detailed study of α exponent is required for more detailed predictions to be made(*e.g.* heart diseases, finances, *etc.*). The plot in Fig. 9 also suggests that $\Delta_{\mathrm{DFA-DMA}} \to 0$ when $L \to \infty$. The latter cases have not been explored by us.

3. Conclusions

We report from the analysis of artificial Brownian integer time series and from the collected data that, on the average, DMA method overestimates Hurst exponent values in comparison with DFA technique. This result contradicts to some previous hypothesis in literature. The DMA method seems to be also more sensitive to the presence of random fluctuations in autocorrelations in time series than DFA analysis does. In many practical situations it is an disadvantage leading to the false signal of not existing, global, noise free autocorrelations in time series.

The mean distance between two methods, *i.e.* the mean difference between $\alpha_{\rm DFA}$ and $\alpha_{\rm DMA}$ exponents calculated for the time series of given length L is a smooth decreasing function of L. For shorter series ($L \leq 6000$) this distance reaches $\sim 15\%$ what might be important in precise determination of α exponent for such series.

There are some open questions. It is not exactly clear where the scaling law exactly starts or terminates, so one needs a more strict requirements how the scaling range should be determined for DFA and DMA techniques and how uncertainties in the choice of scaling range are related to uncertainties in the scaling exponent α. This work is now in progress [14].

REFERENCES

[1] A.-L. Barabasi, H.E. Stanley, *Fractal Concepts in Surface Growth*, Cambridge University Press, Cambridge 1995.

[2] B.J. West, B. Deering, *The Lure of Modern Science: Fractal Thinking*, World Scientific, Singapore 1995.

[3] P.S. Addison, *Fractals and Chaos*, Institute of Physics, Bristol 1997.

[4] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Golberger, *Phys. Rev.* **E49**, 1685 (1994).

[5] E. Alessio, A. Carbone, G. Castelli, V. Frappietro, *Eur. Phys. J.* **B27**, 197 (2002).

[6] A. Carbone, G. Castelli, *Noise in Complex Systems and Stochastic Dynamics*, Proc. of SPIE Vol. **5114**, 406 (2003).

[7] A. Carbone, G. Castelli, `cond-mat/0303465`.

[8] L. Xu, P.Ch. Ivanov, K. Hu, Z. Chen, A. Carbone, H.E. Stanley, *Phys. Rev.* **E**, (2005); `cond-mat/0408047`.

[9] N. Vandewalle, M. Ausloos, *Physica A* **246**, 454 (1997).

[10] M. Ausloos, N. Vandewalle, Ph. Boveroux, A. Minguet, K. Ivanova, *Physica A* **274**, 229 (1999).

[11] N. Vandewalle, M. Ausloos, *Int. J. Comput. Anticipat. Syst.* **1**, 342 (1998).

[12] M. Ausloos, *Physica A* **285**, 48 (2000).

[13] D. Grech, Z. Mazur, *Physica A* **336**, 133 (2004).

[14] D. Grech, Z. Mazur, in preparation.