Physica A 451 (2016) 63–69

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Evenly spacing in Detrended Fluctuation Analysis

Zainy M.H. Almurad [a,b], Didier Delignières [a,*]

[a] EA 2991 Euromov, University of Montpellier, France

[b] Faculty of Physical Education, University of Mossul, Iraq

HIGHLIGHTS

- We analyze the effects of evenly spacing in Detrended Fluctuation Analysis (DFA).
- We compare evenly vs. logarithmic spacing with synthetic signals.
- Evenly spacing decreases by 36% the dispersion of estimates.
- Evenly spacing could allow the analysis of shorter time series.
- Evenly spacing should be systematically adopted in the application of DFA.

ARTICLE INFO

Article history:
Received 30 August 2015
Received in revised form 29 November 2015
Available online 1 February 2016

Keywords:
Detrended Fluctuation Analysis
Power laws
Diffusion plot
Evenly spacing

ABSTRACT

Detrended Fluctuation Analysis is a widely used method, which aims at assessing the level of self-similarity in time series. This method analyzes the diffusion properties of the signal, by computing the linear regression slope in the diffusion plot, representing in log–log coordinates the relationship between the variability of the signal and the length of the intervals over which this variability is computed. We compare in this paper the results obtained with logarithmically spaced and evenly spaced diffusion plots. The study shows the substantial benefits of evenly spacing, especially in the reduction of the variability of estimation.

1. Introduction

The Detrended Fluctuation Analysis (DFA), initially introduced by Peng et al. [1], is a widely used analysis method which aims at determining the level of self-similarity in a time series. A better understanding of this method supposes a short introduction to the underlying model.

The DFA algorithm is based on the diffusion property of fractional Brownian motion (fBm), a family of stochastic processes introduced by Mandelbrot and Van Ness [2]. Here we focus on the discrete version of fBm, which corresponds to the nature of the series analyzed in experimental research. In such process variance is a power function of the length (n) of the interval over which it is computed. Consider a process x_i:

$$Var(x_i) \propto n^{2H}, \tag{1}$$

where H is the Hurst exponent, which can take any real value within the interval]0, 1[. For $H = 1/2$, x_i corresponds to ordinary Brownian motion, and variance is just proportional to the elapsed time (normal diffusion). For $H \neq 1/2$, x_i is characterized by anomalous diffusion: The process is said subdiffusive for $H < 1/2$, and superdiffusive for $H > 1/2$.

* Correspondence to: Euromov, UFR STAPS, 700 avenue du Pic Saint Loup, 34090 Montpellier, France.
E-mail address: didier.delignieres@univ-montp1.fr (D. Delignières).

http://dx.doi.org/10.1016/j.physa.2015.12.155
0378-4371/

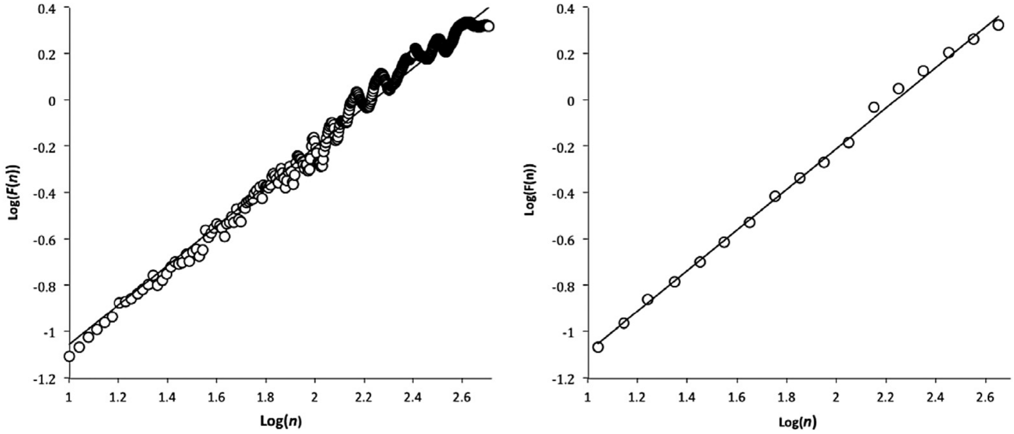

Fig. 1. Two example diffusion plots, obtained with the same series. The left panel represents the logarithmically spaced plot, and the right panel an evenly spaced plot. The slopes of the regression lines are 0.85 and 0.88, respectively.

A second family of stochastic processes, fractional Gaussian noise (fGn), is defined as the series of increments within a fBm. By definition, a fGn is the series of differences in a fBm, and conversely the summation of a fGn gives a fBm. Each fBm series is then related to a specific fGn, and both are characterized by the same H exponent. H determines the nature of correlations between successive values in the fGn: for $H < 0.5$, successive values are negatively correlated, and the series is said to be anti-persistent. Conversely for $H > 0.5$ successive values are positively correlated, and the series is persistent. For $H = 0.5$, successive values are uncorrelated, and the series corresponds to white noise. An important difference between these two classes of processes is that fBm are non-stationary processes, as suggested by Eq. (1), whereas fGn series present stationary mean and variance over time.

As presented in the first paragraphs of this article, fGn and fBm are defined as two distinct families, which can be considered superimposed, with their relationships of summation/differencing. A number of authors, however, have proposed to consider these two families as a continuum, surrounding the mythical border of "ideal" $1/f$ noise [3–6].

The *Detrended Fluctuation Analysis* (DFA), Ref. [1] can be applied to both fGn and fBm signals. The details of the DFA algorithm will be detailed later in this paper. Here we just present its general principles, in order to introduce the hypotheses underlying the present work. This method exploits a typical scaling law, which states that the standard deviation of the integrated series is a power function of the interval length over which it is computed, with an exponent α. Considering a time series x_i:

$$\begin{cases} X_i = \sum_{k=0}^{i} x_k \\ SD(X_i) \propto n^{\alpha}. \end{cases} \tag{2}$$

This scaling just derives from the original definition of fBm, expressed in Eq. (1). fGn series are characterized by α exponents ranging from 0 to 1, and fBm by exponents ranging from 1 to 2. $\alpha = 1$ corresponds to $1/f$ noise. If the series x_i is a fGn, X_i is the corresponding fBm and α is the Hurst exponent. If x_i is a fBm, X_i belongs to another family of over-diffusive processes, characterized by α exponents ranging from 1 to 2, and in that case $\alpha = H + 1$ [1].

The DFA algorithm works as follows: The series is first integrated, and then this integrated series is divided into non-overlapping intervals of length n. Within each interval the series is detrended, and the standard deviation of the residuals is computed. Then one calculates the average (detrended) standard deviation for the intervals of length n, noted $F(n)$. This computation is repeatedly performed over all n values. In practice, the shortest interval length is chosen for allowing a valid estimate of standard deviation (for example $n = 10$), and the lengthiest for allowing at least two distinct estimates (e.g., $n = N/2$). Then $F(n)$ is plotted against n in log–log coordinates, forming the so-called *diffusion plot*, and the exponent α is obtained as the slope of the linear regression (see Fig. 1, left panel).

An important consequence of this logarithmic transformation is that the density of points along the abscissa axis increases as interval length increases (see Fig. 1, left panel). And as regression analysis works on plane geometric principles, the weight of long intervals in the calculation of the slope becomes higher than that of short intervals. Moreover, the long-term region of the diffusion plot often presents irregularities around the linear trend, especially because the number of intervals involved in the computation of $F(n)$ is smaller for long time scales [7]. This results in a greater uncertainty in the determination of the slope in the long-term region of the diffusion plot, where most points are concentrated. A solution for overcoming this problem is to define the diffusion plot over a set of points evenly spaced in the logarithmic scale (Fig. 1, right panel).

Several methods have been proposed for obtaining a series of evenly spaced points in the log–log plot. Some authors have proposed to select a set of interval lengths, evenly spaced on the logarithmic scale. This idea was initially introduced by