

The Effectiveness Evaluation of Two Kinds of Fractal Sequences on Detrended Fluctuation Analysis

Danying Xie^{1, #}, Li Wan^{1, 2}, Yongqiang Zhu¹

1. School of Mathematics and information Science, Guangzhou University, Guangzhou, 510006, China

2. Key Laboratory of Mathematics and Interdisciplinary Sciences Guangzhou Higher Education Institutes, Guangzhou University, Guangzhou, 510006, China

#Email: danying_xie@163.com

Abstract

Used of the fractional Brownian motion and fractional Gaussian noise sequence, the detrended fluctuation analysis (DFA) applied to estimate the Hurst exponent to verify the stability and dependability of the method by changing the data length and regression trend order. The result shows that the Hurst exponent estimate is stable and efficient with the length of data for fractional Brownian motion and fractional Gaussian noise sequence. The influence on the Hurst exponent is not obvious when the regression trend order was changed, and the estimate accuracy is improved with the increasing of Hurst exponent value.

Keywords: Fractional Brownian Motion; Fractional Gaussian Noise; Detrended Fluctuation Analysis; Stability

1 INTRODUCTION

Fractional time series is the sequence which has statistical distribution similarity under the different time scales and is characterized by self-similarity and long-range correlation (long-term memory) in the time or space domain. Fractional Brownian motion (FBM) was firstly put forward by an American mathematician, B.B. Mandelbrot, and developed from general Brownian motion^[1]. It is a continuous non-stationary Gaussian stochastic process that mean function is zero, but endowed with stationary increments, its first-order differential sequence is called fractional Gaussian noise (FGN), a stationary time series. Quantitative description of the degree of their self-similarity and long-range correlation is Hurst exponent (Hurst parameter), H values correspond to correlation coefficient $R(t)$ of past and future increment of zero hour, reflecting the autocorrelation degree of the sequence.

When $0 < H < 0.5$, the correlation coefficient $R(t)$ is negative, indicating that the sequence was negatively correlated, and successive increments tend to have opposite signs. When $H = 0.5$, the correlation coefficient $R(t)$ is 0, indicating that was completely irrelevant. When $0.5 < H < 1$, the correlation coefficient $R(t)$ is positive, indicating that the sequence was positively correlated and long-range correlation, successive increments are more likely to have the same signs.

Currently, the value of Hurst exponent is estimated by various methods such as detrended fluctuation analysis(DFA), rescaled range analysis(R/S), modified rescaled range analysis(MR/S), wavelet analysis, and so on^[2]. Among them, detrended fluctuation analysis is a common way which proposed by Peng in 1994 in order to study the correlation between DNA bases and sort order^[3]. This method overcomes the shortcomings of R/S; it can effectively estimate the scaling exponent and describe long-range correlation of the time series. Hence, DFA is rapidly and broadly applied in many different fields, for instance, DNA sequences, equipment failures, stock market, physics, meteorology, and geology^[4-5]. If the order of polynomial fitting in detrended process is m , it is denoted as DFA m .

Sequences can be used to describe the natural and social phenomena, for example, price fluctuation of stock market, fluctuation of heart rates and brain wave, and the noise in electronic components, natural landscapes^[6-8]. Therefore, and the constant further researches of fractal time series have important practical significance. By setting different

data length of two kinds of fractal time series and different regression trend order of DFA, we investigate the influence on the estimation of Hurst exponent and then explore the effectiveness of DFA with respect to fractal time series.

2 DETRENDED FLUCTUATION ANALYSIS

Detrended fluctuation analysis is a method for quantitatively measuring the extent of long-range correlation in time series and estimating the scaling exponent, it can be calculated as follows [9-10]:

Given a bounded time series $\{x_i\}$ ($i=1, 2, \dots, N$), N is the length of the series.

(i) We integrate the series and cumulative sum $Y(i)$ is obtained:

$$Y(i) = \sum_{k=1}^i (x_k - \bar{x}), i = 1, 2, \dots, N. \quad (1)$$

(ii) $Y(i)$ is divided into N_s non-intersecting boxes of equal length s , where $N_s \equiv \text{int}(N/s)$. N is not always divisible by s , so we repeat the segmentation process from the tail of the series for avoiding discarding any remainder, and then gets $2N_s$ boxes of equal length s .

(iii) In each box, we fit the series $Y(i)$ at least squares principle by using a polynomial function $P_j^m(i)$, which is called the local trend, where $j=1, 2, \dots, N_s$, m is the regression trend order. If order- m polynomial function is applied for fitting, denoted as DFA m .

(iv) By subtracting the local trend $P_j^m(i)$, detrended fluctuation function is calculated by

$$Y_s(i) = Y(i) - P_j^m(i), i = 1, 2, \dots, s. \quad (2)$$

(v) The root-mean-square fluctuation of detrended fluctuation function $Y_s(i)$ is calculated by

$$F(s) = \left[\frac{1}{2N_s} \sum_{v=1}^{2N_s} F^2(v, s) \right]^{1/2}, \quad (3)$$

where

$$F^2(v, s) = \frac{1}{s} \sum_{i=1}^s \{Y[(v-1)s + i] - P_j^m(i)\}^2, v = 1, 2, \dots, N_s, \quad (4)$$

or

$$F^2(v, s) = \frac{1}{s} \sum_{i=1}^s \{Y[N - (v - N_s)s + i] - P_j^m(i)\}^2, v = N_s + 1, N_s + 2, \dots, 2N_s. \quad (5)$$

If the sequence has long-range correlation, there exists exponential relation between $F(s)$ and s , that is

$$F(s) \propto s^\alpha, \quad (6)$$

with α is the scaling exponent, calculated as the slope of a straight line fit to the log-log graph of s against $F(s)$. If the time series is stationary:

$$\alpha = H; \quad (7)$$

If the time series is non-stationary:

$$\alpha = H + 1. \quad (8)$$

3 INFLUENCES ON H ESTIMATE BY DFA

3.1 Generation of Fractional Gaussian Noise and Fractional Brownian Motion

Fractional Gaussian noise (FGN) and fractional Brownian (FBM) motion are not only two important stochastic processes, but also typical fractal time series. FGN is stationary, whose standard deviation does not vary as the time goes, while FBM is non-stationary, whose standard deviation assumes the trend of escalation.

Fractional Brownian motion is a self-similar Gaussian process and can be regarded as the result of fractal integral of

general Brownian motion, its Hurst exponent $B_H(t)$ has the following characteristics ^[11]: $B_H(t)$ has stationary increments; $B_H(t)=0$, $EB_H(t)=0$, $t>0$; $E[B_H(t)]^2=t^{2H}$, $t>0$; Finite dimensional distribution $B_H(t)$ is multidimensional Gaussian distribution. The covariance function of $B_H(t)$ is given by

$$R_B(t,s)=\frac{1}{2}\left[|t|^{2H}+|s|^{2H}-|t-s|^{2H}\right], 0<H<1. \quad (9)$$

It is clear from formula (9) that correlation function depends on time delay $t-s$ and the starting point t . Consequently, FBM is non-stationary time series. In this paper, the FBM sequence is generated by the function of *wfbm* (H, N) in MATLAB, based on the method of wavelet transform. The parameter H represents the Hurst exponent of fractional Brownian motion and N is the data length.

First-order difference sequence of FBM is called the fractional Gaussian noise, denoted as FGN. Discrete fractional Gaussian noise is expressed as

$$G_H(t)=B_H(t+1)-B_H(t), \quad (10)$$

This is a stationary Gaussian process with mean zero and covariance function is given by

$$R_G(t)=\frac{1}{2}\left[|t+1|^{2H}+|t-1|^{2H}-2|t|^{2H}\right]. \quad (11)$$

The formula (11) proves that the autocorrelation function is only related to t , thus FGN is stationary time series. Random midpoint displacement (RMD) is one of the useful methods to generate FGN, whose basic idea is to subdivide the interval $[0, T]$ constantly, get the arithmetic average value of left and right endpoint of each interval, and then iterate and calculate the midpoint value, in the end, add an offset $\sqrt{1-2^{2H-2}}$. FGN sequence in this paper is produced by self-complied program of RMD method in MATLAB ^[12-14].

3.2 Influences on Hurst Exponent by Data Length

Given theoretical value of H ($H=0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9$), 10 sets FBM and FGN of each data length N ($N=2^9, 2^{10}, 2^{11}, 2^{12}$) were produced in MATLAB. We obtained the scaling exponent α with detrended fluctuation analysis method, and then calculated the H value. Finally, the mean and standard deviation of 10 sets data were computed. The result is shown in Table 1.

TABLE 1 MEAN AND STANDARD DEVIATION OF H ESTIMATE OF FBM AND FGN IN DIFFERENT DATA LENGTH

H	Sequence	$N=2^9$ Mean $\pm$ SD	$N=2^{10}$ Mean $\pm$ SD	$N=2^{11}$ Mean $\pm$ SD	$N=2^{12}$ Mean $\pm$ SD
0.6	FBM	0.5583 $\pm$ 0.0894	0.5788 $\pm$ 0.059	0.6071 $\pm$ 0.0380	0.5993 $\pm$ 0.0373
	FGN	0.6268 $\pm$ 0.0787	0.583 $\pm$ 0.0379	0.6206 $\pm$ 0.0318	0.5933 $\pm$ 0.0241
0.65	FBM	0.5982 $\pm$ 0.1025	0.6376 $\pm$ 0.0831	0.6387 $\pm$ 0.0473	0.6424 $\pm$ 0.0394
	FGN	0.7037 $\pm$ 0.0632	0.6254 $\pm$ 0.0305	0.6308 $\pm$ 0.0289	0.6526 $\pm$ 0.0242
0.7	FBM	0.6872 $\pm$ 0.0928	0.6975 $\pm$ 0.0726	0.6969 $\pm$ 0.0549	0.6914 $\pm$ 0.0332
	FGN	0.6639 $\pm$ 0.0492	0.666 $\pm$ 0.0488	0.6797 $\pm$ 0.0342	0.6853 $\pm$ 0.015
0.75	FBM	0.7819 $\pm$ 0.0990	0.7234 $\pm$ 0.0689	0.7241 $\pm$ 0.04	0.7393 $\pm$ 0.0311
	FGN	0.7037 $\pm$ 0.0632	0.7257 $\pm$ 0.0497	0.7263 $\pm$ 0.0384	0.7351 $\pm$ 0.0346
0.8	FBM	0.7469 $\pm$ 0.0604	0.7744 $\pm$ 0.0603	0.7751 $\pm$ 0.0553	0.8034 $\pm$ 0.0335
	FGN	0.7403 $\pm$ 0.0965	0.7788 $\pm$ 0.0577	0.7887 $\pm$ 0.0396	0.791 $\pm$ 0.0259
0.85	FBM	0.7796 $\pm$ 0.1064	0.8334 $\pm$ 0.0403	0.8340 $\pm$ 0.0377	0.8516 $\pm$ 0.0328
	FGN	0.8178 $\pm$ 0.056	0.8178 $\pm$ 0.0503	0.835 $\pm$ 0.0326	0.846 $\pm$ 0.0243
0.9	FBM	0.8144 $\pm$ 0.0638	0.8865 $\pm$ 0.0509	0.8637 $\pm$ 0.0402	0.879 $\pm$ 0.0147
	FGN	0.8271 $\pm$ 0.0994	0.8722 $\pm$ 0.0493	0.8779 $\pm$ 0.0373	0.8886 $\pm$ 0.0256

With respect to non-stationary time series FBM and stationary time series FGN, the mean and standard deviation of H value apparently varies with the data length in Table 1. The mean of H value of both fractal time series are gradually getting closer to theoretical value of H with the increasing of data length. Furthermore, the standard deviation of H value is becoming smaller and smaller, indicating that the stability and validity of Hurst exponent by

detrended fluctuation analysis becomes greater with data lengthened.

3.3 Influences on Hurst Exponent by Regression Trend Order

Given fixed data length $N (N=2^{10})$, we utilized MATLAB to generate 10 sets FBM and FGN of each theoretical value of H ($H=0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9$). Under the situation of different regression trend order- m ($m=1, 2, 3, 4, 5$), order- m detrended fluctuation analysis (DFAm) is applied to get the scaling exponent α and H value, and calculate the mean, standard deviation and the range of mean.

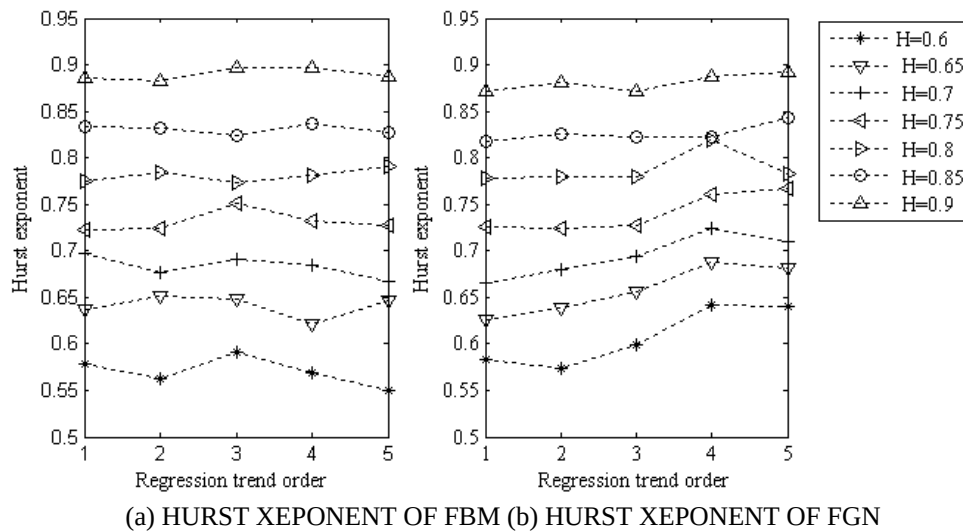


FIG. 1 EFFECTS ON H VALUE OF TWO FRACTAL TIME SERIES BY REGRESSION TREND ORDE

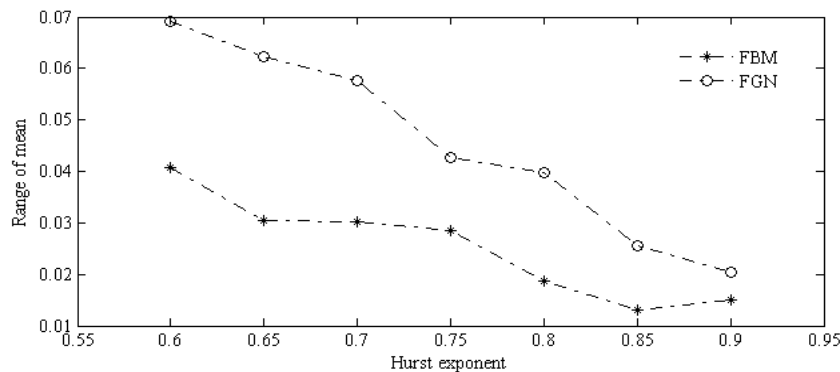


FIG.2 RANGE OF MEAN OF H VALUE OF TWO FRACTAL TIME SERIES

How the regression trend order m affects Hurst exponent is shown in Fig. 1. It can be seen that when the regression trend order m is changed, H value of two fractal time series always fluctuates around the theoretical value, which indicates that the nonlinearity of sequence is not strong, and the influence on the Hurst exponent is not distinct from both stationary and non-stationary time series. Fig. 2 reveals the relation between the theoretical value of H and range of mean. We have found that the range of mean is obviously becoming smaller with the increasing of theoretical value; it means the fluctuation range of H value is getting smaller, thus the estimate of accuracy is improved, and the influence on H value is obviously decreasing.

4 CONCLUSIONS

Detrended fluctuation analysis method is used for non-stationary time series FBM and stationary time series FGN to obtain the scaling exponent α , and then estimate the Hurst exponent H and figure out the mean and standard deviation of H value. By changing the data length and regression trend order, we investigate and analyze their influences on H value. The result shows that, the longer the data length is, the greater the stability and validity of Hurst exponent estimate have. While the influence of regression trend order on H value is not obvious, in addition,

the influence is significantly smaller with increasing theoretical value.

ACKNOWLEDGMENT

This research is supported by the National Nature Science Foundation of China (Grant No.41172295).

REFERENCES

- [1] B. B. Mandelbrot. The fractal geometry of nature. San Francisco: W. H. Freeman, 1982
- [2] J. Chen, X. H. Tan, Z. Jia. "Performance analysis of seven estimate algorithms about the Hurst coefficient." Computer Applications, 26 (2006): 945-950
- [3] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, A. L. Goldberger. "Mosaic organization of DNA nucleotides." Physical Review E, 49 (1994): 1685-1689
- [4] H. Xiao, Y. R. Li, Y. Lv. "Detrend Fluctuation Analysis and Application on Fault Recognition of Gear." Coal Mine Machinery, 29 (2008): 202-204
- [5] C. Huang, H. Q. Gong, W. J. Zhong. "Piecewise Detrended Fluctuation Analysis Method and Empirical Study on China Stock Market." Operations Research and Management Science, 18 (2009): 126-130
- [6] Y. C. Qin, K. Liu. "Advancement of Applied Studies of Fractal Theory in Geography." Progress in Geography, 22 (2003): 426-436
- [7] X. H. Jia, X. R. Li, Y. S. Li. "Fractal dimension of soil particle size distribution during the process of vegetation restoration in arid sand dune area." Geographical Research, 26 (2007): 518-525
- [8] A. W. Lo. "Long-term memory in stock market prices." Econometrica, 59 (1999): 1279-1313
- [9] C. K. Peng, J. Mietus, J. M. Hausdorff, S. Havlin, H. E. Stanley, A. L. Goldberger. "Long-range anticorrelations and non-gaussian behavior of the heart beat." Phys Rev Lett, 70 (1993): 1343-1346
- [10] C. K. Peng, S. Havlin, H. E. Stanley, A. L. Goldberger. "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series." Chaos, 5 (1995): 82-87
- [11] Z. X. Liu, G. L. Li, D. Y. Zhang. "Spectrum Density Application in Synthesis and Estimation for Fractional Gaussian Noise." Microelectronics and Computer, 24 (2007): 37-44
- [12] B. B. Mandelbrot. "A fast fractional Gaussian noise generator." Water Resources Research, 7 (1971): 543-553
- [13] L. Liu. "Monte Carlo Simulation Pricing Based on Summation of Fractional Gaussian Noise." Transactions of Beijing Institute of Technology, 31 (2011): 627-630
- [14] F. G. Niu, W. Q. Liu. "Relations of Fractional Brownian Motion and Hurst Exponent." Journal of Shanxi University, 33 (2010): 380-383

AUTHORS

¹**Danying Xie** (1989-), is a female, graduate student. Born in Jieyang, Guangdong Province, China. Currently, she is studying in school of mathematics and information Science of Guangzhou University, China. Her specialty is probability and mathematical statistics and the major research direction is nonparametric statistics and applied mathematics.
Email: danying_xie@163.com

²**Li Wan** (1961-) is a female, professor. From Wuhan, Hubei Province, China. Her major research field is applied mathematics.
Email: wanli03100@tom.com

³**Yongqiang Zhu** (1990-), male, graduate student. He is from Ganzhou, Jiangxi Province, China. And now he is studying in school of mathematics and information Science of Guangzhou University. His main research field is applied mathematics.
Email: umasou07@sina.com