國立中山大學 應用數學 學系（研究所）

碩士論文

自我相似過程之參數估計及適合度檢定之研究

研究生：蔣佩蓉 撰

指導教授：羅夢娜 教授

中華民國 九十五 年 七 月

摘要

近來有些研究報告顯示生理資料具有長相關和自我相似的特性。此二特性可分別用長相關參數 d 和自我相關係數 H 量化來表示。Peng（1995）藉由分類所得到的心律資料進行分析，研究具有致命病變者其長相關係數的特性。分數布朗運動（Fractional Brownian Motion，簡稱 FBM）和分數差分ARMA（Fractional ARIMA，簡稱 FARIMA）是兩個著名具有自我相似特性的隨機過程，我們有興趣了解自我相似過程，是否適用於對心率的資料建模，以用來了解病人的健康狀況。本文利用 Jones 和 Shen（2004）所提出的嵌入分歧過程（Embedded Branching Process，簡稱 EBP）方法估計參數 H，以及利用自我相似過程最適度檢定，對模擬的 FBM 和 FARIMA 過程做檢定，來討論其適用性，並進一步修訂此檢定量之分佈。最後，針對模擬的 FARIMA 過程和從高雄榮總醫院得到的心率資料，比較不同估計方法求得的參數 H。

關鍵字：自我相似過程、Hurst 參數、嵌入分歧過程、 R/S 方法、趨勢波動分析方法、分數差分 ARMA、 I（d）過程、分數布朗運動

A Study on the Estimation of the Parameter and Goodness of Fit Test for the Self-similar Process

by

Pei-Jung Chiang

Advisor

Mong-Na Lo Huang

Department of Applied Mathematics
National Sun Yat-sen University
Kaohsiung, Taiwan, 804, R.O.C.

July, 2006

Contents

Abstract

Recently there have been reports that certain physiological data seem to have the properties of long-range correlation and self-similarity. These two properties can be characterized by a long-range dependent parameter d, as well as a self-similar parameter H. In Peng et al (1995), the alteration of long-range correlations with life-threatening pathologies are studied by analyzing the heart rate data of different groups of subjects. The self-similarity properties of two well-known processes, namely the Fractional Brownian Motion (FBM) and the Fractional ARIMA (FARIMA), are of interest to see if it is suitable to be used to model the heart rate data in order to examine the health conditions of some patients. The Embedded Branching Process (EBP) method for estimating parameter H and a goodness of fit test for examining the self-similarity of a process based on the EBP method are proposed in Jones and Shen (2004). In this work, the performance of the goodness of fit test are examined using simulated data from the FBM and FARIMA processes. A modification of the distribution of the test statistics under null hypothesis is proposed and has been modified to be more appropriate. Some simulation comparisons of different estimation methods of the parameter H for some FARIMA processes are also presented and applied to heart rate data obtained from Kaohsiung Veterans General Hospital.

Keywords: self-similar process, Hurst parameter, Embedded Branching Process (EBP), R/S method, Detrended Fluctuation Analysis (DFA), Fractional ARIMA (FARIMA), $I(d)$ process, Fractional Brownian Motion (FBM).

List of Tables

List of Figures

1. Introduction

Recently there have been reports that certain physiological data seem to have the properties of long-range correlation and self-similarity characterized by a long-range dependent parameter d and a self-similar parameter H respectively. In Beran (1994), a nice description about self-similarity has been given, where it is described that a geometric shape is called self-similar if the same geometric structures are observed, independently of the distance from which one looks at the shape. In Bate (1996), it has been explained in another way that while considering a superbly accurate map of the coast line of Great Britain, if we use a zoom facility to set the scale of magnification we might observe that, no matter what scale we chose, the observed images would appear similar. In Figure 1, an example with self-similar structure taken from physionet is exhibited.

In Kolmogorov (1941), self-similar processes were introduced in theoretical context. At that time, statisticians did not seem to be aware of the existence or statistical relevance of such processes until "self-similar" was first explained in statistics by Mandelbrot et al. (1969a, b, c). In Feder (1988), it has mentioned that Hurst first applied scale invariant techniques to time-series data when he studied the water level of the Nile River. Leland et al. (1994) established in a statistically rigorous manner the self-similarity characteristic of Ethernet traces. They illustrated some of the most striking differences between self-similar models and the standard models for packet traffic considered in the literature. This phenomena is due to the traces being self-similar in nature, which implies that they also exhibit long range dependency (LRD). But, the Poisson, ARMA and Markov processes are unable to exhibit LRD. Now, self-similarity has been observed in time series obtained from network traffic, high-frequency finance, electroencephalograph and electrocardiograph traces, wind and rainfall patterns, etc..

This work is structured as follows. Self-similarity properties of two well-known processes, namely the Fractional Brownian Motion (FBM) and the Fractional ARIMA (FARIMA), are introduced in Section 2. The Embedded Branching Process (EBP) or Crossing Tree method for estimating the Hurst parameter is described in Section 3. The goodness of fit test based on the EBP method for examining a self-similar process is discussed and a modification of the goodness of fit test is proposed in Section 4. Some statistical analysis results of the simulated data and heart rate data are presented in Section 5. Some conclusion remarks are given in Section 6.

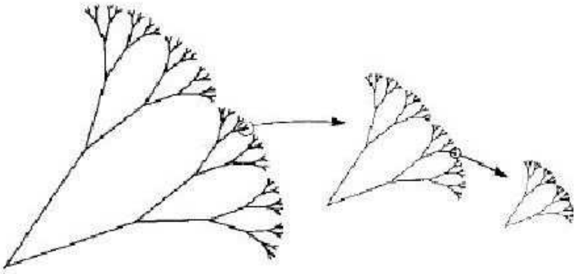

Figure 1: Self-similar structure. (This figure is excerpted from Physionet.)

2. Self-similar processes

In this section, first we introduce definitions of stationary increments, independent increment, and self-similarity. Next are the properties of stationary increment of self-similar processes. In the last part of this section, two processes with self-similar properties are presented, namely the FBM and the FARIMA. For more details about the theory of self-similarity see Beran (1994).

First, we introduce three definitions as follows:

Definition 1 : stationary increments
If for any k time points $t_1, \ldots, t_k$ and for any $k \geq 1$, the distribution of $(Y_{t_1+c} - Y_{t_1+c-1}, \ldots, Y_{t_k+c} - Y_{t_k+c-1})$ does not depend on $c \in R$, then we can say that the process $\{Y_t, t > 0\}$ has *stationary increments.*

Definition 2 : independent increments
If for any k time points $t_1, \ldots, t_k$ and for any $k \geq 1$, $Y_{t_1} - Y_{t_0}, \ldots, Y_{t_k} - Y_{t_{k-1}}$ are independent, then we can say that the process $\{Y_t, t > 0\}$ has *independent increments.*

Definition 3 : self-similarity
A stochastic process $\{Y_t, t > 0\}$ is defined as self-similar process iff

$$Y_t \overset{d}{=} a^{-H} Y_{at} \text{ for all } a > 0 \text{ and } t > 0, \tag{1}$$

where the Hurst index H is a measure of the relative rates of space and time scaling present in such a process, and $\overset{d}{=}$ denotes equal in distribution.

According to Definition 3, $\{Y_t, t > 0\}$ is a self-similar process if for any positive scalar a, the rescaled process, $a^{-H} Y_{at}$, is equal in distribution to the original process $\{Y_t, t > 0\}$. This means that, for any sequence of time points $t_1, \ldots, t_k$, and any positive constant a, $a^{-H}(Y_{at_1}, Y_{at_2}, \ldots, Y_{at_k})$ has the same distribution as

$(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_k})$. Thus, typical sample paths of a self-similar process look qualitatively the same, irrespective of the distance from which we look at them.

2.1. Properties of stationary increment of self-similar processes

For the purpose of illustration, we summarize some of the useful results on the properties of stationary increment of self-similar processes from Beran (1994).

Suppose that Y_t is a self-similar process with self-similarity parameter H. To simplify notation, assume $E(Y_t) = 0$ and a of the equation (1) is equal to t^{-1}, i.e. $Y_t \overset{d}{=} t^H Y_1$. If $H = 0$, then Y_t is equal to Y_1 for all $t > 0$. The trivial case where Y_t is a constant almost surely for every t, i.e. Y_t is not stationary unless $H = 0$ is excluded in the work. For the purpose of modeling data that look stationary, we only consider self-similar process with stationary increments. If $H < 0$, Y_t is not a measurable process.

Hence, in the following, consider a self-similar process $\{Y_t, t > 0\}$ with stationary increment X_t, $H > 0$ and $Y_0 = 0$ with probability 1 in particular. The variance of the increment process $X_t = Y_t - Y_{t-1}$ is denoted by

$$\sigma^2 = E[X_t^2] - (E[X_t])^2 = E[(Y_t - Y_{t-1})^2] = E[(Y_1 - Y_0)^2] = E[Y_1^2].$$

Then for $s < t$,

$$\begin{aligned} E[(Y_t - Y_s)^2] &= E[(Y_{t-s} - Y_0)^2] = E[Y_{t-s}^2] = E[(t-s)^{2H} Y_1^2] \\ &= (t-s)^{2H} E[Y_1^2] = \sigma^2 (t-s)^{2H}. \end{aligned}$$

On the other hand,

$$E[(Y_t - Y_s)^2] = E[Y_t^2] + E[Y_s^2] - 2E[Y_t Y_s] = \sigma^2 t^{2H} + \sigma^2 s^{2H} - 2\gamma_y(t,s),$$

where $\gamma_y(t,s) = \mathrm{Cov}(t,s)$ is the covariance function of Y_t. Hence,

$$\gamma_y(t,s) = \frac{1}{2}\sigma^2 [t^{2H} - (t-s)^{2H} + s^{2H}]. \tag{2}$$

The covariance between X_t and X_{t+k} is equal to

$$\begin{aligned} \gamma(k) &= \mathrm{Cov}(X_t, X_{t+k}) = \mathrm{Cov}(X_1, X_{k+1}) \\ &= \frac{1}{2}\sigma^2 \left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right] \quad \text{for } k \geq 0. \end{aligned} \tag{3}$$

The correlation are given by

$$\rho(k) = \frac{1}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right] \text{ for } k \geq 0. \tag{4}$$

Now, we consider the different cases of H:

1. For $0 < H < 1/2$, the correlations are summable, i.e.

   $$\sum_{k=-\infty}^{\infty} \rho(k) = 0.$$

   The process $\{X_t, t > 0\}$ has short-range dependence.

2. For $H = 1/2$, all correlations are equal to zero, i.e. the process $\{X_t, t > 0\}$ are uncorrelated.

3. For $1/2 < H < 1$, this means that the correlations decay to zero so slowly that

   $$\sum_{k=-\infty}^{\infty} \rho(k) = \infty.$$

   The process $\{X_t, t > 0\}$ has long memory or long-range dependence.

4. For $H = 1$, all correlations are equal to 1, i.e. no matter how far apart in time the observations are.

5. For $H > 1$, $\rho(k)$ diverges to infinity. This contradicts that $\rho(k)$ must be between -1 and 1. (see Appendix B.1.)

Finally, it can be seen that if covariances exist and $\lim_{k\to\infty} \rho(k) = 0$, then $0 < H < 1$.

2.2. Processes with self-similar properties

There are two important processes with self-similar properties illustrated in this part, namely the Fractional Brownian Motion (FBM) and the Gaussian Fractional ARIMA (FARIMA) or Gaussian ARFIMA.

2.2.1. Brownian Motion and Fractional Brownian Motion

Let $B(t)$ be a stochastic process with continuous sample paths and satisfies

(i) $B(t)$ is Gaussian,

(ii) $B(0) = 0$ almost surely,

(iii) $B(t)$ has independent increments,

(iv) $E[B(t) - B(s)] = 0$, and

(v) $\text{Var}[B(t) - B(s)] = \sigma^2(t - s)$ for $s < t$.

Then $B(t)$ is called (standard) Brownian motion.

In fact, Brownian motion $B(t)$ is self-similar with $H = 1/2$. It can also be shown that for all $a > 0$, $\{a^{-1/2}B(at)\}$ is also a Brownian motion. Note that as $B(t)$ is a Gaussian process, the distribution of the process is fully specified by the mean and covariances. From (ii) and (iv) we have

$$E[B(at)] = E[B(at) - B(0)] = 0 = a^{\frac{1}{2}}E[B(t)].$$

Consider the covariance $\text{Cov}(B(t), B(s))$ for $t > s$. Because $B(t) - B(s)$ is independent of $B(s) - B(0) = B(s)$, we can write

$$\begin{aligned} \text{Cov}(B(t), B(s)) &= \text{E}[B(t)B(s)] = \text{E}[(B(t) - B(s))(B(s) - B(0)) + B^2(s)] \\ &= \text{E}[B^2(s)] = \text{Var}(B(s) - B(0)) = \sigma^2 s. \end{aligned}$$

Therefore, for any $a > 0$

$$\begin{aligned} \text{Cov}(B(at), B(as)) &= \text{E}[a^{1/2}(B(t) - E(B(t))) \cdot a^{1/2}(B(s) - E(B(s)))] = a \cdot \text{Cov}(B(t), B(s)) \\ &= a\sigma^2 s = \text{Cov}(a^{1/2}B(t), a^{1/2}B(s)). \end{aligned}$$

In Section 2.1., we know that if Y_t is a self-similar process with stationary increments and suppose that $E[Y_1^2] < \infty$. Then

$$E[Y_tY_s] = \frac{1}{2}[t^{2H} - (t - s)^{2H} + s^{2H}]E[Y_1^2] \qquad \text{for } s < t.$$

Let $0 < H \leq 1$. A Gaussian process $\{B_H(t), t \geq 0\}$ is called *fractional Browian motion* if $E[B_H(t)] = 0$ and

$$E[B_H(t)B_H(s)] = \frac{1}{2}[t^{2H} - (t - s)^{2H} + s^{2H}]E[B_H(1)^2] \qquad \text{for } s < t.$$

Hence, for $H = 1/2$, the self-similar process $B_{\frac{1}{2}}(t)$ turns out to be ordinary Brownian motion. One of the important properties of the FBM is its exact self-similarity. In Figure 2, simulated series of fractional Brownian motion with
$H = 0.5$, $H = 0.7$, and $H = 0.9$ are presented.

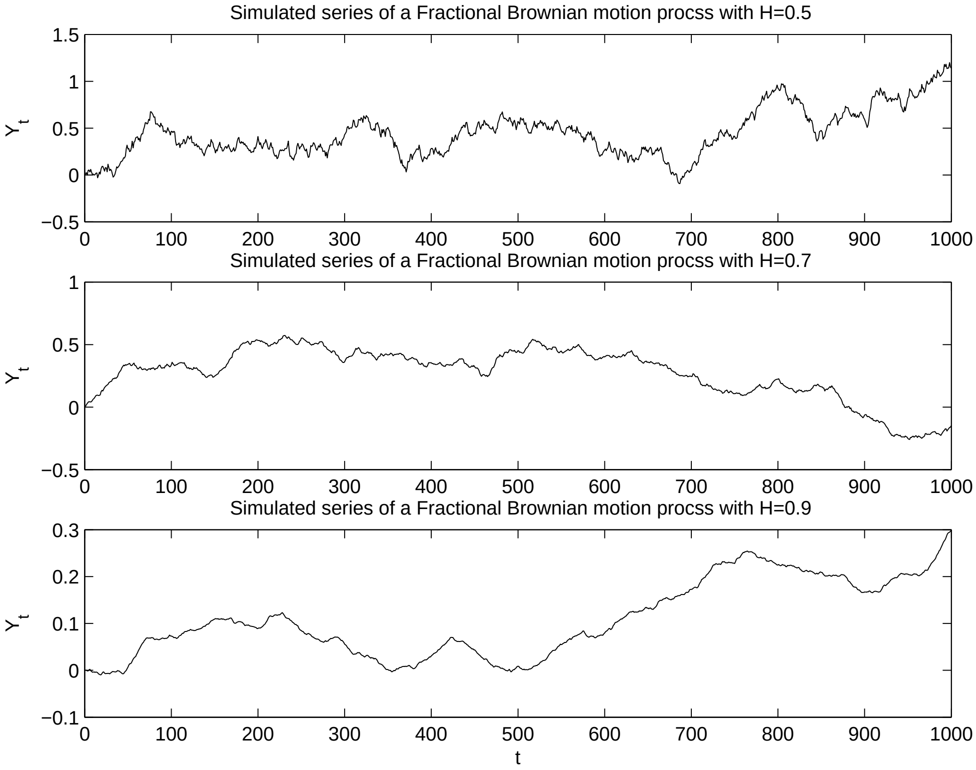

Figure 2: Simulated series of fractional Brownian motion with $H = 0.5$, $H = 0.7$, and $H = 0.9$.

2.2.2. Gaussian Fractional ARIMA (FARIMA)

A Gaussian FARIMA(p, d, q) process with $\phi(B)$ and $\theta(B)$ respectively being the autoregressive and the moving average coefficients and together with the white noise is defined by

$$\phi(B)(1-B)^d X_i = \theta(B)\epsilon_i, \qquad i \geq 1,$$

where B is the backward operator, $B\epsilon_i = \epsilon_{i-1}$, and the ϵ_t are independent, identically distributed Gaussian random variables with mean zero and variance σ^2 i.e. $\epsilon_i \sim WN(0, \sigma^2)$ and

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \cdots - \phi_p B^p, \\ \theta(B) &= 1 - \theta_1 B - \cdots - \theta_q B^q. \end{aligned}$$

For fractional d we interpret $(1-B)^{-d}$ by using formal power series expansion, as follows:

$$(1-B)^{-d} = \sum_{i=0}^{\infty} \frac{\Gamma(i+d)}{\Gamma(d)\Gamma(i+1)} B^i, \qquad i = 1, 2, \ldots,$$

where Γ denotes the gamma function. The auto-covariance function of this process satisfies, for $-1/2 < d < 1/2$,

$$\gamma(h) \sim C_d h^{2d-1} \qquad \text{as } h \to \infty,$$

where $C_d = \pi^{-1}\Gamma(1-2d)\sin \pi d$.

A FARIMA process is a symptotically self-similar process. A FARIMA$(0, d, 0)$ is a special case of FARIMA(p, d, q) processes and it is also called a $I(d)$ process. The difference between a FARIMA and an ARIMA lies in d of FARIMA is a fraction and d of ARIMA is an integer.

Figure 3 shows sample paths of several FARIMA processes and we can see there are many different types of behavior. The parameter d determines the behaviors of long-range dependence, whereas p, q, and the corresponding parameters in $\phi(B)$ and $\theta(B)$ take account of more flexible modelling of short-range properity.

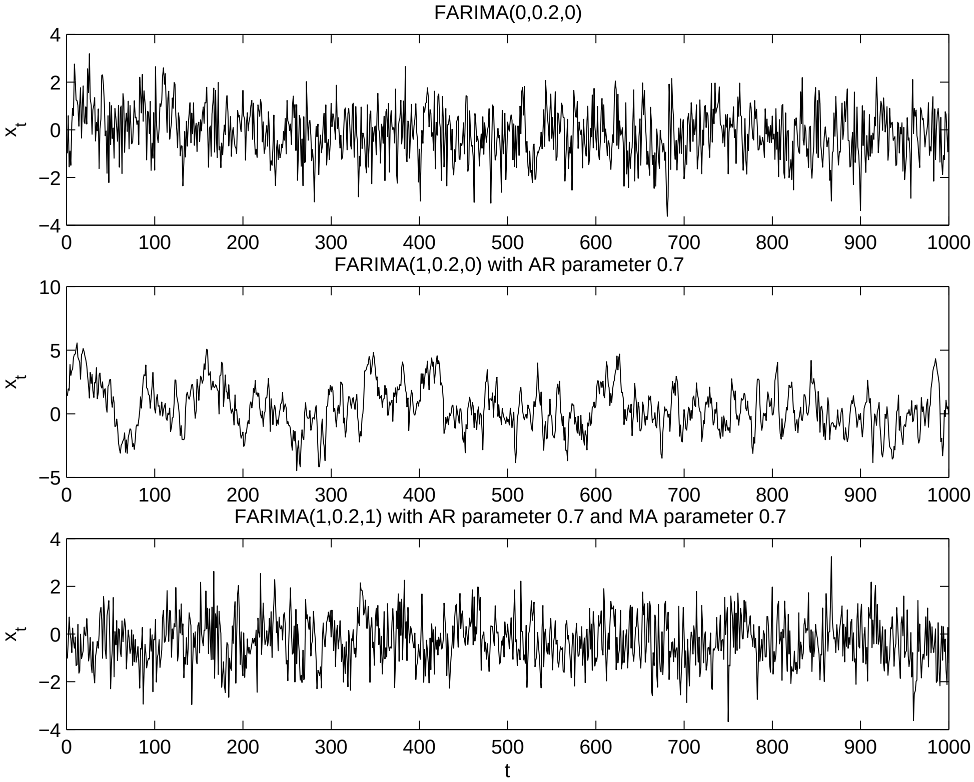

Figure 3 : Simulated series of a fractional ARIMA(0,0.2,0) process (above), a fractional ARIMA(1,0.2,0) process with AR parameter 0.7 (median), and a fractional ARIMA(1,0.2,1) process with AR parameter 0.7 and MA parameter 0.7 (below).

3. Method for estimation of the Hurst parameter

A measure of a self-similar process is the Hurst parameter. In this section, the main method for estimating the Hurst parameter to be introduced is the EBP method. In addition, three estimation methods are summarized in Appendix B.3., namely the R/S method, DFA method, and Moment method. An example of using the above methods to estimate the Hurst parameter for the data of yearly minimal water level of the Nile River is presented in Appendix B.4..

3.1 Embedded branching process (EBP) or crossing tree

Most of the current estimators view the process at regularly spaced points in time. However, many processes are observed only when there is a change in value. That is, we observe crossings of the process rather than observing it at regular times. If we only observe the process at regular time points, then we may not see all the small crossings (see Figure 4). Therefore, Jones and Shen (2004) proposed the method of Embedded Branching Process (EBP), which builds a tree of crossings that encodes the sample path.

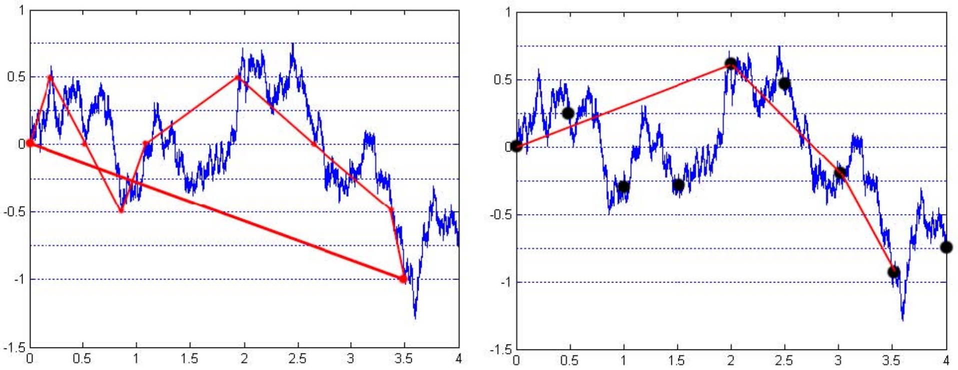

Figure 4 : The left figure gives a sample path and all its crossings and the right figure gives crossings observed if the process is sampled at regular time (the solid dots). (This figure is excerpted from the manuscript of Jones and Shen, namely self-similar processes via the crossing tree.)

To begin with, we construct a crossing tree from a given continuous process $X(t)$ and fix a base scale δ. Without loss of generality, we assume that $X(0) = 0$. For $n = 0, 1, \ldots$, let $T_0^n = 0$ and $T_{k+1}^n = \inf\{t > T_k^n : X(t) \in \delta 2^n \mathbb{Z}, X(t) \neq X(T_k^n)\}$ be the hitting times of the $\delta 2^n$-size crossings of the process. There is a natural tree structure to the crossings, as each crossing of size can be decomposed into a sequence of crossings of size. Let Z_k^n be the number of subcrossings of size $2^{n-1}\delta$ that make up the k-th crossing of size $2^n\delta$ and $N(n)$ is the total number of crossings of size $2^n\delta$. Then, $N(n) \geq \sum_{k=1}^{N(n+1)} Z_k^{n+1}$.

If the process $X(t)$ is self-similar, spatially, and temporally homogeneous, then the Z_k^n, $n \geq 0, k \geq 1$, are identically distributed. Build the crossing tree up from the bottom, then count family sizes, $Z^n = \{Z_1^n, Z_2^n, \ldots\}$. If Z^n is ergodic, then we can estimate the distribution of Z_k^n empirically. Let $\mu = E(Z_k^n)$ and if X is self-similar, then

$$X(t) \overset{d}{=} 2^{-k} X(\mu^k t) = (\mu^k)^{-\log 2/\log \mu} X(\mu^k t).$$

Hence, $H = \log 2/\log \mu$.

In practice, if Z^n is ergodic, then $\widehat{\mu}_n = \sum_{k=1}^{N(n)} Z_k^n / N(n)$ is a consistent estimator for μ, i.e. $E(\widehat{\mu}_n) \to \mu$ as $N(n) \to \infty$. The estimator at scale $2^n\delta$ is

$$\widehat{H}_n = \frac{\log 2}{\log \widehat{\mu}_n}.$$

If we believe that self-similarity holds over scales $2^m\delta$ to $2^n\delta$, then we can combine these levels to get a more accurate estimate

$$\begin{aligned}
\widehat{\mu}_{m,n} &= \frac{N(m)\widehat{\mu}_m + N(m+1)\widehat{\mu}_{m+1} + \cdots + N(n)\widehat{\mu}_n}{N(m) + N(m+1) + \cdots + N(n)} \\
&= \frac{\sum_{k=1}^{N(m)} Z_k^m + \sum_{k=1}^{N(m+1)} Z_k^{m+1} + \ldots + \sum_{k=1}^{N(n)} Z_k^n}{N(m) + N(m+1) + \ldots + N(n)}, \\
\widehat{H}_{m,n} &= \frac{\log 2}{\log \widehat{\mu}_{m,n}}.
\end{aligned}$$

In fact, $\widehat{\mu}_{m,n}$ is the mean of the total number of subcrossings for scales $2^m\delta$ to $2^n\delta$.

The analysis about the $100(1-\alpha)\%$ confidence interval for μ and more details are presented in Jones and Shen (2004). On the other hand, the method of the test statistic for examining a self-similar process is introduced in Section 4. Based on the above illustration, the estimation of parameter H for the EBP method can be summarized as follows:

Step 1. Select a δ.

Step 2. When at time t the difference with the former value is equal to $2^n\delta$, $n = 0, 1, 2, \ldots$, we keep a record of time t.

Step 3. Calculate the number of subcrossings, Z_k^n, of size $2^{n-1}\delta$ that make up the k-th crossing of size $2^n\delta$ and $N(n)$ is the total number of crossings of size $2^n\delta$.

Step 4. Calculate $\widehat{\mu}_n = \sum_{k=1}^{N(n)} Z_k^n / N(n)$.

Step 5. Calculate $\widehat{\mu}_{m,n}$ which is the mean of all Z_k^j across scale $2^m\delta$ to $2^n\delta$.

Step 6. Calculate $\widehat{H}_{m,n} = \frac{\log 2}{\log \widehat{\mu}_{m,n}}$.

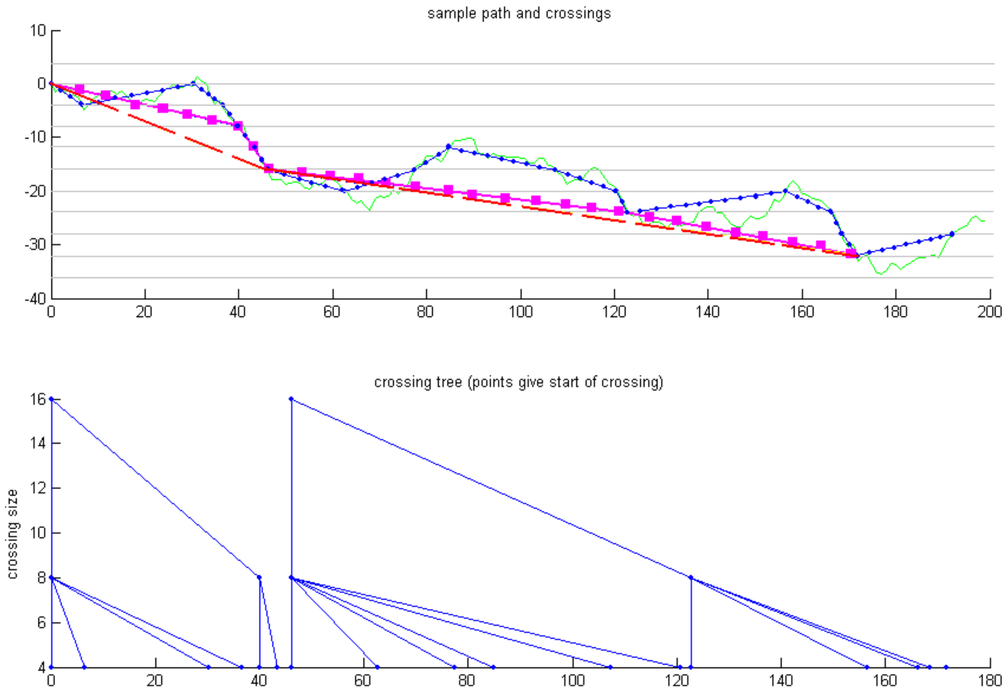

Figure 5 : The figures for EBP method. Top figure gives a sample path and all its crossings. Bottom figure gives a crossing tree.

Next, in order to understand the procedures to estimate parameter H for the EBP method, an example is given:

Example 1: Estimation of the Hurst parameter

For a simulated process, choose $\delta = 4$. Using the EBP method, we keep a record of time t when the difference with the former value is equal to $2^n\delta$, $n = 0, 1, 2, \ldots$. Then, we can obtain Figure 5.

In Figure 5, the solid line is a sample path. We use the dotted line, the square line, and the dashed line to show the change of $2^0\delta = \delta$, $2^1\delta = 2\delta$, and $2^2\delta = 4\delta$, respectively.

By calculation, the results of the number of subcrossings are as follows:

$$Z_1^1 = 4, Z_2^1 = 2, Z_3^1 = 6, Z_4^1 = 4;$$
$$Z_1^2 = 2, Z_2^2 = 2.$$

In this case, $N(1) = 4$, $N(2) = 2$,

$$\widehat{\mu}_1 = \sum_{k=1}^{4} \frac{Z_k^1}{N(1)} = 4, \ \text{ and } \widehat{\mu}_2 = \sum_{k=1}^{2} \frac{Z_k^2}{N(2)} = 2.$$

Hence,

$$\widehat{\mu}_{1,2} = \frac{N(1)\widehat{\mu}_1 + N(2)\widehat{\mu}_2}{N(1) + N(2)} = \frac{10}{3},$$

and

$$\widehat{H}_{1,2} = \frac{\log 2}{\log(10/3)} \approx 0.575717.$$

4. Goodness of fit test for examining a self-similar process

In Jones and Shen (2004), a goodness of fit test for the self-similar process based on the EBP method is introduced first, and we evaluate the proposed test in the section. The goodness of fit test is stated first in the following.

4.1 The method of goodness of fit test and an example

Let $p^n(x) = \mathbb{P}(Z_k^n = x)$, and let $\widehat{p}^n$ be the empirical distribution of Z_k^n obtained from Z^n. To test the hypothesis $p^m = p^{m+1} = \cdots = p^n$, results for testing with contingency tables are employed. Take h bins $\{2\}, \{4\}, \ldots, \{2h-2\}, \{2h, 2h+2, \ldots\}$. Because the number of the subcrossings with larger increments is quite sparse, we merge some of them into the same bin. It means that the number of subcrossings equals to $2h, 2h+2$ and the larger ones are merged. Hence, the observed value of the number of subcrossings can be divided into h bins.

Let $\widehat{p}_k^j$ be the frequency Z^j falling into bin k and $\widehat{p}_k$ be the frequency for the combined sequences $Z^m \cup \cdots \cup Z^n$ falling into bin k. Then the test statistic used is

$$\begin{aligned} T^{m,n} &= \sum_{j=m}^{n} \sum_{k=1}^{h} N(j) \frac{(\widehat{p}_k^j - \widehat{p}_k)^2}{\widehat{p}_k} \\ &= \sum_{j=m}^{n} \sum_{k=1}^{h} \frac{(N(j)\widehat{p}_k^j - N(j)\widehat{p}_k)^2}{N(j)\widehat{p}_k} \sim \chi^2_{(n-m-1)(h-1)}, \end{aligned}$$

where $N(j)\widehat{p}_k^j$ and $N(j)\widehat{p}_k$ are the corresponding observed and expected values respectively . If $Z^m, \ldots, Z^n$ are independent, then $T^{m,n}$ is asymptotically chi-squared distributed, with $(n-m-1)(h-1)$ degrees of freedom. If the observed value of $T^{m,n}$ is larger than the $\chi^2_{\alpha,(n-m-1)(h-1)}$ percentage point, then we will reject the hypothesis of self-similarity across scales $2^m\delta$ to $2^n\delta$, at the $100(1-\alpha)\%$ level.

Based on the above illustration, the goodness of fit test for examining a self-similar process can be summarized as follows:

Step 1. Choose a δ.

Step 2. When at time t the difference with the former value is equal to $2^j\delta$, $j = 0, 1, 2, \ldots$, we keep a record of time t.

Step 3. Calculate the number of subcrossings, Z_k^j, of size $2^{j-1}\delta$ that makes up the k-th crossing of size $2^j\delta$ and $N(j)$ is the total number of crossings of size $2^j\delta$.

Step 4. Using the contingency tables to compute $\widehat{p}_k^j$ and $\widehat{p}_k$ for level j and bin k.

Step 5. Evaluate the test statistic

$$T^{m,n} = \sum_{j=m}^{n}\sum_{k=1}^{h} N(j)\frac{(\widehat{p}_k^j - \widehat{p}_k)^2}{\widehat{p}_k}.$$

If $T^{m,n} > \chi^2_{\alpha,(n-m-1)(h-1)}$, then we reject the hypothesis at the $100(1-\alpha)\%$ level.

Next, in order to understand the procedure of the goodness of fit test for examining a self-similar process, details of the above computation procedure are illustrated in the following example.

Example 2: Test statistic for self-similarity
Assume that we have the number of subcrossings as follows:

$$\begin{aligned}
Z^1 = (Z_1^1, Z_2^1, \ldots, Z_{n_1}^1) &= (2,2,2,2,4,2,2,2,2,4,2,4,2,6,4,2,4,6,2,2,8,2,2,4,4,4,6,\\
&\quad\ 2,2,2,2,6,4,6,8,2,6,2,2,2,6)\\
Z^2 = (Z_1^2, Z_2^2, \ldots, Z_{n_2}^2) &= (2,2,2,6,4,2,2,2,4,2,2,4,6,2,2,8,2,4,6,2,8,6,4,2,2,2)\\
Z^3 = (Z_1^3, Z_2^3, \ldots, Z_{n_3}^3) &= (2,2,2,6,4,2,4,2,2,4,6,2,2,8,2,4,6,2)
\end{aligned}$$

According to Z^1, there are 23 points of the number of subcrossings equal to 2, i.e. $Z_1^1, Z_2^1, \dots, Z_{n_1}^1$ fall in bin $\{2\}$ is equal to 23. The number of $Z_r^1, r = 1, 2, \dots, n_1$, fall in bin $\{4\}$, $\{6\}$ and $\{8\}$ respectively are 9, 7, 2. Calculations for other scales are the same. Then we can obtain a table of the summary of the number of subcrossings as follows:

Table 1 : The summary of the number of subcrossings.

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$N(j)$
$j = 1$	23	9	7	2	41
$j = 2$	15	5	4	2	26
$j = 3$	10	4	3	1	18
$N(k)$	48	18	14	5	85

In this case, there are four bins, i.e. $h = 4$. For the purpose of convenient illustration, we define some more notation as follows:

$$\begin{aligned} N(j,k) &= \text{the total number of the subcrossings equal to } 2k \text{ for scale } 2^j\delta, \\ N(k) &= \text{the total number of the subcrossings equal to } 2k \text{ for scale } 2^m\delta \text{ to } 2^n\delta, \text{ and} \\ N &= \text{the total number of the subcrossings for all } j \text{ and } k, \end{aligned}$$

where

$$\begin{aligned} N(j) &= N(j,1) + N(j,2) + \cdots + N(j,h), \\ N(k) &= N(m,k) + N(m+1,k) + \cdots + N(n,k), \text{and} \\ N &= \sum_{j=m}^{n} \sum_{k=1}^{h} N(j,k) = \sum_{j=m}^{n} N(j) = \sum_{k=1}^{h} N(k). \end{aligned}$$

Then, we can calculate

$$\widehat{p}_k^j = \frac{N(j,k)}{N(j)} \quad \text{and} \quad \widehat{p}_k = \frac{N(k)}{N}.$$

The contingency table turns to

Table 2: The contingency table of Example 2.

	$k = 1$	$k = 2$	$k = 3$	$k = 4$
$\widehat{p}_k^1$	0.56098	0.21951	0.17073	0.04878
$\widehat{p}_k^2$	0.57692	0.19231	0.15385	0.07692
$\widehat{p}_k^3$	0.55556	0.22222	0.16667	0.05556
$\widehat{p}_k$	0.56471	0.21176	0.16471	0.05882

Hence, the test statistic is

$$T^{1,3} = \sum_{j=1}^{3} \sum_{k=1}^{4} N(j) \frac{(\widehat{p}_k^j - \widehat{p}_k)^2}{\widehat{p}_k} = 0.32439 < \chi^2_{0.05,3} = 7.81473.$$

Therefore, according to the statement mentioned above, we will not reject the hypothesis of self-similarity across scales $2^1\delta$ to $2^3\delta$ at the 95% level.

4.2. A modification of the goodness of fit test

As mentioned above, the statistic $T^{m,n} = \sum_{j=m}^{n} \sum_{k=1}^{h} N(j) \frac{(\widehat{p}_k^j - \widehat{p}_k)^2}{\widehat{p}_k}$ is stated to have the chi-square distribution with degree of freedom $(n-m-1)(h-1)$. We are interested in the performances of the goodness of test provided in Jones and Shen (2004). Hence, we use the simulated data from the FBM and FARIMA processes to examine the statement. The δ we choose is equal to a constant times E_{FBM} and E_{FARIMA}, respectively, which are defined as follows:

Assume $\{X_t, t > 0\}$ is a FBM process, then E_{FBM} is estimated by

$$E_{\mathrm{FBM}} = \widehat{E}(|X_t - X_{t-1}|) = \frac{1}{N-1} \sum_{t=2}^{N} |X_t - X_{t-1}|,$$

where N is the length of the process.

Because the FARIMA process is the stationary increment of a self-similar process, hence the corresponding self-similar process is the accumulation of FARIMA process. Assume $\{W_t, t > 0\}$ is a FARIMA process, then E_{FARIMA} is estimated by

$$E_{\mathrm{FARIMA}} = \widehat{E}(|Y_t - Y_{t-1}|) = \widehat{E}(|W_t|) = \frac{1}{N-1} \sum_{t=2}^{N} |W_t|,$$

where $Y_t = \sum_{i=1}^{t} W_i$ and N is the length of the process. In fact, $\widehat{E}(|W_t|)$ is approximately equal to the mean of the absolute value of the FARIMA process.

The δ we choose is respectively equal to E_{FBM} times $3, 2.5, 2, 1.5, 1,$ and 0.5. According to the different δ, the acceptance percentage of the chi-square distribution with the degree of freedom $(n-m-1)(h-1)$ are computed.

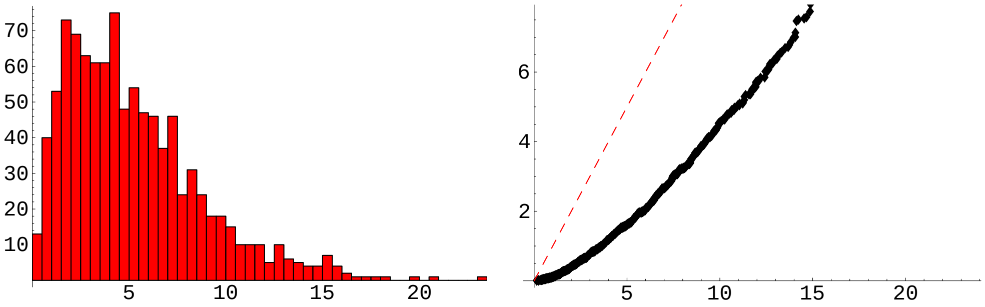

Figure 6 : Histogram (left) and quantile-quantile plot of chi-square distribution (right) for the FBM processes with $H = 0.5$, $\delta = 3E$, level$= 1 \sim 3$, $mh = 6$, $N = 10000$ and 1000 replications by using the EBP method.

From the results for the simulated data, we find that the histogram of the statistics T obtained from 1000 replications and the chi-square distribution with degree of freedom $(n-m-1)(h-1)$ does not seem to work. One result is presented in Figure 6 where data are from the FBM processes with $H = 0.5$, $\delta = 3E$, level= $1 \sim 3$, $mh = 6$, $N = 10000$ and 1000 replications.

Hence, we think the distribution of T should be modified, and conjecture that it may be as a constant times the chi-square distribution with degree of freedom ν, i.e.

$$T \overset{d}{\sim} c\chi^2_\nu.$$

Then, $\mathrm{E}(T) = c\nu$ and $\mathrm{Var}(T) = c^2(2\nu)$. Moment estimators for $\widehat{c}$ and $\widehat{\nu}$ may be computed by equating $\mathrm{E}(T)$ and $\mathrm{Var}(T)$ respectively. Therefore, we can obtain estimates for $\widehat{c}$ and $\widehat{\nu}$ as

$$\widehat{c} = \frac{S^2}{2\overline{T}}$$

and

$$\widehat{\nu} = \frac{\overline{T}}{\widehat{c}} = \frac{2\overline{T}^2}{S^2},$$

where $\overline{T}$ and S^2 are the mean and variance of the simulated data with many replications, respectively.

Figure 7 is a quantile-quantile plot of chi-square distribution after adjustment for the FBM processes with $H = 0.5$, $\delta = 3E$, level= $1 \sim 3$, $mh = 6$, $N = 10000$ and 1000 replications, where the data are the same as in Figure 6.

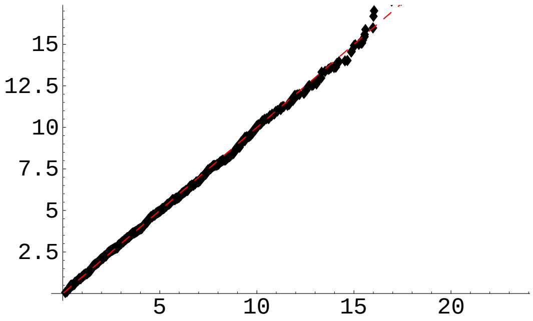

Figure 7 : Quantile-quantile plot of chi-square distribution after adjustment for the FBM processes with $H = 0.5$, $\delta = 3E$, level= $1 \sim 3$, $mh = 6$, $N = 10000$ and 1000 replications.

Next, we use the $\widehat{c}$ and $\widehat{\nu}$ to recompute the acceptance percentage of the FBM and FARIMA processes. For the simulated data of the FBM processes with $H = 0.5$, the results are summarized in Table 3. First, some notation of Table 3 are explained. The level=$m \sim n$ means the hypothesis of self-similarity across scales $2^m\delta$ to $2^n\delta$. mh means that we take $mh/2 = h$ bins, i.e. $\{2\}$, $\{4\}$, $\ldots$, $\{mh, mh+2, \ldots\}$. d.f. is the degree of freedom, $(n-m-1)(h-1)$. Percentage and new percentage respectively are the acceptance percentage of the chi-square distribution with the degree of freedom $(n-m-1)(h-1)$ and ν. The results of the FBM processes with $H = 0.6, 0.7, 0.8$ and 0.9 are presented in Appendix A.1. In addition, the results for the FARIMA processes with $H = 0.5, 0.6, \ldots, 0.9$ are showed in Appendix A.2.

Table 3 : The simulated data for FBM processes with $H = 0.5$, $N = 10000$, and 1000 replications. ($E = E_{\text{FBM}}$)

δ	level	mh	d.f.	$\overline{T}$	percentage	$\widehat{c}$	$\widehat{\nu}$	new percentage
($3E$)	1∼ 3	6	2	5.23148	65.60%	1.19301	4.38510	94.70%
0.02394		8	3	7.90665	59.10%	1.49233	5.29818	94.90%
($2.5E$)	1∼ 3	6	2	6.50947	53.50%	1.39602	4.66288	95.00%
0.01995		8	3	9.23086	48.30%	1.46594	6.29689	94.70%
($2E$)	1∼ 3	6	2	9.37397	31.20%	1.74337	5.37693	95.20%
0.01596		8	3	12.63190	25.70%	1.83356	6.88928	94.60%
($1.5E$)	1∼ 3	6	2	16.65948	5.50%	1.90437	8.74803	94.90%
0.01197		8	3	21.01984	3.80%	2.09537	10.03156	94.80%
	1∼ 4	6	4	20.91565	5.90%	1.85678	11.26448	94.70%
		8	6	27.60088	5.70%	2.30902	11.95350	95.20%
	2∼ 4	6	2	5.23148	65.60%	1.19301	4.38510	94.70%
		8	3	7.90665	59.10%	1.49233	5.29818	94.90%
(E)	1∼ 3	6	2	44.44594	0.00%	2.12666	20.89937	95.10%
0.00798		8	3	52.34066	0.00%	2.25752	23.18500	95.30%
	1∼ 4	6	4	54.50931	0.00%	2.06316	26.42034	95.20%
		8	6	65.94730	0.00%	2.32671	28.34359	95.10%
	1∼ 5	6	6	59.28502	0.00%	2.06048	28.77241	95.70%
		8	9	73.44489	0.00%	2.39851	30.62108	94.50%
	2∼ 4	6	2	9.37397	31.20%	1.74337	5.37693	95.20%
		8	3	12.63187	25.70%	1.83356	6.88928	94.60%
	2∼ 5	6	4	12.13049	39.00%	1.68749	7.18847	95.10%
		8	6	16.96073	32.20%	1.83050	9.26561	94.40%
	3∼ 5	6	2	4.44821	77.10%	1.06475	4.17771	95.20%
		8	3	6.79273	69.80%	1.13548	5.98223	94.80%
($0.5E$)	1∼ 3	6	2	238.41102	0.00%	2.20166	108.28708	95.90%
0.00399		8	3	262.80240	0.00%	2.33707	112.44960	95.30%
	1∼ 4	6	4	304.39327	0.00%	2.42363	125.59402	94.90%
		8	6	342.67274	0.00%	2.82784	121.17805	94.50%
	1∼ 5	6	6	331.56042	0.00%	2.48617	133.36211	95.30%
		8	9	379.45595	0.00%	3.08415	123.03406	94.80%
	2∼ 4	6	2	44.44594	0.00%	2.12666	20.89937	95.10%
		8	3	52.34066	0.00%	2.25752	23.18500	95.30%
	2∼ 5	6	4	54.50931	0.00%	2.06316	26.42034	95.20%
		8	6	65.94730	0.00%	2.32671	28.34359	95.10%
	3∼ 5	6	2	9.37397	31.20%	1.74337	5.37693	95.20%
		8	3	12.63187	25.70%	1.83356	6.88928	94.60%

From the results of FBM and FARIMA processes, we can find that the percentage of acceptance with larger δ are larger than that with smaller δ. After adjusting $\widehat{c}$ and $\widehat{\nu}$, the new percentages of acceptance are close to 95% for all conditions chosen. It means that the adjustment is useful for the FBM and FARIMA processes. Since the method is quite robust for different δ, we choose δ according to computer time for computation. Hence, in the following comparisons, the choice of δ and mh respectively are $3E$ and 6.

Because the FBM process and FARIMA$(0, d, 0)$ process or $I(d)$ process are with the same distribution under $H = 0.5$ or $d = 0$, the results should be similar by using the EBP method. From Tables 3 and 16 (see Appendix A.2.), the results of them are indeed consistent. In the following, we would like to compare them under the same criterion and choose $\widehat{c}$ and $\widehat{\nu}$ from the FBM process under different H since the FBM and FARIMA process respectively are an exact and symptotically self-similar process. Before making the comparison, we feel that the number of replications may be not enough, hence the data of FBM process are simulated for 1000, 4000 and 9000 replications where 4000 replications are new 3000 replications add to the previous 1000 replications and 9000 replications are new 5000 replications add to the previous 4000 replications. After calculations of $\widehat{c}$ and $\widehat{\nu}$, the results are shown in Figure 8.

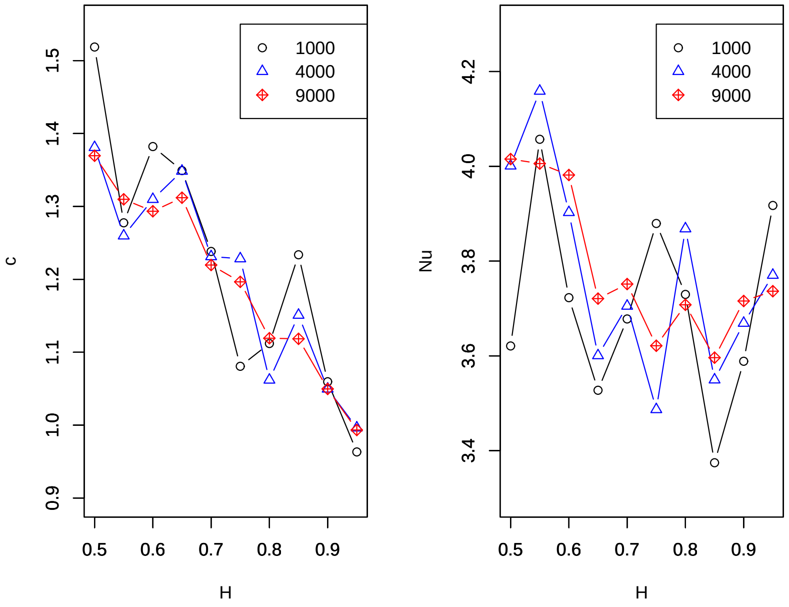

Figure 8 : The different $\widehat{c}$ and $\widehat{\nu}$ values are obtained from 1000, 4000 and 9000 replications.

From Figure 8, it seems the number of replications equal to 9000 is better than that for 1000 and 4000. We think the behavior of $\widehat{c}$ should be monotone. According to Figure 8, there seems to be a linear relationships between $\widehat{c}$ and H. However, there are still variations at $H = 0.55, 0.65, 0.75$, 0.85 and 0.95. Therefore, it seems the number of replications equal to 9000 may not be enough. The values of $\widehat{c}$ and $\widehat{\nu}$ for 9000 replications are presented in Table 4.

Table 4 : The values of $\widehat{c}$ and $\widehat{\nu}$ for different H.
(9000 replications)

H	criterion $\widehat{c}$	criterion $\widehat{\nu}$	critical value $(c\chi^2_\nu)$
0.50	1.36955	4.01526	13.02780
0.55	1.30975	4.00576	12.43880
0.60	1.29337	3.98165	12.23260
0.65	1.31194	3.72080	11.84870
0.70	1.21961	3.75153	11.07650
0.75	1.19652	3.62136	10.60960
0.80	1.11920	3.70779	10.08400
0.85	1.11840	3.59623	9.87032
0.90	1.04966	3.71579	9.47126
0.95	0.99333	3.73646	8.99681

We use the least square fit to obtain the relationships between $\widehat{c}$ and H and $\widehat{\nu}$ and H for 9000 replications. Figure 9(a) suggests that there seems to be a strong statistical relationship between $\widehat{c}$ and H and the linear regression model appears to be reasonable. The regression line of $\widehat{c}$ and H is $\widehat{c} = 1.78541 - 0.810033H$ and $R^2 = 0.96045$. The relationship between $\widehat{\nu}$ and H is illustrated in Figure 9(b). The regression function of $\widehat{\nu}$ and H is $\widehat{\nu} = 6.35639 - 6.51247H + 3.93668H^2$. Because the R^2 of quadratic and cubic regression model for the relationship between $\widehat{\nu}$ and H respectively are 0.972509 and 0.972548, it seems the quadratic regression model may be enough to illustrate.

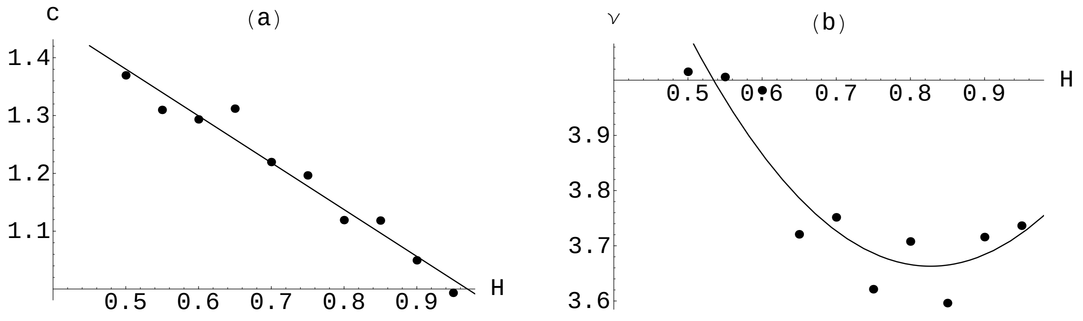

Figure 9 : The relationships between $\widehat{c}$ and H and $\widehat{\nu}$ and H with 9000 replications.

In the following, we recompute the acceptance percentage of the FBM processes for 1000, 4000 and 9000 replications according to the values of criterion $\widehat{c}$ and $\widehat{\nu}$ mentioned above and the results are presented in Table 5.

Table 5 : The acceptance percentage of FBM processes under criterion. (replications)

H	FBM(1000)	FBM(4000)	FBM(9000)
0.50	93.90%	94.45%	94.68%
0.55	95.50%	95.08%	94.88%
0.60	94.20%	94.90%	94.88%
0.65	94.40%	94.93%	95.02%
0.70	94.80%	94.88%	94.93%
0.75	96.20%	94.93%	94.87%
0.80	95.80%	95.60%	94.90%
0.85	93.80%	94.70%	95.08%
0.90	95.20%	95.23%	95.03%
0.95	94.79%	94.72%	94.97%

Next, the data of FARIMA$(0, d, 0)$, FARIMA$(1, d, 1)$, FARIMA$(1, d, 0)$ and FARIMA$(0, d, 1)$ processes are simulated with 1000 replications and AR parameter ($\phi_1 = 0.5$) or MA parameter ($\theta_1 = 0.5$). The results of acceptance percentage of the above processes are presented in Table 6.

Table 6 : The acceptance percentage of FARIMA(p, d, q) processes with 1000 replications under criterion.

H (d)	FARIMA$(0, d, 0)$ or $I(d)$	FARIMA$(1, d, 1)$	FARIMA$(1, d, 0)$	FARIMA$(0, d, 1)$
0.5 (0.0)	94.40%	14.90%	32.70%	78.40%
0.6 (0.1)	96.40%	95.70%	44.50%	98.00%
0.7 (0.2)	96.30%	97.00%	63.10%	94.60%
0.8 (0.3)	96.40%	94.10%	79.20%	92.20%
0.9 (0.4)	94.50%	96.70%	88.40%	94.80%

According to Table 6, the acceptance percentage of FARIMA$(0, d, 0)$ or $I(d)$ processes is approximate to 95% under the criterion $\widehat{c}$ and $\widehat{\nu}$. It means that the goodness of fit test is also suitable for the $I(d)$ process. Now, we consider the corresponding parameters in $\phi(B)$ and $\theta(B)$. In the case with AR parameter and MA parameter are included in FARIMA(p, d, q) processes, the acceptance percentages do not seem to be good when there are only AR parameters present. Hence, the effect of AR parameters of the FARIMA$(1, d, 0)$ processes is eliminated through AR filter with AR parameter equal to 0.5 and the acceptance percentage should be similar to $I(d)$ process. From Tables 6 and 7, the differences of acceptance percentage between them are smaller.

Table 7 : The acceptance percentage of FARIMA$(1, d, 0)$ processes and after AR filter with 1000 replications under criterion.

H (d)	FARIMA$(1, d, 0)$	after AR filter
0.5 (0.0)	32.70%	95.60%
0.6 (0.1)	44.50%	96.50%
0.7 (0.2)	63.10%	96.50%
0.8 (0.3)	79.20%	95.80%
0.9 (0.4)	88.40%	95.40%

In Table 6, the acceptance percentage of FARIMA$(1, 0, 1)$ is 14.90% which differs from the FARIMA$(1, d, 1)$ processes with $d = 0.1, 0.2, 0.3$ and 0.4. It is conjectured that there may be some questions. We use the same method for the FARIMA$(1, d, 1)$ processes with $H = 0.51$ or $d = 0.01$ and 1000 replications. The result of the acceptance percentage is 96.00%. Hence, it seems that there may be some interesting features about the acceptance percentage of the FARIMA$(1, 0, 1)$ worth being investigated in the future.

5. Numerical Comparison and Application

In this section, we use six different methods to estimate the parameter H for the simulated data of the FARIMA process and latter apply to the heart rate data collected from ICU (intensive care unit) of the department of cardiovascular surgery of Kaohsiung Veterans General Hospital. The six methods respectively are rescaled adjusted range (R/S), EBP, Detrended Fluctuation Analysis (DFA), Absolute, Variance, and approximate MLE, where EBP method is introduced in Section 3 and others are illustrated in Appendix.

5.1. Numerical Comparison

In this part, the Gaussian FARIMA$(0, d, 0)$ or $I(d)$ process with $d = 0.2$ are simulated and the six methods are used to estimate the Hurst parameter, where the so called absolute and variance method are based on the corresponding moment method. In the expression of FARIMA(p, d, q),

$$\phi(B)(1 - B)^d X_i = \theta(B)\epsilon_i, \qquad i \geq 1,$$

if the MLE method is used to estimate the parameter of FARIMA, then extensive computations are required. Hence, Haslett and Raftery (1989) proposed the method for estimating MLE under the normality assumption for the samples. Hence, the relation between H and d is $H = d + 1/2$. It can be evaluated by the code "arima.fracdiff" of the software S-plus. The results are illustrated in Table 8.

Table 8 : The results of numerical evaluation for $I(d)$ processes with $d = 0.2$ ($H = 0.7$), $N = 10000$, and 20 replications.

	R/S	EBP	DFA1	Absolute	Variance	approximate MLE
Mean	0.69724	0.72430	0.68440	0.68963	0.68893	0.697547605
Bias	-0.00276	0.02430	-0.01560	-0.01037	-0.01107	-0.002452395
Standard deviation	0.01277	0.01313	0.01529	0.03948	0.03728	0.006116568

Based on the limited simulation results, it seems the MLE method has the smallest bias and standard deviation. The results of the estimation of parameter H for other methods are also close to $H = 0.7$. The estimation by the EBP method tends to overestimate, and the others tend to underestimate the true parameters.

5.2. Application

The motivation for studying the self-similar process is from the need of analyzing the heart rate data to examine the degree of self-similarity with respect to the health condition of a person. It is observed that the variation of heart rate data is non-stationary for instance there are variations in the heart rate when a person takes a break. It is not easy to fit in with the assumption of stationary for the methods of traditional analysis. Recently there have been several reports that certain physiological data may display the properties of long-range correlation and self-similarity which have some information in clinical medical research. After understanding the self-similarity for heart rate data, we are interested in the relation between the degree of self-similar and health. In Peng et al (1995), the relation between health and H are as follows:

Table 9 : The relation between health and H.

Group	state	Number of people	Age range (years)	H (Mean $\pm S.D.$)
1	healthy adults	29	$20 \sim 64$	1.00 ± 0.10
2	heart failure	15	$22 \sim 71$	1.24 ± 0.22
3	heart danger	10	$35 \sim 82$	1.22 ± 0.25

In Table 9, data from each subject contained approximately 24 hours of ECG recording encompassing $\sim 10^5$ heartbeats. The similar results are obtained when the time series are divided into three consecutive subsets (of $\sim$ 8 hours each) and repeated the above analysis, i.e. the value of H should not be affected by the sampling time.

Similar analysis is applied to study the effect of physiologic aging. The relation between H and age are presented in Table 10, where data from healthy subjects underwent 2 hours of continuous supine resting ECG recording. In the group of healthy young subjects, the value of H closes to 1.0. In healthy elderly subjects, the interbeat interval time series showed two scaling regions. The the values of H are 0.5 and 1.5 over the short range and the longer range, respectively.

Table 10 : The relation between age and H.

Group	state	Number of people	Age range (years)	H value
1	healthy young	10	$21 \sim 34$	≈ 1.00
2	healthy elder	10	$68 \sim 81$	0.50 or 1.50

In the following, based on the assumption of self-similarity for the heart rate data, the EBP method, R/S, DFA, and Moment method are used to estimate the Hurst parameter for the heart rate data obtained from ICU of the department of cardiovascular surgery of Kaohsiung Veterans General Hospital. The interval of time is calculated after the operation until the patient leaves the ICU. All patients are survivors except the patient of number hr10. According to the limited results, the parameter H estimated from the DFA method are larger than others which is of interest to see why this phenomena occurs on the real data.

Table 11 : Estimation of the Hurst parameter for different methods.

number	N	R/S	EBP	DFA	Absolute	Variance
hr1	1885	0.93163	0.973	1.1540	0.92561	0.90840
hr2	1439	1.03365	0.942	1.3748	0.81583	0.82765
hr3	916	1.05206	0.915	1.2049	0.75417	0.70349
hr4	1360	0.96912	0.994	1.3033	0.97632	0.97776
hr5	1460	0.98111	0.963	1.1272	0.94161	0.92978
hr6	3409	0.98145	0.980	1.3541	0.97920	0.96076
hr7	5910	0.91935	0.970	1.1500	0.95655	0.93985
hr8	7953	0.91536	0.902	1.0735	0.82384	0.80674
hr9	1264	0.95699	0.955	1.1598	0.91345	0.91952
hr10	2616	0.98161	0.976	1.4980	0.96137	0.94661

N is the length of the heart rate data for the patients.

6. Conclusion

In this thesis, a modification of the goodness of fit test is proposed in Section 4.2 and it is useful for the FBM and FARIMA processes. We should use the same method for the processes without the property of self-similarity to see whether it is sensitive to non self-similar processes, in other words whether it is a test with high power. As the relationships between $\widehat{c}$ and H and $\widehat{\nu}$ and H are illustrated, it seems that the number of replications equal to 9000 may be not enough. Hence, larger number of replications is needed.

From some results of simulation, the acceptance percentage of the FARIMA processes may be affected by AR parameter or MA parameter. Therefore, the FARIMA(p, d, q) processes with different coefficients of the parameter AR and MA are needed to do the simulation. We recompute the acceptance percentage to see the relationship between parameters and acceptance percentages. In last part of Section 4.2, some interesting features about the acceptance percentage of FARIMA$(1, 0, 1)$ do exist. Hence, more theoretical investigations and simulations are needed to see why this phenomena occurs.

If the real data can be fitted in $I(d)$ process, we can obtain the estimation of d and find the corresponding values of $\widehat{c}$ and $\widehat{\nu}$ through the relationships between $\widehat{c}$ and H and $\widehat{\nu}$ and H. The real data can be tested whether it is a self-similar process or not.

References

1. Bates, S. and McLaughlin, S. (1996). An investigation of the impulsive nature of Ethernet data using stable distributions. In *Proceedings of the 12th UK Performance Engineering Workshop* (Edited by J. Hillston and R. Pooley), 17-32.

2. Bates, S. and McLaughlin, S. (1997). Testing the Gaussian assumption for self-similar teletraffic models. *IEEE Signal Processing Workshop on Higher-Order Statistics*, 21-23.

3. Beran, J. (1994). *Statistics for Long-Memory Processes.* Chapman and Hall, New York.

4. Embrechts, P. and Maejima, M. (2002). *Selfsimilar Processes.* Princeton Series in Applied Mathematics, Princeton University Press.

5. Feder, J. (1988). *Fractals.* Plenum Press, New York.

6. Guo, C.-Y. (2004). Studies in the electrocardiogram monitoring indices. Master thesis, Department of Applied Mathematics, National Sun Yat-sen University.

7. Haslett, J. and Raftery, A.E. (1989). Space-time modeling with long-memory dependence: assessing ireland's wind power resource. *Appl. Stat.*, 38, 1, 1-50.

8. Jones, O.D. and Shen, Y. (2004). Estimating the Hurst index of a self-Similar process via the crossing tree. *IEEE Signal Processing Letters*, 11, 4, 416-419.

9. Kantelhardt, J.W. , Bunde, E.K., Rego, H.A., Havlin, S. and Bunde, A. (2001). Detecting long-range correlations with detrended fluctuation analysis. *Physica A*, 294, 441.

10. Kantelhardt, J.W., Zschiegner, S.A. , Bunde, E.K., Bunde, A., Havlin, S., and Stanley, H.E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, 316, 87-114.

11. Kolmogorov, A.N. (1941) Local structure of turbulence in fluid for very large Reynolds numbers. *Transl. in Turbulence. S.K.Friedlander and L.Topper (eds.) (1961), Interscience Publishers, New York*, 151-155.

12. Leland, W.E., Taqqu, M.S., Willinger, W. and Wilson, D.V. (1994). On the self-similar nature of Ethernet traffic (extended version). ACM Transactions on Networking, 2, 1-14.

13. Mandelbrot, B.B. and Wallis, J.R. (1969a) Computer experiments with fractional Gaussian noises. *Water Resources Res.*, 5, 1, 228-267.

14. Mandelbrot, B.B. and Wallis, J.R. (1969b) Some long-run properties of geophysical records. *Water Resources Res.*, 5, 321-340.

15. Mandelbrot, B.B. and Wallis, J.R. (1969c) Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Res.*, 5, 967-988.

16. Peng, C.-K., Havlin, S., Stanley, H.E. and Goldberger, A.L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, 5, 82-87.

17. Sakalauskien, G. (2003). The Hurst Phenomenon in Hydrology. Environmental Research, Engineering and Management, 3, 16-20.

18. Shawki, Y., Attaia, K., Naggar, O., Elwan, Y., Kamel, S. (2005). Floods and their influence on the Nile River system. *GIS Modelling Application in River Engineering Research Cluster, Nile Basin Capacity Building Network 'NBCBN'*

19. Taqqu, M.S., Teverovsky, V., and Willinger, W. (1995). Estimators for long-range dependence: an empirical study. *Fractals*, 3, 4, 785-788.

20. Taqqu, M.S. and Teverovsky, V. (1998). On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series. A Practical Guide to Heavy Tails: Statistical Techniques and Applications Boston, MA: Birkhauser, 177-217.

21. Physionet, available at http://www.physionet.org/tutorials/ndc/.

22. Matlab code for estimating the Hurst index H of a self-similar process, available at http://www.ms.unimelb.edu.au/ odj/.

Appendix A.

A.1. The simulated data for FBM processes with $H = 0.6, \ldots, 0.9$, $N = 10000$, and 1000 replications.

Table 12 : The simulated data for FBM processes with $H = 0.6$, $N = 10000$ and 1000 replications. ($E = E_{\mathrm{FBM}}$)

δ	level	mh	d.f.	$\overline{T}$	percentage	$\widehat{c}$	$\widehat{\nu}$	new percentage
($3E$)	1∼ 3	6	2	5.07511	68.80%	1.33150	3.81158	94.50%
0.00953		8	3	7.36337	62.60%	1.37636	5.34989	94.70%
($2.5E$)	1∼ 3	6	2	5.75921	61.10%	1.41374	4.07373	94.40%
0.00794		8	3	8.24590	55.40%	1.40521	5.86808	95.00%
($2E$)	1∼ 3	6	2	7.86477	41.00%	1.59560	4.92903	94.70%
0.00635		8	3	10.61740	36.90%	1.75029	6.06607	95.20%
($1.5E$)	1∼ 3	6	2	12.42827	15.70%	1.69255	7.34292	96.00%
0.00476		8	3	15.63542	13.70%	1.73973	8.98728	95.70%
	1∼ 4	6	4	16.59289	18.40%	1.82868	9.07372	95.40%
		8	6	21.60027	16.40%	2.01796	10.70401	95.40%
	2∼ 4	6	2	5.10358	69.00%	1.35440	3.76815	94.70%
		8	3	7.45608	62.00%	1.44338	5.16569	94.90%
(E)	1∼ 3	6	2	29.71977	0.10%	1.97296	15.06355	94.50%
0.00318		8	3	34.50849	0.10%	2.08953	16.51495	95.20%
	1∼ 4	6	4	38.35180	0.00%	2.10939	18.18150	94.30%
		8	6	45.57338	0.10%	2.32321	19.61660	94.40%
	1∼ 5	6	6	43.30282	0.20%	2.18188	19.84652	94.80%
		8	9	52.72637	0.20%	2.46561	21.38471	94.50%
	2∼ 4	6	2	7.66515	42.20%	1.53659	4.98843	95.20%
		8	3	10.38906	36.90%	1.51996	6.83507	95.40%
	2∼ 5	6	4	10.60203	48.90%	1.48261	7.15093	94.80%
		8	6	14.73018	42.00%	1.56556	9.40886	94.80%
	3∼ 5	6	2	4.42083	74.30%	1.15686	3.82140	94.40%
		8	3	6.50468	70.40%	1.21264	5.36409	94.70%
($0.5E$)	1∼ 3	6	2	148.23837	0.00%	2.13321	69.49061	95.00%
0.00159		8	3	159.90620	0.00%	2.26585	70.57216	95.00%
	1∼ 4	6	4	194.07488	0.00%	2.45699	78.98895	95.30%
		8	6	212.53213	0.00%	2.75134	77.24674	95.50%
	1∼ 5	6	6	216.66349	0.00%	2.58407	83.84589	95.70%
		8	9	240.72707	0.00%	3.07215	78.35791	95.70%
	2∼ 4	6	2	29.71977	0.10%	1.97296	15.06355	94.50%
		8	3	34.50849	0.10%	2.08953	16.51495	95.20%
	2∼ 5	6	4	38.35180	0.00%	2.10939	18.18150	94.30%
		8	6	45.57338	0.10%	2.32321	19.61660	94.40%
	3∼ 5	6	2	7.66515	42.20%	1.53659	4.98843	95.20%
		8	3	10.38906	36.90%	1.51996	6.83507	95.40%

A.1. (continued)

Table 13 : The simulated data for FBM processes with $H = 0.7$, $N = 10000$ and 1000 replications. ($E = E_{\text{FBM}}$)

δ	level	mh	d.f.	$\overline{\overline{T}}$	percentage	$\widehat{c}$	$\widehat{\nu}$	new percentage
($3E$)	1∼ 3	6	2	4.57172	74.40%	1.24992	3.65760	95.00%
0.00379		8	3	6.97619	66.60%	1.41862	4.91759	94.40%
($2.5E$)	1∼ 3	6	2	5.13301	67.80%	1.35788	3.78017	95.00%
0.00316		8	3	7.59092	61.10%	1.41147	5.37803	94.90%
($2E$)	1∼ 3	6	2	6.04199	58.70%	1.48699	4.06323	94.70%
0.00253		8	3	8.58621	51.20%	1.49858	5.72955	94.40%
($1.5E$)	1∼ 3	6	2	8.78456	34.30%	1.51903	5.78301	95.30%
0.00190		8	3	11.36885	31.90%	1.53707	7.39646	95.40%
	1∼ 4	6	4	12.32379	37.80%	1.69649	7.26428	94.60%
		8	6	16.48180	35.20%	1.76782	9.32325	95.60%
	2∼ 4	6	2	4.54993	75.60%	1.10507	4.11733	94.60%
		8	3	6.87375	65.60%	1.18582	5.79662	94.80%
(E)	1∼ 3	6	2	18.43814	3.60%	1.92606	9.57300	95.50%
0.00126		8	3	21.45353	3.70%	1.93123	11.10874	95.70%
	1∼ 4	6	4	24.76186	4.80%	2.26749	10.92040	94.90%
		8	6	29.78593	4.20%	2.33320	12.76612	95.60%
	1∼ 5	6	6	29.07373	6.80%	2.47252	11.75876	94.70%
		8	9	36.16144	5.80%	2.66935	13.54689	96.00%
	2∼ 4	6	2	6.13322	57.90%	1.46439	4.18825	94.90%
		8	3	8.59203	51.60%	1.47507	5.82483	95.10%
	2∼ 5	6	4	8.81527	62.20%	1.62108	5.43791	94.50%
		8	6	12.73613	55.80%	1.69367	7.51983	94.60%
	3∼ 5	6	2	4.08534	79.60%	1.12290	3.63822	95.00%
		8	3	6.43525	71.70%	1.22644	5.24708	95.80%
($0.5E$)	1∼ 3	6	2	83.22702	0.00%	2.00095	41.59376	94.70%
0.00063		8	3	89.05101	0.00%	2.01822	44.12348	94.70%
	1∼ 4	6	4	111.35274	0.00%	2.40799	46.24303	96.60%
		8	6	120.45567	0.00%	2.47996	48.57167	95.60%
	1∼ 5	6	6	127.00740	0.00%	2.85067	44.55352	95.80%
		8	9	139.76516	0.00%	2.99990	46.58993	95.70%
	2∼ 4	6	2	18.43814	3.60%	1.92606	9.57300	95.50%
		8	3	21.45353	3.70%	1.93123	11.10874	95.70%
	2∼ 5	6	4	24.76186	4.80%	2.26749	10.92040	94.90%
		8	6	29.78593	4.20%	2.33320	12.76612	95.60%
	3∼ 5	6	2	6.13322	57.90%	1.46439	4.18825	94.90%
		8	3	8.59203	51.60%	1.47507	5.82483	95.10%

A.1. (continued)

Table 14 : The simulated data for FBM processes with $H = 0.8$, $N = 10000$ and 1000 replications. ($E = E_{\mathrm{FBM}}$)

δ	level	mh	d.f.	$\overline{\overline{T}}$	percentage	$\widehat{c}$	$\widehat{\nu}$	new percentage
($3E$)	1∼ 3	6	2	3.99189	80.40%	1.05178	3.79535	95.50%
0.00151		8	3	6.24993	73.00%	1.10624	5.64971	94.40%
($2.5E$)	1∼ 3	6	2	4.61250	73.30%	1.24326	3.71000	95.70%
0.00126		8	3	6.91246	64.80%	1.20755	5.72438	95.90%
($2E$)	1∼ 3	6	2	5.00371	68.40%	1.20403	4.15580	94.70%
0.00101		8	3	7.24305	63.20%	1.28010	5.65819	95.20%
($1.5E$)	1∼ 3	6	2	6.27262	56.70%	1.52076	4.12466	94.70%
0.00075		8	3	8.68735	50.80%	1.55040	5.60329	95.80%
	1∼ 4	6	4	9.30544	61.60%	1.84320	5.04853	94.60%
		8	6	13.10811	54.70%	1.89650	6.91174	94.40%
	2∼ 4	6	2	4.11568	78.30%	1.22034	3.37258	94.70%
		8	3	6.35077	70.90%	1.15352	5.50557	95.60%
(E)	1∼ 3	6	2	10.75517	24.50%	1.85112	5.81008	95.40%
0.00050		8	3	13.21664	21.40%	1.70521	7.75076	95.60%
	1∼ 4	6	4	15.50002	26.70%	2.44288	6.34497	95.10%
		8	6	19.52748	25.20%	2.32759	8.38957	94.20%
	1∼ 5	6	6	19.36691	30.20%	2.80386	6.90723	95.80%
		8	9	24.92560	28.40%	2.72467	9.14813	94.40%
	2∼ 4	6	2	5.10101	68.60%	1.44201	3.53742	93.90%
		8	3	7.30662	64.10%	1.37004	5.33316	94.00%
	2∼ 5	6	4	7.75085	72.00%	1.59500	4.85948	94.50%
		8	6	11.13397	68.60%	1.57489	7.06967	94.10%
	3∼ 5	6	2	3.98554	80.30%	1.09499	3.63979	95.30%
		8	3	6.09806	75.50%	1.13802	5.35848	93.80%
($0.5E$)	1∼ 3	6	2	39.35629	0.20%	2.36639	16.63136	94.50%
0.00025		8	3	42.90730	0.10%	2.32076	18.48845	94.60%
	1∼ 4	6	4	54.39615	0.10%	3.29947	16.48632	95.20%
		8	6	60.05414	0.10%	3.22475	18.62287	95.00%
	1∼ 5	6	6	64.47194	0.20%	4.16585	15.47629	95.50%
		8	9	72.52717	0.30%	4.14421	17.50084	95.50%
	2∼ 4	6	2	10.75517	24.50%	1.85112	5.81008	95.40%
		8	3	13.21664	21.40%	1.70521	7.75076	95.60%
	2∼ 5	6	4	15.50002	26.70%	2.44288	6.34497	95.10%
		8	6	19.52748	25.20%	2.32759	8.38957	94.20%
	3∼ 5	6	2	5.10101	68.60%	1.44201	3.53742	93.90%
		8	3	7.30662	64.10%	1.37004	5.33316	94.00%

A.1. (continued)

Table 15 : The simulated data for FBM processes with $H = 0.9$, $N = 10000$ and 1000 replications. ($E = E_{\text{FBM}}$)

δ	level	mh	d.f.	$\overline{T}$	percentage	$\widehat{c}$	$\widehat{\nu}$	new percentage
($3E$)	1∼ 3	6	2	3.85514	82.90%	0.96228	4.00626	95.20%
0.00060		8	3	5.90565	76.30%	1.02682	5.75141	95.10%
($2.5E$)	1∼ 3	6	2	3.86292	81.20%	0.94136	4.10356	95.70%
0.00050		8	3	5.98147	74.40%	0.93428	6.40227	95.40%
($2E$)	1∼ 3	6	2	4.28507	76.30%	1.09466	3.91453	94.70%
0.00040		8	3	6.57474	70.10%	1.25425	5.24195	95.00%
($1.5E$)	1∼ 3	6	2	4.59586	72.20%	1.09216	4.20805	95.30%
0.00030		8	3	6.72947	68.00%	1.17143	5.74468	95.60%
	1∼ 4	6	4	7.26470	74.90%	1.48312	4.89826	95.40%
		8	6	10.56919	70.80%	1.59307	6.63447	95.00%
	2∼ 4	6	2	3.85514	82.90%	0.96228	4.00626	95.20%
		8	3	5.90565	76.30%	1.02682	5.75141	95.10%
(E)	1∼ 3	6	2	6.15659	56.60%	1.52943	4.02541	94.70%
0.00020		8	3	8.45277	52.90%	1.51878	5.56551	94.60%
	1∼ 4	6	4	9.42001	59.80%	1.90750	4.93840	94.90%
		8	6	13.06353	55.40%	1.93777	6.74151	94.80%
	1∼ 5	6	6	12.47332	60.60%	2.28777	5.45219	94.80%
		8	9	17.42220	56.80%	2.38674	7.29958	94.50%
	2∼ 4	6	2	4.28507	76.30%	1.09466	3.91453	94.70%
		8	3	6.57474	70.10%	1.25425	5.24195	95.00%
	2∼ 5	6	4	6.63658	80.00%	1.38491	4.79205	95.00%
		8	6	9.99664	74.10%	1.50568	6.63930	94.10%
	3∼ 5	6	2	3.71570	83.50%	1.08586	3.42188	95.40%
		8	3	5.80025	77.40%	1.10110	5.26766	95.70%*
($0.5E$)	1∼ 3	6	2	13.82863	19.00%	2.58687	5.34571	95.40%
0.00010		8	3	16.41252	16.90%	2.50088	6.56269	95.80%
	1∼ 4	6	4	20.59780	19.90%	3.67980	5.59754	95.20%
		8	6	24.73788	18.50%	3.50514	7.05760	95.10%
	1∼ 5	6	6	25.95951	20.80%	4.47957	5.79510	95.60%
		8	9	31.90258	20.00%	4.34381	7.34439	95.40%
	2∼ 4	6	2	6.15659	56.60%	1.52943	4.02541	94.70%
		8	3	8.45277	52.90%	1.51878	5.56551	94.60%
	2∼ 5	6	4	9.42001	59.80%	1.90750	4.93840	94.90%
		8	6	13.06353	55.40%	1.93777	6.74151	94.80%
	3∼ 5	6	2	4.28507	76.30%	1.09466	3.91453	94.70%
		8	3	6.57474	70.10%	1.25425	5.24195	95.00%

*: The number of T is equal to 999.

Remark: When the statistic T is calculated, it is possible that the denominator of T may be equal to zero, i.e. the number of subcrossings equal to 8 may be zero. Hence, in that case, the statistic T computed during that replication is deleted from the analysis.

A.2. The simulated data for FARIMA processes with $H = 0.5, 0.6, \ldots, 0.9$, $N = 10000$ and 1000 replications.

Table 16 : The simulated data for FARIMA processes with $H = 0.5$, $N = 10000$ and 1000 replications. ($E = E_{\text{FARIMA}}$)

δ	level	mh	d.f.	$\overline{T}$	percentage	$\widehat{c}$	$\widehat{\nu}$	new percentage
($3E$)	1∼ 3	6	2	5.45923	65.20%	1.48005	3.68855	94.80%
2.39282		8	3	8.04830	57.90%	1.69822	4.73926	95.90%
($2.5E$)	1∼ 3	6	2	6.53906	53.90%	1.56405	4.18086	94.20%
1.99402		8	3	9.37502	46.20%	1.71560	5.46456	94.70%
($2E$)	1∼ 3	6	2	9.26109	33.00%	1.83721	5.04085	94.80%
1.59521		8	3	12.69380	26.30%	2.04923	6.19441	95.60%

Table 17 : The simulated data for FARIMA processes with $H = 0.6$, $N = 10000$ and 1000 replications. ($E = E_{\text{FARIMA}}$)

δ	level	mh	d.f.	$\overline{T}$	percentage	$\widehat{c}$	$\widehat{\nu}$	new percentage
($3E$)	1∼ 3	6	2	4.67579	72.70%	1.24395	3.75882	94.90%
2.41752		8	3	7.07621	64.40%	1.30537	5.42084	94.50%
($2.5E$)	1∼ 3	6	2	5.46942	64.60%	1.42752	3.83141	94.30%
2.01460		8	3	7.90563	58.20%	1.48871	5.31038	94.20%
($2E$)	1∼ 3	6	2	6.72928	52.80%	1.52129	4.42341	95.10%
1.61168		8	3	9.26218	48.70%	1.58972	5.82630	94.70%

Table 18 : The simulated data for FARIMA processes with $H = 0.7$, $N = 10000$ and 1000 replications. ($E = E_{\text{FARIMA}}$)

δ	level	mh	d.f.	$\overline{T}$	percentage	$\widehat{c}$	$\widehat{\nu}$	new percentage
($3E$)	1∼ 3	6	2	4.28453	76.30%	1.15525	3.70875	95.60%
2.50850		8	3	6.61417	68.00%	1.24320	5.32029	95.30%
($2.5E$)	1∼ 3	6	2	4.46514	75.20%	1.22465	3.64606	94.70%
2.09041		8	3	6.73874	66.70%	1.26097	5.34410	96.00%
($2E$)	1∼ 3	6	2	5.26264	67.40%	1.46524	3.59166	95.20%
1.67233		8	3	7.53580	61.20%	1.39705	5.39409	95.10%

Table 19 : The simulated data for FARIMA processes with $H = 0.8$, $N = 10000$ and 1000 replications. ($E = E_{\text{FARIMA}}$)

δ	level	mh	d.f.	$\overline{T}$	percentage	$\widehat{c}$	$\widehat{\nu}$	new percentage
($3E$)	1∼ 3	6	2	3.99750	81.20%	1.00471	3.97878	95.20%
2.74580		8	3	6.14255	74.20%	1.13846	5.39547	95.30%
($2.5E$)	1∼ 3	6	2	3.85176	80.90%	1.02708	3.75019	96.30%
2.28817		8	3	6.02159	76.40%	1.14192	5.27323	95.40%
($2E$)	1∼ 3	6	2	4.34523	74.90%	1.17471	3.69898	94.60%
1.83054		8	3	6.51744	68.90%	1.22103	5.33765	94.80%

Table 20 : The simulated data for FARIMA processes with $H = 0.9$, $N = 10000$ and 1000 replications. ($E = E_{\text{FARIMA}}$)

δ	level	mh	d.f.	$\overline{T}$	percentage	$\widehat{c}$	$\widehat{\nu}$	new percentage
($3E$)	1∼ 3	6	2	3.87077	81.60%	1.24899	3.09913	95.30%
3.43626		8	3	5.92926	75.60%	1.17384	5.05117	95.20%
($2.5E$)	1∼ 3	6	2	4.03521	80.90%	1.10244	3.66024	95.00%
2.86355		8	3	6.03239	76.00%	1.09951	5.48643	94.80%
($2E$)	1∼ 3	6	2	4.08460	79.20%	1.14305	3.57343	94.90%
2.29084		8	3	6.12230	74.30%	1.08895	5.62222	94.50%

Appendix B.

B.1. For $H > 1$, $\rho(k)$ diverges to infinity. This contradicts that $\rho(k)$ must be between -1 and 1. (see Section 2.1.)

Proof of the $\rho(k)$ diverges to infinity for $H > 1$. (Beran (1994))

$$\begin{aligned}\rho(k) &= \frac{1}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right] \\ &= \frac{1}{2}k^{2H}\left[(1+\frac{1}{k})^{2H} - 2 + (1-\frac{1}{k})^{2H}\right].\end{aligned}$$

The asymptotic behavior of $\rho(k)$ follows by Taylor expansion:

$$\rho(k) = \frac{1}{2}k^{2H} f(\frac{1}{k})$$

where $f(x) = (1+x)^{2H} - 2 + (1-x)^{2H}$.

If $0 < H < 1$ and $H \neq 1/2$, then we expand $f(x)$ at the origin.

$$\begin{aligned}f'(x) &= 2H(1+x)^{2H-1} - 2H(1-x)^{2H-1}, \\ f''(x) &= 2H(2H-1)(1+x)^{2H-2} + 2H(2H-1)(1-x)^{2H-2}, \\ f(0) &= 0,\ \ f'(0) = 0,\ \text{ and }\ f''(0) = 4H(2H-1).\end{aligned}$$

Hence, the first non-zero term in the Taylor expansion of $f(x)$ is equal to $2H(2H-1)x^2$, i.e.

$$f(x) \approx f(0) + f'(0)x + \frac{f''(0)x^2}{2} = 2H(2H-1)x^2.$$

As k tends to infinity, $\rho(k)$ is equivalent to $H(2H-1)k^{2H-2}$, i.e.

$$\frac{\rho(k)}{H(2H-1)k^{2H-2}} \to 1 \qquad\qquad k \to \infty.$$

B.2. Aggregated series

Because of the aggregated series are used in the moment method in Appendix B.3.3., some properties taken are summarized in this part from Leland et al. (1994).

Let $\{X_t, t > 0\}$ be a stationary stochastic process with mean μ, variance σ^2, and autocorrelation function $\rho(k)$. Now, we define the new covariance stationary time series or the corresponding aggregated series,

$$X^{(m)}(k) := \frac{1}{m} \sum_{t=(k-1)m+1}^{km} X_t, \qquad\qquad k = 1, 2, \ldots,$$

where m is the length of each block.

If the process $\{X_t, t > 0\}$ is self-similar and a property of stationary increments of self-similar processes of Section 2 can be used, then the sample mean can be written as

$$\begin{aligned} X^{(m)}(k) &= \frac{1}{m} \sum_{t=(k-1)m+1}^{km} X_t = \frac{1}{m}(Y_{km} - Y_{(k-1)m}) = \frac{1}{m}(Y_m - Y_0) \\ &\overset{d}{=} \frac{1}{m} \cdot m^H (Y_1 - Y_0) = m^{H-1}(Y_1 - Y_0). \end{aligned}$$

Therefore,

$$\mathrm{Var}(X^{(m)}(k)) = m^{2(H-1)}\mathrm{Var}(Y_1 - Y_0) = m^{2(H-1)}\sigma^2.$$

For N/m and m large enough ,

$$X^{(m)} \overset{d}{\sim} m^{H-1} S,$$

where $X^{(m)} = \{X^{(m)}(k), k = 1, 2, \ldots\}$ and S is a process which depends on the distribution of X but does not depend on m. Hence, for each m, the process $X^{(m)}$ is self-similar with self-similarity parameter H.

The process $\{X_t, t > 0\}$ is called *exactly self-similar* if for all m, $\mathrm{Var}(X^{(m)}) = m^{2(H-1)}\sigma^2$ and

$$\rho^{(m)}(k) = \rho(k), \;\; k \geq 0.$$

On the other hand, the process $\{X_t, t > 0\}$ is called *asymptotically self-similar* if for all k large enough,

$$\rho^{(m)}(k) \to \rho(k) \;\; as \;\; m \to \infty,$$

where $\rho(k)$ is given by (4).

For finite variance processes,

$$H = d + 1/2,$$

and for infinite variance processes,

$$H = d + 1/\alpha.$$

B.3. Some methods for estimation of the Hurst parameter

B.3.1. R/S method

The methods of time series analysis have been recognized as important tools for assisting in solving problems related to the management of water resources. R/S method is one of them and proposed in early periods. For the purpose of illustration, we summarize the method of R/S from Beran (1994).

The Nile River has been known for its long-term behavior. The famous hydrologist Hurst noticed these characteristics when he was investigating the question of how to regularize the flow of the Nile River. His discovery can be described as follows: Suppose we want to calculate the capacity of a reservoir such that it is ideal for the time span between t and $t+k$. To simplify matters, assume that time is discrete and that there are no storage losses (caused by evaporation, leakage, etc.). By ideal capacity we mean that we want to achieve the following: that the outflow is uniform, that at time $t+k$ the reservoir is as full as at time t, and the reservoir never overflows.

Let X_i denote the inflow at time i and $Y_j = \sum_{i=1}^{j} X_i$ the cumulative inflow up to the time j. Then the ideal capacity can be shown to be equal to

$$R(t,k) \;=\; \max_{0\le i\le k}\left[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)\right] - \min_{0\le i\le k}\left[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t)\right].$$

$R(t,k)$ is called the adjusted range where $(Y_{t+i} - Y_t) - \frac{i}{k}(Y_{t+k} - Y_t)$ is equal to $\sum_{j=t+1}^{t+i} X_j - \frac{i}{k}\sum_{j=t+1}^{t+k} X_j$. It can be regarded as the difference of the real total inflow and estimative total inflow for i units.

In order to study the properties that are independent of the scale, $R(t,k)$ is standardized by

$$S(t,k) = \sqrt{\frac{1}{k}\sum_{i=t+1}^{t+k}(X_i - \overline{X}_{t,k})^2},$$

where $\overline{X}_{t,k} = k^{-1}\sum_{i=t+1}^{t+k} X_i$. Note that $S^2(t,k)$ is equal to $(k-1)/k$ times the usual sample variance of $X_{t+1}, \ldots, X_{t+k}$. The ratio

$$R/S(t,k) = \frac{R(t,k)}{S(t,k)}$$

is called the *rescaled adjusted range* or *R/S-statistic.*

Calculate the ratio $R/S(t,k)$ for all possible values of t and k, i.e. for each k, there are $n-k+1$ replicates $R/S(k) = \{R/S(0,k), \ldots, R/S(n-k,k)\}$.

Hurst plotted the logarithm of R/S against several value of k. He observed that $\log[R/S]$ was scatter around a straight line with a slope that exceed 1/2 for large k. If X_i is long-range dependent, then

$$\log E[R/S] \approx a + H \log k, \qquad \text{with } H > \tfrac{1}{2} \text{ as } k \to \infty,$$

where H is the Hurst parameter.

From a statistical point of view it, the following difficulties aries (Embrechts and Maejima (2002)):

(1) It is difficult to decide from which k the asymptotic behavior starts, and so how many points are to be included in the least squares regression.

(2) For finite samples, the distribution of R/S is neither normal nor symmetric, and the values of R/S for different time points and lags are not independent from each other. This raises the equation of whether least squares regression is appropriate.

(3) Only very few values of R/S can be calculated for large values k, thus making the inference less reliable even at large lags.

Because of these problems, Beran (1994) concludes that it seems difficult to derive the results of statistical inference based on the R/S method. In Appendix B.4.1., we use the Nile River data to explain the difficulty of choosing the cut-off point by the estimate of H obtained by fitting a least squares line. In addition, we also point out the distribution of R/S is neither normal nor symmetric.

In the case of the Nile River data, Hurst observed that the parameter H is equal to 0.91 (Gaudenta Sakalauskien (2003)). Today the R/S analysis is mostly used for the hydrological studies such as river flow, precipitation, temperature, etc. Based on the above illustration, the R/S method can be summarized as follows:

Step 1. Formulate the partial sum of the series W_{ik}, where

$$W_{ik} = (X_{t+1} + X_{t+2} + \ldots + X_{t+i}) - \frac{i}{k}(X_{t+1} + X_{t+2} + \ldots + X_{t+k})$$

for $i = 1, 2, \ldots, k$. For $i = 0$, $W_{0k} = 0$.

Step 2. Find $R(t,k) = \max(0, W_{1k}, W_{2k}, \ldots, W_{kk}) - \min(0, W_{1k}, W_{2k}, \ldots, W_{kk})$.

Step 3. Calculate $S(t,k)$, which is equal to the square root of $(k-1)/k$ times the usual sample variance of $X_{t+1}, \ldots, X_{t+k}$, ie.

$$S(t,k) = \sqrt{\frac{1}{k}\sum_{i=t+1}^{t+k}(X_i - \overline{X}_{t,k})^2},$$

where $\overline{X}_{t,k} = k^{-1}\sum_{i=t+1}^{t+k} X_i$.

Step 4. Calculate the ratio $R/S = R(t,k)/S(t,k)$.

Step 5. Step 1 to 4 are repeated for several k.

Step 6. Plot $\log(R/S)$ against $\log k$ and use least square fit to evaluate the slope of the straight line which is equal to the parameter $\widehat{H}$.

Next, in order to understand the procedures for the R/S method, an example is given:

Example 3 : Estimation of parameter H for R/S method

Assume that the process $X(t)$ is
1157, 1088, 1169, 1169, 984, 1322, 1178, 1103, 1211, 1292, 1124, 1171, 1133, 1227, 1142, 1216, 1259, 1299, 1232, 1117.

In this case, the length of time series, n, is equal to 20.

For $k = 3$:

$$\begin{aligned}
W_{13} &= X_{t+1} - \frac{1}{3}(X_{t+1} + X_{t+2} + X_{t+3}), \\
W_{23} &= (X_{t+1} + X_{t+2}) - \frac{2}{3}(X_{t+1} + X_{t+2} + X_{t+3}), \\
W_{33} &= (X_{t+1} + X_{t+2} + X_{t+3}) - \frac{3}{3}(X_{t+1} + X_{t+2} + X_{t+3}) = 0, \\
R(t,3) &= \max(0, W_{13}, W_{23}, W_{33}) - \min(0, W_{13}, W_{23}, W_{33}), \\
S(t,3) &= \sqrt{\frac{1}{3}\sum_{i=t+1}^{t+3}\left(X_i - \overline{X}_{t,3}\right)^2}, \text{ where } \overline{X}_{t,3} = \frac{X_{t+1} + X_{t+2} + X_{t+3}}{3}, \text{ and} \\
R/S(t,3) &= R(t,3)/S(t,3).
\end{aligned}$$

The methods are the same for other k, $k = 4, \dots, 15$.

Therefore,

$$\begin{aligned}
R/S(3) &= \{R/S(0,3), R/S(1,3), \dots, R/S(17,3)\} \\
&= \{1.40083, 1.41421, 1.41421, 1.26151, 1.28048, 1.33152, 1.3499, 1.27872, \\
&\quad 1.23906, 1.36124, 1.39101, 1.29503, 1.40888, 1.40417, 1.31759, 1.23923, \\
&\quad 1.29585, 1.31723\}, \\
&\vdots \\
R/S(15) &= \{R/S(0,15), R/S(1,15), \dots, R/S(5,15)\} \\
&= \{3.59363, 3.31974, 2.74109, 2.64638, 2.7511, 3.435\}.
\end{aligned}$$

Now, we plot $\log(R/S)$ against $\log k$ and find the slope using least square fit. By the result of calculation in this case, the regression line is $\log[R/S] = -0.146643 + 0.593122 \log[k]$ and the slope is $\widehat{H} = 0.593122$.

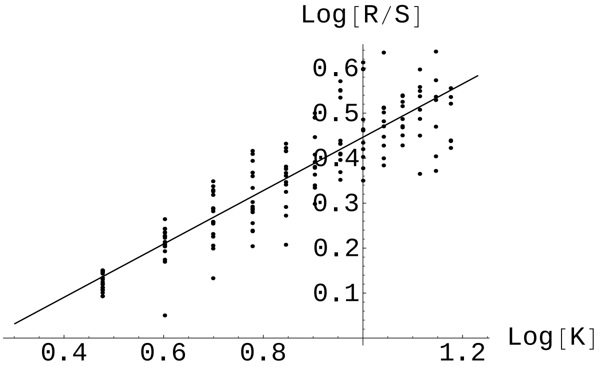

Figure 10 : The result of the example for R/S method.

B.3.2. Detrended Fluctuation Analysis

The Detrended Fluctuation Analysis (DFA) proposed by Peng et al. (1994) has been established as an important tool for the detection of long-range (auto-)correlations in time series with non-stationarities. It has been applied to diverse fields of DNA, heart rate dynamics, human gait, long-time weather records, cloud structure, economical time series, etc..

In Kantelhardt et al. (2002), the method of the multifractal characterization of nonstationary time series is proposed, namely multifractal DFA (MF-DFA), which is based on a generalization of the detrended fluctuation analysis (DFA). The different orders of the DFA technique are studied, that allow to eliminate different orders of trends. For stationary time series, $h(2)$ is identical to the Hurst parameter H (see Feder (1988)). Here, in order to get the parameter H of self-similarity and simplify the question, we consider the $q = 2$ and disregard a short part at the end of the profile may remain. Hence, the MF-DFA method with $q = 2$ consists of five steps which is illustrated as follows:

Consider a stochastic process $\{X_t, t > 0\}$ with equidistant measurements. In most application, the index t will correspond to the time of the measurements.

Step 1: Determine the profile

$$Y(i) = \sum_{t=1}^{i} \left[X_t - \overline{X}\right], \qquad i = 1, \ldots, N,$$

where $\overline{X}$ is the mean of the series $X_1, \ldots, X_N$.

Step 2: Divide the profile $Y(i)$ into $N_s \equiv \mathrm{int}(N/s)$ nonoverlapping segments of equal length s. Since the length N of the series is often not a multiple of the considered time scale s, a short part at the end of the profile may remain. For the purpose of simplifying the question, this part of the series is disregard.

Step 3. For each segment ν, $\nu = 1, \ldots, N_s$,

$$F^2(s, \nu) \equiv \frac{1}{s} \sum_{i=1}^{s} \left\{Y[(\nu - 1)s + i] - y_\nu(i)\right\}^2,$$

where $y_\nu(i)$ is the fitting polynomial in segment ν. Linear, quadratic, cubic, or higher order polynomials can be used in the fitting procedure (conventionally called DFA1, DFA2, DFA3, ...). By construction, $F_2(s)$ is only defined for $s \geq m + 2$, where m is the order of fitting polynomial.

Step 4. Calculate the fluctuation function

$$F_2(s) = \sqrt{\frac{1}{N_s} \sum_{\nu=1}^{N_s} F^2(s, \nu)}.$$

Step 5. We are interested in how $F_2(s)$ depends on the time scale s. Hence, Steps 2 to 4 must be repeated for several time scales s.

Step 6. Plot $\log(F_2(s))$ against $\log(s)$ and use least square fit to evaluate the slope of the straight line which is equal the parameter H.

In fact, Steps 3 and 4 can be merged into

$$F_2(s) = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}[Y(i) - y_\nu(i)]^2},$$

where $N_t = s \times N_s$ and $F_2(s)$ is the square root of MSE.

Next, in order to understand the procedures to estimate the parameter H for the DFA method, an example is given:

Example 4 : Estimation of the parameter H for the DFA method

Assume that the process $X(t)$ is
1157, 1088, 1169, 1169, 984, 1322, 1178, 1103, 1211, 1292, 1124, 1171, 1133, 1227, 1142, 1216, 1259, 1299, 1232, 1117. 1157, 1155, 1232, 1083, 1020, 1394, 1196, 1148, 1083, 1189, 1133.

In this case, the length N of the series is 30 and mean $\overline{X}$ is 1174.2.

For $s = [10^{0.7}] = 5$, then we divide the time series into blocks of length 5 and the number of blocks is $N_s = 6$.

The cut-off points we chose are $s = [10^{0.7}], [10^{0.8}], \ldots, [10^{1.1}]$ and the fitting polynomial in segment is linear, i.e. DFA1. Then, the $F_2(s)$ are

$$\begin{aligned} & (F_2(5), F_2(6), F_2(7), F_2(10), F_2(12)) \\ = & (48.4352, 45.2424, 62.6238, 67.227, 83.1478). \end{aligned}$$

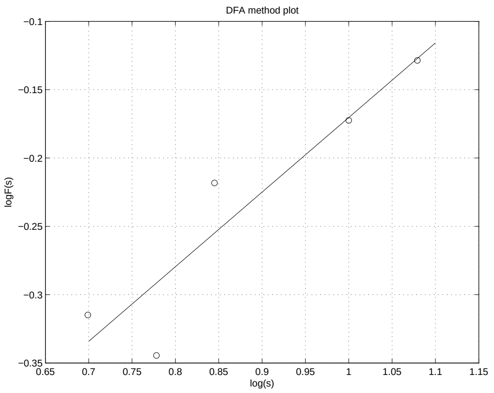

Figure 11 : The figure for the example of DFA method.

Hence, we can plot $\log(F_2(s))$ against $\log(s)$ and use least square fit to evaluate the slope of the straight line which is equal the parameter $\widehat{H} = 0.54568$.

B.3.3. Moment method

Given a time series $\{X_t, t > 0\}$, of length N, and consider the aggregated series, i.e.

$$X^{(m)}(k) := \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \;\; k = 1, 2, \ldots, [N/m],$$

where m is the length of each block.

We take the n^{th} absolute moment of this series,

$$AM_n^{(m)} = \frac{1}{N/m} \sum_{k=1}^{N/m} |X^{(m)}(k) - \overline{X}|^n,$$

where $\overline{X}$ is the overall series mean. The aggregated series $X^{(m)}$ asymptotically behaves like $Cm^{n(H-1)}$ for large m, and thus $AM_n^{(m)}$ is proportional to $m^{n(H-1)}$.

For successive values of m, the sample absolute moment of the aggregated series is plotted versus m on a log-log plot. The result should be a straight line with a slope of $n(H-1)$. Hence, the estimation of H is equal to (slope$/n$)$+1$. This method is used for $n = 1$, and it reduces to Absolute Value method (in short Absolute). For $n = 2$, it reduces to the Variance method (in short Variance).

We assume that both N and N/m are large to ensures that both the length of each block and the number of blocks is large. In practice, the points at the very low and high ends of the plots are not used to fit the least-squares line. If the low end of the plot is used, the short-range effects can distort the estimates of H and there are too few blocks to get reliable estimates of $AM_n^{(m)}$ at the very high end of the plot. The choices of the cut-offs are $10^{0.7}$ and $10^{2.5}$ (see Taqqu (1998)). In Section 5, we use $m = [10^{0.7}], \ldots, [10^{2.5}]$ to estimate the parameter H for the FBM and FARIMA processes, where [ ] denotes the greatest integer function.

Based on the above illustration, the Moment method can be summarized as follows:

Step 1. Calculate the overall series mean $\overline{X}$ and divide the time series of length N into blocks of length m, i.e.

$$(X_1, X_2, \ldots, X_m), (X_{m+1}, \ldots, X_{2m}), \ldots, (X_{([N/m]-1)m+1}, \ldots, X_{[N/m]m}).$$

Step 2. Average the series over each block, then we obtain the aggregated series, i.e.

$$X^{(m)}(1), X^{(m)}(2), \ldots, X^{(m)}([N/m]),$$

where $X^{(m)}(k) = \frac{1}{m}(X_{(k-1)m+1} + \ldots + X_{km})$.

Step 3. Subtract $\overline{X}$ from the series of the Step 2 and then take absolute value and power n. Then we obtain the new series

$$|X^{(m)}(1) - \overline{X}|^n, \;\; |X^{(m)}(2) - \overline{X}|^n, \;\; \ldots, \;\; |X^{(m)}([N/m]) - \overline{X}|^n.$$

Step 4. Calculate the mean of the series of Step 3, then we can get the value of $AM_n^{(m)}$.

Step 5. Steps 1 to 4 are repeated for several length m.

Step 6. Plot $\log(AM_n^{(m)})$ against $\log(m)$ and use least square fit to evaluate the slope of the straight line.

Step 7. Calculate $\widehat{H} = \frac{\text{slope}}{n} + 1$.

Next, in order to understand the procedures to estimate the parameter H for the moment method, an example is given:

Example 5 : Estimation of the parameter H by momoent method

Assume that the process $X(t)$ is
1157, 1088, 1169, 1169, 984, 1322, 1178, 1103, 1211, 1292, 1124, 1171, 1133, 1227, 1142, 1216, 1259, 1299, 1232, 1117. 1157, 1155, 1232, 1083, 1020, 1394, 1196, 1148, 1083, 1189, 1133.

In this case, we use the absolute value method ($n = 1$). The length N of the series is 30 and mean $\overline{X}$ is 1174.2.

For $m = [10^{0.7}] = 5$, then we divide the time series into blocks of length 5, i.e.

$$(X_1, X_2, \ldots, X_5), (X_6, \ldots, X_{10}), \ldots, (X_{25}, \ldots, X_{30}).$$

Then, the aggregated series is

$$\begin{aligned} & (X^{(5)}(1), X^{(5)}(2), \ldots, X^{(5)}(6)) \\ = \; & (1113.4, 1221.2, 1159.4, 1224.6, 1176.8, 1149.8), \end{aligned}$$

and we can obtain

$$AM_1^{(5)} = \frac{1}{6}\sum_{k=1}^{6} |X^{(5)}(k) - \overline{X}| = 33.3333.$$

Repeat the same ways for other m, $m = [10^{0.8}], \ldots, [10^{1.1}]$.

Therefore, we can get

$$\begin{aligned} & (AM_1^{(5)}, AM_1^{(6)}, AM_1^{(7)}, AM_1^{(10)}, AM_1^{(12)}) \\ = \; & (33.3333, 24.16, 16.3571, 11.8667, 6.125). \end{aligned}$$

Finally, we plot $\log(AM_1^m)$ against $\log m$ and find the slope using least square fit.

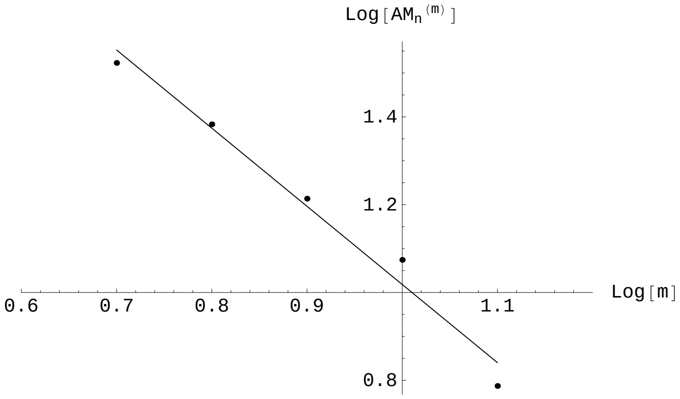

Figure 12 : The result of the example for moment method with $n = 1$.

B.4. Nile River data

The Nile River is one of the world's great assets. Since ancient time, the Nile River has been known for its long-term behavior. The long periods of dryness are followed by the long periods of floods. As far as floods are concerned, it has positive and negative significant impacts on the Nilotic countries. On the positive side, it can provide water for drinking, industrial activities, agricultural activities and etc.. On the negative side, floods have great destructive forces on the people and environment. In addition to knowing flood impacts and mitigating the negative impacts, people should properly utilize the floodwater (see Shawki et al. (2005)).

According to data of yearly minimal water levels of the Nile River for the years 622-1284, the time series, sample autocorrelations, and sample partial autocorrelations are sketched for illustrations. From Figure 13, the sample autocorrelations indeed exhibit strong long-range dependence. Following are estimations of the parameter H of the Nile River data by using some methods.

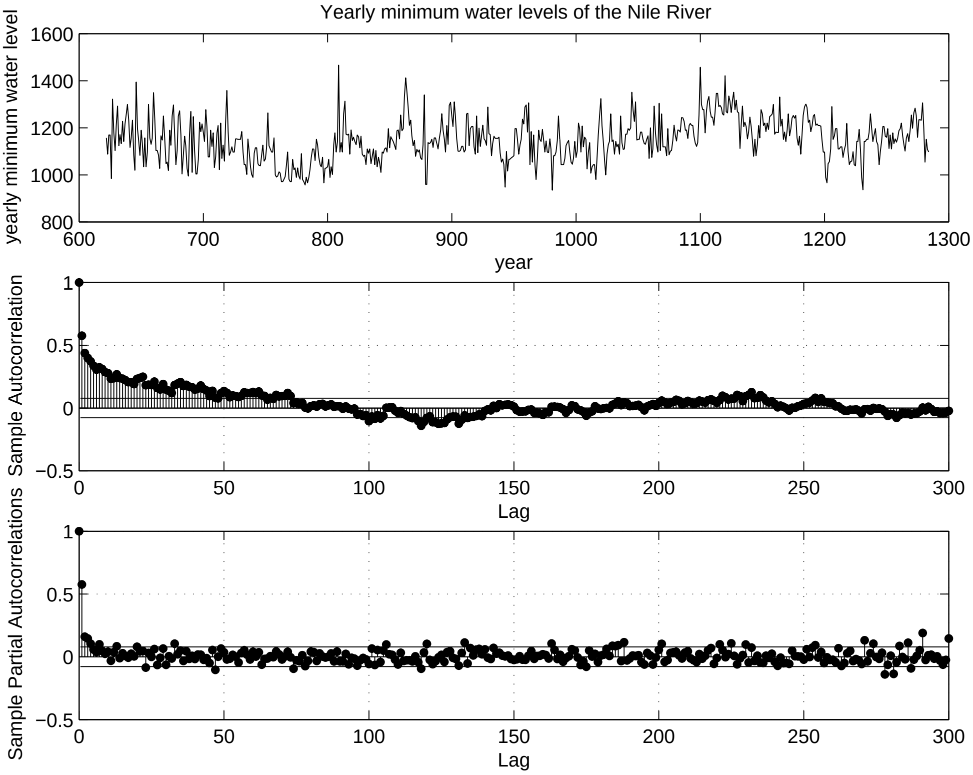

Figure 13 : Yearly minimum water levels of the Nile River.

B.4.1. R/S method for the Nile River data

In this part, the R/S method with $k = 10, 20, ..., 300$, $t = 0, 50, ..., 50[\frac{N-k}{50}]$ is used to estimate the parameter H for the Nile River data. The R/S plot is presented in Figure 14 and the estimation of H is equal to 0.905822. Figure 15 shows that the distribution of R/S is neither normal nor symmetric as mentioned in Appendix B.3.1., for Nile River data with $k = 10, 20, ..., 300$, and $t = 0, 50, ..., 50[\frac{N-k}{50}]$.

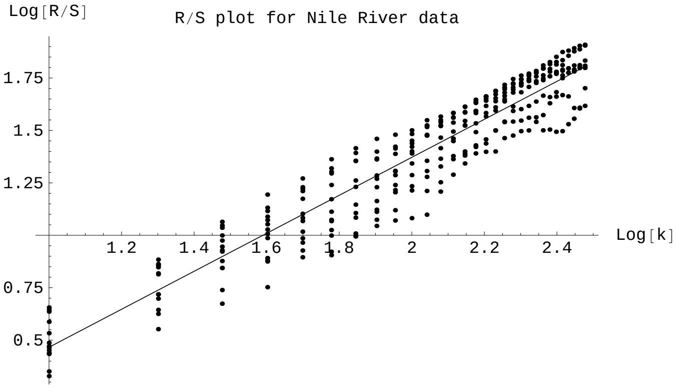

Figure 14 : R/S plot for Nile River data with $k = 10, 20, ..., 300$, $t = 0, 50, ..., 50[\frac{N-k}{50}]$ and $\widehat{H} = 0.905822$.

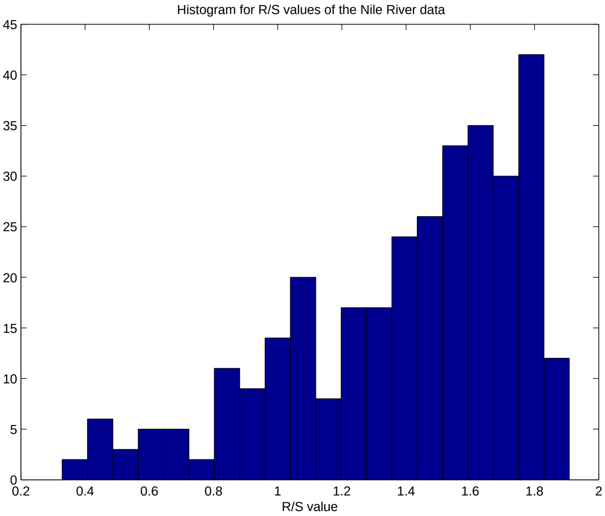

Figure 15 : Histogram for R/S values of the Nile River data with $k = 10, 20, ..., 300$, and $t = 0, 50, ..., 50[\frac{N-k}{50}]$.

In the following, the different cut-off points of k and t are utilized to evaluate the values of H for the Nile River data and the results are listed in Table 21. It seems difficult to derive the results of statistical inference based on the R/S method indeed.

Table 21 : The estimate of H of different cut-off point of k and t for the Nile River data.

k	t	H
10, 20,. . . , 60	0, 10, 20,. . . , A	0.803083
10, 20,. . . , 60	0, 50, 100,. . . , B	0.833445
10, 20,. . . , 60	0, 100, 200,. . . , C	0.754114
10, 20,. . . , 100	0, 10, 20,. . . , A	0.821269
10, 20,. . . , 100	0, 50, 100,. . . , B	0.829434
10, 20,. . . , 100	0, 100, 200,. . . , C	0.715886
10, 20,. . . , 150	0, 10, 20,. . . , A	0.863121
10, 20,. . . , 150	0, 50, 100,. . . , B	0.869631
10, 20,. . . , 150	0, 100, 200,. . . , C	0.802018
10, 20,. . . , 200	0, 10, 20,. . . , A	0.893457
10, 20,. . . , 200	0, 50, 100,. . . , B	0.900732
10, 20,. . . , 200	0, 100, 200,. . . , C	0.856659
10, 20,. . . , 250	0, 10, 20,. . . , A	0.903146
10, 20,. . . , 250	0, 50, 100,. . . , B	0.909349
10, 20,. . . , 250	0, 100, 200,. . . , C	0.884909
10, 20,. . . , 300	0, 10, 20,. . . , A	0.901177
10, 20,. . . , 300	0, 50, 100,. . . , B	0.905822
10, 20,. . . , 300	0, 100, 200,. . . , C	0.884932
10, 20,. . . , 350	0, 10, 20,. . . , A	0.89726
10, 20,. . . , 350	0, 50, 100,. . . , B	0.901345
10, 20,. . . , 350	0, 100, 200,. . . , C	0.89191
10, 20,. . . , 400	0, 10, 20,. . . , A	0.893802
10, 20,. . . , 400	0, 50, 100,. . . , B	0.898102
10, 20,. . . , 400	0, 100, 200,. . . , C	0.889256

[ ] denotes the greatest integer function.
$A = 10[\frac{N-k}{10}]$, $B = 50[\frac{N-k}{50}]$, and $C = 100[\frac{N-k}{100}]$.

B.4.2. DFA method for the Nile River data

In this part, we use DFA method to estimate the parameter H for the Nile River data with $n = [10^{0.7}]$, $[10^{0.8}]$,...,$[10^{2.5}]$ where [ ] denotes the greatest integer function. The result is showed in Figure 16 and $\widehat{H}$ is equal to 0.89714.

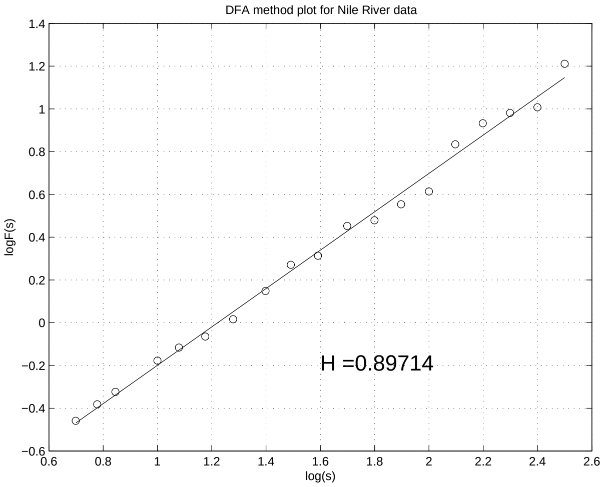

Figure 16 : DFA plot for Nile River data with $n = [10^{0.7}]$, $[10^{0.8}]$,...,$[10^{2.5}]$.