

# TESTING CONSTANCY OF THE HURST EXPONENT OF SOME LONG MEMORY STATIONARY GAUSSIAN TIME SERIES

*F. Lombard*

Centre for Business Mathematics and Informatics, North–West University,  
Potchefstroom, South Africa  
email: *fredl@telkomsa.net*

*and*

*J.L. Robbertse*

Department of Statistics, University of Johannesburg, South Africa  
email: *wickesr@uj.ac.za*

---

**Key words:** Hurst exponent, Fractional Gaussian noise, Changepoints

---

**Summary:** Long-range dependence is often observed in stationary time series. The *Hurst exponent* then characterizes the long term features of the data, which implies that changes in its value could have implications for the long term behaviour of the series. In this paper we propose and apply tests to detect changes over time in the Hurst exponent of long memory Gaussian time series, in particular fractional Gaussian noise and fractionally integrated Gaussian white noise.

---

## 1. Introduction

A strictly stationary time series  $X_1, \dots, X_n$  is said to exhibit *long range dependence* or *long memory* if its autocorrelations  $\rho(k)$  decrease to zero as a power of the lag  $k$  but so slowly that their sum is not absolutely convergent. Such series have been observed in many contexts, for instance hydrology (Hurst, 1951), ethernet traffic (Leland, et al. 1994, Willinger et al., 1995), wind speeds (Haslett and Raftery, 1989) and stock returns (Granger and Hyung, AMS:62F03, 62M10)

2004), to name just a few. Fractional Gaussian noise (FGN), first studied in depth by Mandelbrot and Van Ness (1968), is possibly the best known among the long range dependent time series. FGN has autocorrelations  $\rho(k)$  proportional to  $k^{2(\alpha-1)}$ ,  $0 < \alpha < 1$ . The parameter  $\alpha$  is known as the "Hurst exponent" of the series and we are particularly interested in the case  $1/2 < \alpha < 1$  where the autocorrelations, all positive, sum to infinity. Another well known class of long memory processes are the fractionally integrated Gaussian white noises, first introduced by Granger and Joyeux (1980). These series, generally denoted by  $I(d)$ , also have autocorrelations  $\rho(k) \propto k^{2(\alpha-1)}$ , where  $\alpha = d + 1/2$ .

Time series that have been shown to exhibit long memory typically consist of rather large numbers of observations. Granger and Hyung (2004) put forward a theory that partially attributes long memory to occasional structural breaks in long time series. They also note that the resulting long memory parameter can exhibit a time dependence. Given a long FGN or  $I(d)$  series, it is therefore natural to enquire whether  $\alpha$  is constant over the full extent of the data or whether its value perhaps changes one or more times. It is this question that we investigate in the present paper. A test to detect changes in the value of  $\alpha$  has been proposed by Beran and Terrin (1996) but their results were shown to be incorrect by Horvath and Shao (1999).

To check constancy of  $\alpha$  across the full data set one can estimate  $\alpha$  by maximum likelihood (ML) from each of the series  $X_1, \dots, X_k$  and  $X_{k+1}, \dots, X_n$  for  $k = m, \dots, n - m$  where  $m$  is small compared to  $n$ . Denoting the difference between the ML estimates obtained at each  $k$  by  $\Delta_k$ , a (likelihood ratio) test of constancy against an alternative of a single upward or downward jump in the value of  $\alpha$  can then be based on the largest observed

value of  $|\Delta_k|$ . This is the approach put forward by Beran and Terrin (1996) and which is implicit in the work of Horvath and Shao (1999). However, when  $n$  is large there are often computational problems involved in finding the required ML estimates. In a stretch of data of length  $n = 4,000$  for instance, a covariance matrix of order  $16 \times 10^6$  must be inverted at each of 2,800  $k$ -values (if we take  $m = 100$ , for instance) in order to implement the methodology. In the three data sets analysed in the present paper, we have  $n = 3,121$ ,  $n = 4,000$  and  $n = 660$  respectively.

The purpose of the present paper is to show that substantial further conceptual simplifications and computational benefits can accrue if one works instead with the first differences of the series in question,  $Y_t = X_t - X_{t-1}$ ,  $t = 1, 2, \dots$ . In Section 2 we note that  $Y_t$  is a short memory process and that we may assume approximate independence between the maximum likelihood estimates of  $\alpha$  obtained from relatively short disjoint stretches of data. In Section 3 we show how a number of existing change point test statistics can be applied more or less directly to test constancy of  $\alpha$ . Section 4 contains change point analyses of three long memory series, each illustrating a specific aspect involved in the analysis of such data. For simplicity of presentation we restrict our attention in this paper in the main to fractional Gaussian noise.

## 2. Differenced fractional Gaussian noise

Standard fractional Gaussian noise (FGN) is a stationary Gaussian process  $X_t$  with unit variance and autocorrelation function

$$\rho_X(k) = (|k+1|^{2\alpha} + |k-1|^{2\alpha} - 2|k|^{2\alpha})/2$$

for  $k \geq 0$ . By Taylor expansion, for  $k$  large,

$$\rho_X(k) \sim \alpha(2\alpha - 1)k^{2(\alpha-1)} \quad (1)$$

where  $\sim$  indicates that the ratio of the two sides converges to 1 as  $k \rightarrow \infty$ .

Then

$$\sum_{k \in \mathbb{Z}} \rho_X(k) = \infty$$

for  $1/2 \leq \alpha < 1$ . Henceforth, we confine our attention to the latter range of  $\alpha$  values. Differenced fractional Gaussian noise (DFGN), defined by

$$Y_t = X_t - X_{t-1}, t = 1, 2, \dots,$$

has variance

$$\text{var}(Y_t) = 4(1 - 2^{2\alpha-2}) \quad (2)$$

and covariance function

$$\gamma_\alpha(k) = (4|k+1|^{2\alpha} + 4|k-1|^{2\alpha} - 6|k|^{2\alpha} - |k-2|^{2\alpha} - |k+2|^{2\alpha})/2,$$

which is negative for all  $k \geq 1$ . It follows from (1), after some calculation, that for large  $k$

$$\gamma_\alpha(k) \sim -Ck^{2(\alpha-2)} \quad (3)$$

where

$$C = 2\alpha(2\alpha-1)(2\alpha-2)(2\alpha-3) > 0.$$

Thus,

$$-\infty < \sum_{k \geq 1} k\gamma_\alpha(k) < 0$$

so that DFGN is seen to be a genuine *short memory* stationary process. Table 1 shows values of the autocorrelations  $\gamma_\alpha(k)/\gamma_\alpha(0)$  for  $k = 1, \dots, 5$  and four values of  $\alpha$ . The rapid decrease in autocorrelation with increasing lag is evident.

**Table 1** Autocorrelations of DFGN at lags  $k = 1, \dots, 5$ .

$k$	$\alpha = 0.60$	$\alpha = 0.70$	$\alpha = 0.80$	$\alpha = 0.90$
1	-0.454	-0.404	-0.348	-0.286
2	-0.033	-0.065	-0.093	-0.116
3	-0.006	-0.014	-0.024	-0.034
4	-0.002	-0.006	-0.011	-0.017
5	-0.001	-0.003	-0.006	-0.010

Let  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$  be an observed DFGN series. The part of the log likelihood function involving unknown parameters is

$$\ell(\alpha, \sigma^2; \mathbf{Y}) = -\frac{1}{2} \left[ n \log(\sigma^2) + \log \det(\boldsymbol{\Sigma}) + \frac{1}{\sigma^2} \mathbf{Y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{Y} \right] \quad (4)$$

where the matrix  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\alpha)$  has  $(i, j)$  th element equal to  $\rho(|i - j|, \alpha) \times (8 - 2^{2\alpha+1})$  and where  $\sigma$  is an unknown scale parameter. Given a (long) series  $\mathbf{Y}$  of length  $n$ , we consider  $B$  disjoint contiguous shorter series of length  $m \ll n$ ,  $\mathbf{Y}_b = (Y_{(b-1)m+1}, Y_{(b-1)m+2}, \dots, Y_{bm})^\top$ ,  $b = 1, 2, \dots, B$  with  $mB = n$ . For  $b = 1, \dots, B$  denote by  $\hat{\alpha}_b$  the ML estimate of  $\alpha$  obtained from the series  $\mathbf{Y}_b$ . There are no problems, computational or otherwise, in obtaining these estimates from relatively short series of lengths  $m$ . A further benefit of working with DFGN rather than FGN relates to the presence of slowly varying trends in the mean of the FGN series. Differencing largely eliminates such trends, hence also the effect they would have if  $\alpha$  were estimated directly from the FGN series.

We argue in the Appendix on theoretical grounds, supported by Monte Carlo simulation results, that the series of estimates  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_B$  can often be treated as statistically independent observations provided the block lengths are "sufficiently large". The idea to use such blocks of data is not new, however - see, for instance, Beran and Terrin (1994). Then the constancy of  $\alpha$  can be tested by applying existing tests, discussed in Section 3,

that are applicable to independent observations to the series  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_B$ . Sometimes, however, the latter series will exhibit some short range negative autocorrelation inherited from the negative autocorrelation in the DFGN series. If this negative autocorrelation is not negligible, the tests will lose some power to detect changepoints. The loss can be avoided by making a relatively simple adjustment to the test statistics - see Sections 3 and 4.2 for details and an example.

Finally, suppose the hypothesis of constancy is rejected. Then we can estimate the block in which the putative change occurs, but not the *changepoint* within such a block. An additional step is required to find such an estimate. Thus, it is advantageous from an estimation point of view that the blocks be as small as possible without forfeiting the independence property of the time series  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_B$  but also large enough to provide estimates that are not overly variable.

### 3. Test statistics

Let  $\bar{\alpha}$  denote the sample mean and  $s$  the sample standard deviation of the estimates  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_B$ . Tests for constancy of  $\alpha$  can be based on the standardized cumulative sums (cusums)

$$T_b = \sum_{i=1}^b (\hat{\alpha}_i - \bar{\alpha}) / ((B-1)^{1/2} s), \quad b = 1, 2, \dots, B-1. \quad (5)$$

A cusum plot, which consists of a plot of  $T_b$  against  $b$ , is the best known graphical tool for detecting possible changepoints. If, for instance,  $\alpha$  increases in value after some point  $\tau$  the cusum plot should reach a minimum near  $\tau$  and change sharply in an upward direction thereafter. This is because the terms  $\hat{\alpha}_b - \bar{\alpha}$ ,  $b \leq \tau$ , would then typically tend to be negative rather than positive, resulting in a downward sloping cusum. After the changepoint is reached the terms  $\hat{\alpha}_b - \bar{\alpha}$  would tend to be positive rather than negative, giving rise to a

change in an upward direction. If no change in  $\alpha$  occurs, the cusum plot will tend to show fluctuations with large variability and no clear trend.

Lombard (1987) proposed statistics involving quadratic forms in  $T_b$  to test various changepoint hypotheses. While these test statistics were framed in the context of ranked data, their asymptotic distributions apply equally well to the standardised cusums  $T_b$  in (5). One well known test statistic (Lombard, 1987, Section 2.3), is

$$m_1 = (B - 1)^{-1} \sum_{b=1}^{B-1} T_b^2. \quad (6)$$

The motivation for using  $T_b^2$  rather than  $T_b$  in the sum is to capture a change irrespective of whether it occurs in an upward or downward direction. The statistic  $m_1$  is appropriate when only one  $\alpha$ -changepoint is present in the data, that is, when the alternative

$$\alpha_b = \begin{cases} \xi_1, & 1 \leq b \leq \tau \\ \xi_2, & \tau < b \leq B, \end{cases}$$

with  $\xi_1 \neq \xi_2$ , is thought to be appropriate. Here  $\alpha_j$  denotes the true value of the Hurst parameter underlying the series  $\mathbf{Y}_j$  that constitutes block  $j$ . An alternative that accommodates two changepoints is

$$\alpha_b = \begin{cases} \xi_1, & 1 \leq b \leq \tau_1 \\ \xi_2, & \tau_1 < b \leq \tau_2 \\ \xi_3, & \tau_2 < b \leq B \end{cases} \quad (7)$$

with arbitrary real numbers  $\xi_1, \xi_2$  and  $\xi_3$ . A test statistic for detecting this type of alternative is (Lombard, 1987, Section 3)

$$m_2 = 2m_1 - \left( B^{-1} \sum_{b=1}^{B-1} T_b \right)^2 \quad (8)$$

A special case of (7), namely the "square wave" alternative is obtained upon requiring that  $\xi_1, \xi_3 < \xi_2$  or  $\xi_1, \xi_3 > \xi_2$ . For this case Lombard (1987, Section 5.3) proposed the statistic

$$U^2 = (B - 1)^{-1} \sum_{b_1=1}^{B-1} \left\{ \sum_{b_2=b_1}^B (T_{b_2} - T_{b_1})^2 \right\} \quad (9)$$

The large sample (i.e. large  $B$ ) distributions of the statistics  $m_1$ ,  $m_2$  and  $U^2$  are known. For ease of use some of the asymptotic percentage points are given in Table 2.

**Table 2** Asymptotic percentage points of three test statistics

	10%	7.5%	5%	2.5%	1%
$m_1$	0.347	0.394	0.461	0.584	0.743
$m_2$	0.486	0.542	0.622	0.764	0.958
$U^2$	0.152	0.166	0.187	0.222	0.268

Rather than relying on asymptotic results one may use a permutational approach, provided the independence assumption holds to a satisfactory degree. Consider, for instance, the statistic  $m_1$  from (6). The permutation approach involves calculating  $m_1$  for each of a large number,  $N$ , of random permutations  $\hat{\alpha}_{\pi(1)}, \hat{\alpha}_{\pi(2)}, \dots, \hat{\alpha}_{\pi(B)}$  of the estimates  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_B$ . Denoting the corresponding statistic values by  $m_1(j)$ ,  $j = 1, \dots, N$ , then the estimated permutation  $p$ -value is

$$\hat{p} = N^{-1} \times \text{number of } m_1(j) \geq m_{1,obs}$$

where  $m_{1,obs}$  denotes the value of the statistic  $m_1$  calculated on the series of estimates  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_B$  in their original time order. We used  $N = 10,000$  permutations in all our applications.

One can check the appropriateness of the independence assumption in any specific instance by inspecting a plot of the periodogram of the  $\hat{\alpha}$  series and testing it for constancy or by calculating the first few autocorrelations and testing these for significance. Failure of the independence assumption

typically manifests itself in a periodogram that exhibits a downward trend towards the lower frequencies. This is caused by some residual negative autocorrelation from the differenced series. In the presence of autocorrelation in the  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_B$  series, the standard deviation  $s$  in (5) should be replaced by the square root of  $\hat{f}(0)$ , an estimate of the spectral density at the zero frequency - see, for instance, Lombard and Hart (1992) and Section 4.2 below.

## 4. Application to data

### 4.1 Ethernet traffic data

These data are taken from Willinger et al. (1995), who discussed their self-similar nature. We confine attention to their "BC-pAug89" data:

<http://ita.ee.lbl.gov/html/contrib/BC.html>,

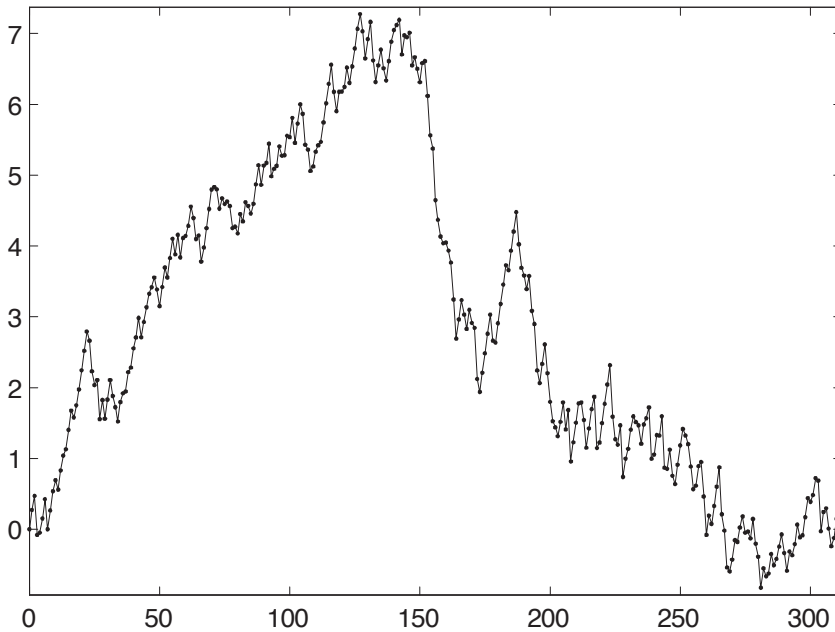
and computed the number of packet arrivals on the ethernet traffic network in each of 3121 consecutive one second time intervals. The series of 3120 successive differences constitute our  $\mathbf{Y}$  data. Table 3 shows the results of applying the three tests considered in Section 3 using  $B = 312$  blocks of length 10. (There is no evidence of serial correlation in the series of  $\hat{\alpha}$  estimates.)

**Table 3** Test for change in  $\alpha$ : Ethernet traffic data.  
Block length = 10,  $B = 312$

statistic	obs. value	perm. $p$ -value	asympt. $p$ -value
$m_1$	0.531	0.033	0.034
$m_2$	0.738	0.030	0.030
$U^2$	0.207	0.034	0.034

The  $p$ -values, which are virtually identical (whether obtained by the permutation method or by asymptotic approximation), indicate at least one change in the  $\alpha$ -value. Indeed, the cusum plot in Figure 1 suggests a downward

change in block  $b = 127$  followed perhaps by an upward change in block  $b = 173$ . The corresponding estimates of  $\alpha$ , computed by maximum likelihood in the three segments, are  $\hat{\alpha}_1 = 0.88$ ,  $\hat{\alpha}_2 = 0.73$  and  $\hat{\alpha}_3 = 0.90$ , suggesting a square wave configuration of  $\alpha$ -values in these data.



**Figure 1.** Cusum plot of ethernet traffic data. Time index on horizontal axis and cusum value on vertical axis.

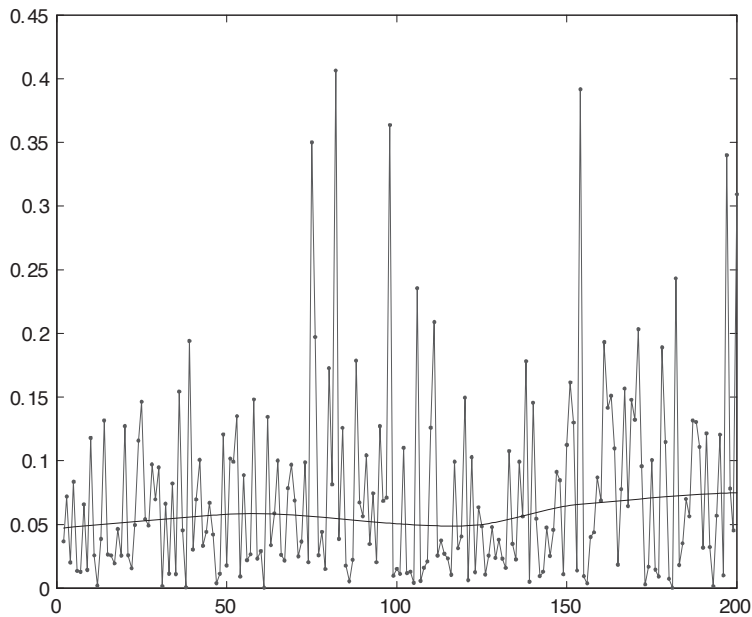
When blocks of length 20 are used, none of the three test statistics produces a significant result. With such a block length the square wave consists of just 23 observations. None of the three test statistics are apparently powerful enough to detect a change of such short duration.

## 4.2 Wind speed data

Haslett and Raftery (1989) analyze a multivariate time series consisting of (the square roots of) daily average wind speeds at 12 meteorological stations in

Ireland. Our interest is in the long memory character of the series of wind speeds at the 12 stations where data are abundant. Here we consider as an example the deseasonalised data from station number 11. The cusum plot exhibits some erratic behaviour towards the end of the series and we therefore restrict attention to the first 4,000 observations. We use blocks of length  $m = 10$ , which yield a series of 400  $\hat{\alpha}$  estimates. The cusum plot of this series of estimates (not shown here) suggests a decrease in the value of  $\alpha$  after block  $b = 237$ .

However, the independence assumption regarding the  $\hat{\alpha}$  series is in some doubt. Figure 2 shows a plot of the periodogram of the 400  $\hat{\alpha}$  values together with a loess estimate of the spectral density function. The decreasing trend toward the lower frequencies is indicative of negative serial correlation. The estimate of the spectral density at the zero frequency,  $\hat{f}(0)$ , is 0.048. The corresponding value of the one-change test statistic  $m_1$ , computed using (5) with  $s$  replaced by  $(\hat{f}(0))^{1/2} = 0.218$ , is 0.583 with a  $p$ -value of 0.033. If the negative autocorrelation is ignored, that is if (5) with  $s = 0.270$  is used, we find  $m_1 = 0.352$  with a  $p$ -value of 0.095.



**Figure 2.** Periodogram and loess smooth for the wind speed data. Frequency (number of cycles in the data set) is on the horizontal axis and spectral power on the vertical axis.

Haslett and Raftery (1989) modeled these data as a FARIMA (fractional ARMA) time series, not as FGN as we did above. We repeated the analysis using 200 blocks of length 20 each and maximum likelihood estimates obtained under a FARIMA model as well as under a FGN model. Use of a larger block size is necessitated by the fact that the FARIMA estimation algorithm uses the periodogram of the data. A block size of 10 yields only 5 periodogram values, which is not sufficient to produce reliable estimates of  $\alpha$  from such blocks. On the other hand, with the larger block size there is no evidence of serial correlation between the block estimates obtained under either of the models. Thus, in the test statistics we use  $T_b$  from (5) without any adjustment for autocorrelation. The results are shown in Table 4.

**Table 4** Test for change in  $\alpha$ : Wind speed data.  
Block length = 20,  $B = 200$ .

statistic	obs. value	perm. $p$ -value	asympt. $p$ -value
$m_1$ FARIMA	0.387	0.081	0.078
$m_1$ DFGN	0.443	0.056	0.056

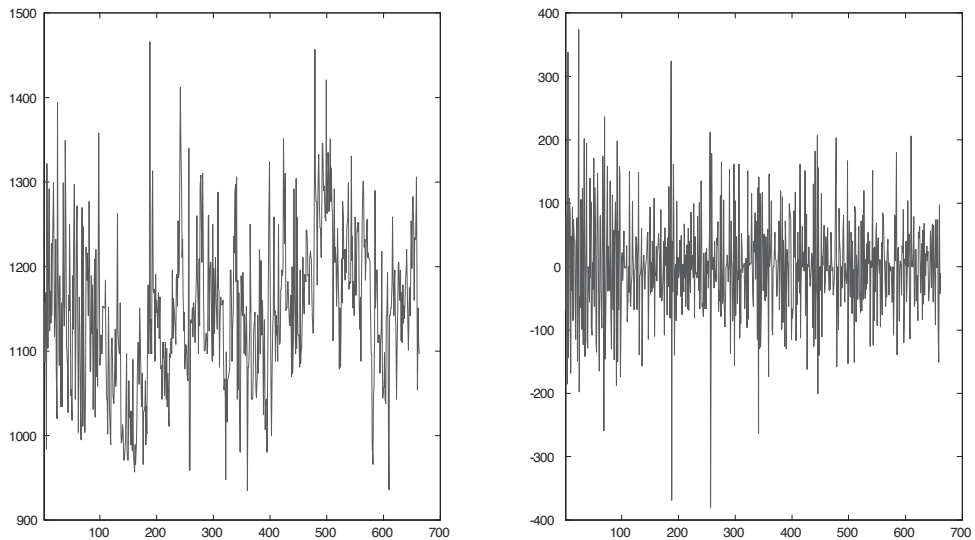
As expected, the larger block size dilutes the significance of the DFGN result - the  $p$ -value has increased from 0.033 to 0.059. The result under a FARIMA model is less significant. Thus, if a sample size of  $B = 200$  is regarded as "large", the statistic  $m_1$  does not provide convincing evidence of non-constancy.

### 4.3 Nile river inflow data

The annual flows into the Nile river are possibly the best known set of data in the context of long memory time series; see Beran (1994). From the plot of the data in the left panel of Figure 3 it would seem that a number of mean changes have occurred. However, others (e.g. Hurst, 1951; Beran, 1994) have demonstrated that the seeming non-stationarity is most likely a manifestation of long memory in the data. These non-stationarities are not visible in the rightmost panel in Figure 3, which is a plot of the first differences,  $Y$ , of the data in the left panel. It is rather clear from the latter plot, though, that the variability among the first 100 or so observations is greater than that among the remaining observations. The apparent change in variability could possibly be explained by a change in the value of  $\alpha$ . For instance, Beran (1994, page 206) reports  $\hat{\alpha} = 0.54$  for the first 100 observations and  $\hat{\alpha} = 0.88$  thereafter, which would imply, see (2), that for  $t \leq 100$  and  $s > 100$

$$\text{var}(Y_t)/\text{var}(Y_s) \approx (1 - 2^{-0.92})/(1 - 2^{-0.24}) = 3.07.$$

Beran (1994, page 208) found strong evidence of at least one changepoint upon using six blocks of length 100, but he points out that the  $p$ -value may be suspect because the blocks were chosen with the increased variability among the first 100 observations in mind.



**Figure 3.** Time series plots of Nile river inflows (vertical axis, left panel) and their first differences (right panel) for 663 consecutive years.

The following, independent, analysis assumes merely that a change, if present, occurs early or late in the series rather than towards the middle. Then a slightly more powerful version of the statistic  $m_1$  is obtained by using instead of the cusum (5) the weighted cusum

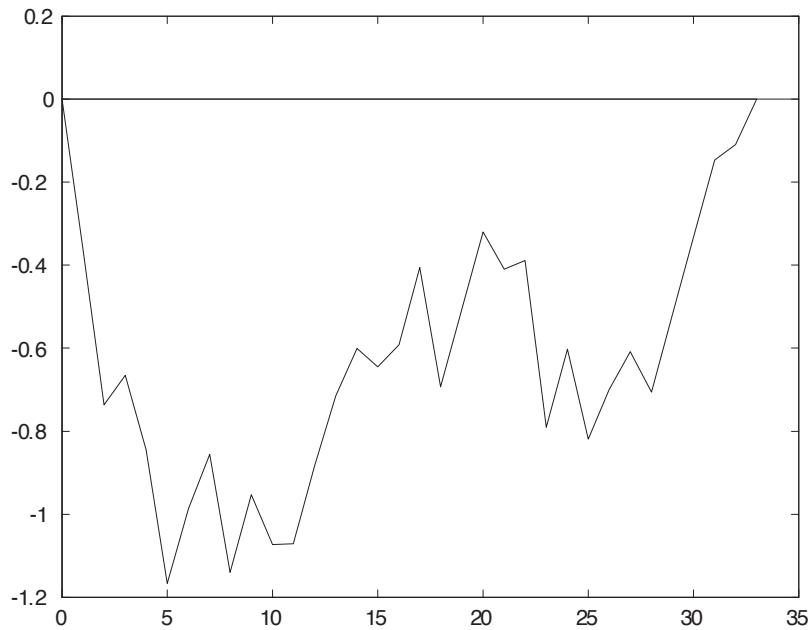
$$T_b^* = T_b / \{w_b(1 - w_b)\}^{1/2}, \quad b = 1, 2, \dots, B - 1$$

where  $w_b = b/B$ . The corresponding weighted version of  $m_1$  in (6) is

$$m_1^* = (B - 1)^{-1} \sum_{b=1}^{B-1} (T_b^*)^2,$$

the large sample distribution of which is that of the Anderson-Darling (1954, page 768) goodness of fit criterion. We used blocks of length  $m = 10$  to obtain a series of 66 MLE's of  $\alpha$  and find  $m_1^* = 1.59$  with a  $p$ -value of 0.15 (computed from formula (8) of Anderson and Darling, 1954). However, the  $\alpha$  estimates obtained from the 66 blocks are highly variable, which gives rise to a rather uninformative cusum plot. Nonetheless, the series of 66 estimates do not exhibit any serial correlation.

More stable estimates can be expected to result if a larger block size were used. In fact, with a block size of 20, we find  $m_1^* = 2.67$  with a permutation  $p$ -value of 0.034. While the corresponding cusum plot in Figure 4 does not show a well-defined minimum, it seems clear that the change in  $\alpha$  has most likely taken place within the first 10 blocks, i.e. within the first 200 observations.



**Figure 4.** Cusum plot of 33  $\hat{\alpha}$  estimates from consecutive disjoint blocks of 20 observations - Nile river inflows data.

## References

- Anderson, T.W. and D.A. Darling. (1954). A test of goodness of fit. *Journal of the American Statistical Association*, **49**, 765-769.
- Beran, J. (1994). *Statistics for long-memory processes*. Chapman and Hall, New York.
- Beran, J. and Terrin, N. (1994). Estimation of the long-memory parameter, based on a multivariate central limit theorem. *Journal of Time Series Analysis*, **15**, 269-278.

- Beran, J. and Terrin, N.(1996). Testing for a change of the long-memory parameter. *Biometrika*, **83**, 627-638.
- Dalhaus, R. (1989). Efficient parameter estimation for self-similar processes, *The Annals of Statistics*, **17**, 1749-1766.
- Granger, C.W.J. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P500 absolute stock returns. *Journal of Empirical Finance*, **11**, 399-421.
- Granger, C.W.J. and Joyeux, R. (1980). An introduction to long memory time series and fractional differencing. *Journal of Time Series Analysis*, **1**, 1-15.
- Haslett, J. and Raftery, A. (1989). Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power. *Applied Statistics*, **38**, 1-50.
- Horváth, L. and Shao, Q-M. (1999). Limit Theorems for Quadratic Forms with Applications to Whittle's Estimate. *The Annals of Applied Probability*, **9**, 146-187.
- Hurst, H. E., (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **116**, 770-99.
- Leland, W.E., Taqqu, M.S., Willinger, W. and Wilson, D.V. (1994). On the Self-Similar Nature of Ethernet Traffic (Extended Version). *IEEE/ACM Transactions on Networking*, **2**, 1-15.
- Lombard, F. (1987). Rank tests for changepoint problems. *Biometrika*, **74**, 615-624.
- Lombard, F. and Hart, J.D. (1994). The analysis of change-point data with dependent errors. In: *Change-point Problems, IMS Lecture Notes-Monograph Series* (Volume 23, 1994), Eds. E. Carlstein, H-G Müller and D. Siegmund. 194-209.
- Willinger, W., Taqqu, M.S., Leland, W.E. and Wilson, D.V., (1995). Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements. *Statistical Science*, **10**, 67-85.

## Appendix

### Asymptotic independence of $\hat{\alpha}_1, \dots, \hat{\alpha}_B$

We set  $\sigma = 1$  without loss of generality and denote by  $\alpha_j^{(0)}$  the true value of  $\alpha_j$ . Since the DFGN series has short memory, it has an infinite moving average representation

$$Y_t = \sum_{k=0}^{\infty} a_{t-k} \zeta_k$$

with  $\sum_{k=0}^{\infty} a_k^2 < \infty$  and a sequence  $\{\zeta_k\}$  of i.i.d. standard normal random variables. This fact places DFGN within the ambit of Theorem 1 of Beran and Terrin (1994). Since  $\partial\ell/\partial\hat{\alpha}_j = 0$  we obtain by Taylor expansion that

$$-\partial\ell/\partial\alpha_j^{(0)} = (\hat{\alpha}_j - \alpha_j^{(0)})\partial^2\ell/\partial\alpha_j^{(*)2} \quad (10)$$

with  $0 \leq |\hat{\alpha}_j - \alpha_j^{(*)}| < |\hat{\alpha}_j - \alpha_j^{(0)}|$ . Upon differentiating (4) with respect to  $\alpha_j$  we find furthermore that

$$\begin{aligned} \partial\ell/\partial\alpha_j &= \frac{1}{2}\mathbf{Y}^\top \mathbf{A}(\alpha_j)\mathbf{Y} - \frac{1}{2}\text{tr}(\boldsymbol{\Sigma}(\alpha_j)^{-1}\boldsymbol{\Sigma}(\alpha_j)') \\ &= \frac{1}{2}\mathbf{Y}^\top \mathbf{A}(\alpha_j)\mathbf{Y} - \mathbf{E}_{\alpha_j} \left[ \frac{1}{2}\mathbf{Y}^\top \mathbf{A}(\alpha_j)\mathbf{Y} \right] \end{aligned} \quad (11)$$

where  $\mathbf{A}(\alpha) = \boldsymbol{\Sigma}(\alpha)^{-1}\boldsymbol{\Sigma}(\alpha)'\boldsymbol{\Sigma}(\alpha)^{-1}$  and where the prime indicates differentiation with respect to  $\alpha$ . Notice that the first term on the right hand side of (11) is a quadratic form of the type considered in Beran and Terrin (1994), denoted by  $Q_{N,j}$  there. It is straightforward to show that the condition (4) required in Theorem 1 of Beran and Terrin (1994, page 271) is satisfied. Thus, it follows from the latter Theorem together with (10) that

$$m^{-1/2}(\hat{\alpha}_j - \alpha_j^{(0)})\partial^2\ell/\partial\alpha_j^{(*)2}, \quad j = 1, \dots, B$$

are for fixed  $B$  asymptotically independent and normally distributed as  $m \rightarrow \infty$ . Since  $m^{-1}\partial^2\ell/\partial\alpha_j^{(*)2}$  converges in probability to  $-E[\partial\ell/\partial\alpha_j^{(0)}]^2$  as

$m \rightarrow \infty$  - see Dalhaus (1987; (v) in the proof of his Theorem 3.2) - it follows that  $m^{1/2}(\hat{\alpha}_j - \alpha_j^{(0)})$ ,  $j = 1, \dots, B$  are asymptotically independent and normally distributed as  $m \rightarrow \infty$ .

We also ran some Monte Carlo simulations to check the appropriateness of the independence assumption. We generated 1,000 DFGN series of length  $n = 1200$  (600) each with Hurst parameters varying between  $\alpha = 0.6$  and  $\alpha = 0.9$ . Each series was divided into  $B = 120$  (60) contiguous blocks of length 10 and the first 5 lag autocorrelations of the resulting time series of estimates  $\hat{\alpha}_1, \dots, \hat{\alpha}_{120}$  were calculated. We did the same for blocks of length 20. The percentage of estimated autocorrelations falling outside the 95.6% ( $\pm 2/B^{1/2} = \pm 0.183$ ) limits are reported in Table 5. With only two exceptions, the observed proportions are in line with the null hypothesis.

**Table 5** Monte Carlo simulation: percentage of autocorrelations outside usual confidence limits.

N=600, Blocklength = 10, Blocks = 60					N=1200, Blocklength = 10, Blocks = 120				
Lag	H=0.60	H=0.70	H=0.80	H=0.90	Lag	H=0.60	H=0.70	H=0.80	H=0.90
1	3.0	2.9	3.6	3.6	1	4.2	4.5	4.6	6.0
2	3.4	3.6	3.6	4.5	2	4.5	4.8	3.4	3.5
3	3.6	2.8	4.6	3.5	3	3.2	3.5	4.8	3.0
4	3.4	3.7	2.6	2.7	4	5.1	3.2	3.2	3.8
5	3.4	3.5	3.7	2.2	5	4.0	3.1	4.6	3.9
N=600, Blocklength = 20, Blocks = 30					N=1200, Blocklength = 20, Blocks = 60				
Lag	H=0.60	H=0.70	H=0.80	H=0.90	Lag	H=0.60	H=0.70	H=0.80	H=0.90
1	2.2	1.8	3.6	2.1	1	4.1	4.3	3.3	3.1
2	2.6	3.9	2.4	2.6	2	4.1	3.5	3.8	3.1
3	2.8	2.2	2.5	2.3	3	3.3	3.3	3.0	3.1
4	2.0	1.6	2.5	1.7	4	3.6	3.3	2.1	2.0
5	1.4	2.2	1.9	1.6	5	3.3	3.8	3.2	2.4