Original article

Not all estimators are born equal: The empirical properties of some estimators of long memory

William Rea [a,b,*], Les Oxley [b,c], Marco Reale [a], Jennifer Brown [a]

[a] *Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand*
[b] *Department of Economics and Finance, University of Canterbury, Christchurch, New Zealand*
[c] *Department of Economics, Waikato University, Hamilton, New Zealand*

Abstract

We present the results of a simulation study into the properties of 12 different estimators of the Hurst parameter, H, or the fractional integration parameter, d, in long memory time series which are available in R packages. We compare and contrast their performance on simulated Fractional Gaussian Noises and fractionally integrated series with lengths between 100 and 10,000 data points and H values between 0.55 and 0.90 or d values between 0.05 and 0.40. We apply all 12 estimators to the Campito Mountain data and estimate the accuracy of their estimates using the Beran goodness-of-fit test for long memory time series.
© 2012 IMACS. Published by Elsevier B.V. All rights reserved.

1. Introduction

The subject of long-memory time series was brought to prominence by Hurst [35] and has subsequently received extensive attention in the literature. See the volumes by Beran [10], Embrechts and Maejima [20], and Palma [45] and the collections of Doukhan et al. [19] and Robinson [49] and the references therein.

There is considerable debate in the literature on the origins of the so-called Hurst phenomenon, see for example Klemes [38] or Diebold and Inoue [18]. It is not the purpose of this paper to address this unresolved issue. Instead, we focus on the empirical properties of several estimators freely available in R [47]. There are some estimators which are known to have excellent finite sample properties such as the local Whittle of Shimotsu and Phillips [51] and the maximum likelihood estimator of Sowell [53] which do not appear to be currently implemented in R and hence are not covered in this paper.

Of critical importance in analyzing and modeling long memory time series is estimating the strength of the long-range dependence. Two measures are commonly used. The parameter H, known as the Hurst or self-similarity parameter,

---

* Corresponding author at: Department of Economics and Finance, University of Canterbury, Private Bag 4800, Ilam, Christchurch 8140, New Zealand. Tel.: +64 3 364 3474; fax: +64 3 364 2635.

E-mail addresses: bill.rea@canterbury.ac.nz (W. Rea), loxley@waikato.ac.nz (L. Oxley), marco.reale@canterbury.ac.nz (M. Reale), jennifer.brown@canterbury.ac.nz (J. Brown).

was introduced to applied statistics by Mandelbrot and van Ness [42] and arises naturally from the study of self-similar processes. The other measure, the fractional integration parameter, d, arises from the generalization of the Box-Jenkins ARIMA(p, d, q) models from integer to non-integer values of the integration parameter d. This generalization was accomplished independently by Granger and Joyeux [28] and by Hosking [34]. The fractional integration parameter d is also the discrete time counterpart to the self-similarity parameter H and the two are related by the simple formula $H = d + 1/2$. These two models are explained in more detail in Section 2.

A number of estimators of H and d have been developed. These are usually validated by an appeal to some aspect of self-similarity, or by an asymptotic analysis of the distributional properties of the estimator as the length of the time series converges to infinity.

A number of theoretical results on the asymptotic properties of various estimators have been obtained. The aggregated variance method was shown to be asymptotically biased of the order $1/\log N$, where N is the number of observations, by Giraitis et al. [26] who also showed the GPH [23] estimator was asymptotically normal and unbiased. Agiakloglou et al. [3] examined the bias of the GPH estimator when the estimator was applied to time series generated by ARMA processes. Robinson [48] proved the averaged periodogram estimator was consistent under very mild conditions. Lobato and Robinson [39] obtained its limiting distribution. The Peng estimator [46] was proved to be asymptotically unbiased by Taqqu et al. [55]. Some theoretical properties of the R/S estimator have been examined by Mandelbrot [40] and Mandelbrot and Taqqu [41]. Mandelbrot [40] proved that the R/S statistic is robust to the increment process having a long-tailed distribution in the sense that $E[X_i^2] = \infty$. However, Bhattacharya et al. [12] proved that the R/S statistic was not robust to departures from stationarity. Thus for a short memory process with slowly decaying deterministic trend the R/S statistic will report an estimate of H which implies the presence of long-memory. An estimator based on wavelets was proved asymptotically unbiased and efficient by Abry et al. [2]. They also showed the traditional variance type estimators were fundamentally flawed and could not lead to good estimators of H. Nevertheless, several variance type estimators are implemented in R hence are freely available. Fox and Taqqu [21] proved the Whittle estimator was consistent and asymptotically normal for Gaussian long range dependent sequences. Dahlhaus [16] proved the Whittle estimator of Fox and Taqqu [21] was efficient. Further theoretical results on the Whittle estimator can be found in Horvath and Shao [33].

Because the finite sample properties of these estimators can be quite different from their asymptotic properties some previous authors have undertaken empirical comparisons of estimators of H and d. Nine estimators were discussed in some detail by Taqqu et al. [55] who carried out an empirical study of these estimators for a single series length of 10,000 data points, five values of both H and d, and 50 replications. Teverovsky and Taqqu [56] showed in a simulation study that the differenced variance estimator was unbiased for five values of H (0.5, 0.6, 0.7, 0.8, and 0.9) for series with 10,000 observations whereas the aggregated variance estimator was downwards biased. Jensen [36] undertook a comparison of two estimators based on wavelets, one proposed by Jensen [36] and the other proposed by McCoy and Walden [43], with the GPH estimator for four series lengths (2^7, 2^8, 2^9, and 2^{10} observations), five values of d and 1000 replications. They reported the wavelet estimators had lower mean squared errors (MSEs) than the GPH estimator for all d values and series lengths investigated. Jeong et al. [37] carried out a comparison of six estimators on simulated fractional Gaussian noises (FGNs) with 32,768 (2^{15}) observations, five values of H and 100 replications. Smith et al. [52] studied the properties of three estimators under model misspecification.

Several of the above empirical investigations would have been limited by the then available computer power which has since increased considerably. We have extended these studies to a larger number of parameters, higher number of replications and 12 estimators as detailed in Section 3.

The remainder of the paper is organized as follows. Section 2 gives a brief outline of Fractional Gaussian Noises and Fractionally Integrated Series. Section 3 gives details of the method. Section 4 presents the results. Section 5 applies the methods to the Campito Mountain data which is regarded as a standard example of a long memory time series. Section 6 contains the discussion and Section 7 gives our conclusions and suggests avenues of future research.

2. Models

A number of models have been proposed to account for the extraordinary persistence of the correlations across time found in long memory series. The most widely used models for long memory series are the Fractional Gaussian Noises (FGNs), a continuous time process, and their discrete time counterparts, the fractionally integrated series of order d. FGNs were introduced into applied statistics by Mandelbrot and van Ness [42] in an attempt to model the

exceptionally slow decay in the autocorrelations observed in long memory series. FGNs are the stationary increments of an Gaussian H self-similar stochastic process which we now define.

Definition 1. A real-valued stochastic process $\{Z(t)\}_{t \in \mathcal{R}}$ is self-similar with index $H > 0$ if, for any $a > 0$,

$$\{Z(at)\}_{t \in \mathcal{R}} =_d \{a^H Z(t)\}_{t \in \mathcal{R}}$$

where $\mathcal{R}$ denotes the real numbers and $=_d$ denotes equality of the finite dimensional distributions. H is also known as the Hurst parameter.

Definition 2. A real-valued process $Z = \{Z(t)\}_{t \in \mathcal{R}}$ has stationary increments if, for all $h \in \mathcal{R}$

$$\{Z(t + h) - Z(h)\}_{t \in \mathcal{R}} =_d \{Z(t) - Z(0)\}_{t \in \mathcal{R}}.$$

The stationary increments of an H-self-similar series with $0.5 < H < 1$ exhibit long memory. It can be shown [see 10, Section 2.3] that the autocorrelations of the increments of an H self-similar process are of the form

$$\rho(k) = \frac{1}{2}[(k + 1)^{2H} - 2k^{2H} + (k - 1)^{2H}]$$

where $k \geq 0$ and $\rho(k) = \rho(-k)$ for $k < 0$. Asymptotically

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

Thus for $1/2 < H < 1$ the autocorrelations decay so slowly that they are not summable, i.e.

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

Independently Granger and Joyeux [28] and Hosking [34] obtained the discrete time counter-parts to FGNs, the Fractionally Integrated processes of order d (FI(d)). These are also a generalization of the "integration" part of the Box-Jenkins ARIMA(p, d, q) (AutoRegressive Integrated Moving Average) models to non-integer values of the integration parameter d, and are often referred to as ARFIMA models (AutoRegressive Fractionally Integrated Moving Average). They have the added advantage of being able to model any additional short term autocorrelation present in the series not accounted for by the long memory component.

Denoting the backshift operator by B, for a non-integer value of d the operator $(1 - B)^d$ can be expanded as a Maclaurin series into an infinite order AR representation

$$(1 - B)^d X_t = \sum_{k=0}^{\infty} \frac{\Gamma(k - d)}{\Gamma(k + 1)\Gamma(-d)} X_{t-k} = \epsilon_t \tag{1}$$

where $\Gamma(\cdot)$ is the gamma function, and ϵ_t is a innovation term drawn from an $N(0, \sigma^2)$ distribution. The operator in Eq. (1) can also be inverted and X_t written in an infinite order MA representation. It can be shown [see 10, Section 2.5] that the autocorrelations for a fractionally integrated process are, asymptotically,

$$\rho(k) \approx c|k|^{2d-1}$$

where c is a constant which depends on d, and on any AR(p), and MA(q) components used in the model. For $0 < d < 1/2$ the autocorrelations decay so slowly that they are not summable. It is this slow rate of decay of the autocorrelations that makes them attractive as models for long memory time series.

As indicated above, the two parameters H and d are related by the simple formula $d = H - 1/2$.

Both FGNs and FI(d)s have been extensively applied as models for long memory time series. A partial list of some areas of application are atmospheric pressures [54], atmospheric temperatures [13], economics [4], exchange rates [8], interest rates [7], inflation [5], ocean surface temperatures [44], political science [14], river flows [35], stock markets [24], and wind flow [31].

2.1. Constrained non-stationary models

Klemes [38] argued that long memory in hydrological time series was a statistical artifact caused by analyzing non-stationarity time series with statistical tools which assume stationarity. Often series which display the long memory property are constrained for physical reasons to lie in a bounded range, but beyond that we have no reason to believe that they are stationary. In the series we study in Section 5, namely tree ring data, there are upper and lower physical limits to a tree ring's growth in any given year.

Models of this type which have been proposed typically have stochastic shifts in the mean, but overall are mean reverting about some long term average.

We define the break model as follows:

$$y_t = \sum_{i=1}^{p} I_{t_{i-1} < t \le t_i} \mu_i + \epsilon_t \tag{2}$$

where y_t is the observation at time t, $I_{t \in S}$ is an indicator variable which is 1 only if $t \in S$ and 0 otherwise, t is the time, t_i, $i = 1, \ldots, p$, are the breakpoints and μ_i is the mean of the regime i, and ϵ_t is a noise term. In this case, a regime is defined as the period between breakpoints.

It is important to note that Eq. (2) is just a way to represent a sequence of different models (i.e. models subjected to structural breaks). This model only deals with breaks in mean. It can be generalized for any kind of break. In series with a structural break the noise process, ϵ_t, may also undergo a change. This class of model has been used by others when studying long memory processes. However, given a true break each regime must be modeled separately.

2.2. Aggregation models

In economic and financial applications many time series are aggregates of numerous individual component series. For example, stock market indices, such as the S&P 500, are just a (possibly weighted) sum of a few tens to thousands of individual stocks. Granger and Morris [29] examined the types of series which resulted from summing a number of AR(p) and MA(q) series. Granger and Morris [29] considered the AR and MA models to be "...intuitively reasonable and could well occur in practice ...". Of particular interest, they reported that the sum of an AR(p_1) and an AR(p_2) series yielded an ARMA($p_1 + p_2$, max(p_1, p_2)) series.

Subsequently Granger [27] considered whether the sum

$$X_t = \sum_{j=1}^{N} X_t(j)$$

where $X_t(j)$; $j = 1, 2, \ldots, N$ are the individual components of an aggregated series X_t, each of which has short-memory, could exhibit long memory. Granger [27] reported this was the case. Beran [10, p. 16] stated "...observing long-range dependence in an aggregated series (macro-level) one cannot conclude that this long memory is due to the genuine occurrence of long memory in the individual series (micro-level). It is possible that instead it is induced artificially by aggregation."

Aggregation models have also appeared in the non-economic literature, albeit indirectly. Hare and Mantua [30] amassed 100 North Pacific climatic and biological time series and used a composite analysis to demonstrate regime shifts in the Pacific Decadal Oscillation (PDO). Rudnick and Davis [50] responded with a paper criticizing their methodology and demonstrating that composite analysis of red noise series produced the step-like changes seen in the PDO. Overland et al. [44] produced a third paper which suggested a long-memory model also fitted the data. It would appear all three sets of authors were discussing, in part, a Granger-style long memory aggregation model without being aware of his work.

Cioczek-Georges and Mandelbrot [15] have studied an alternative aggregation model which exhibits long memory properties. Their model has found application in modeling aggregate traffic on computer networks. Willinger et al. [58] also studied long memory in computer networks from the point of view of the network protocols which govern how traffic from individual sources is aggregated onto shared data networks.
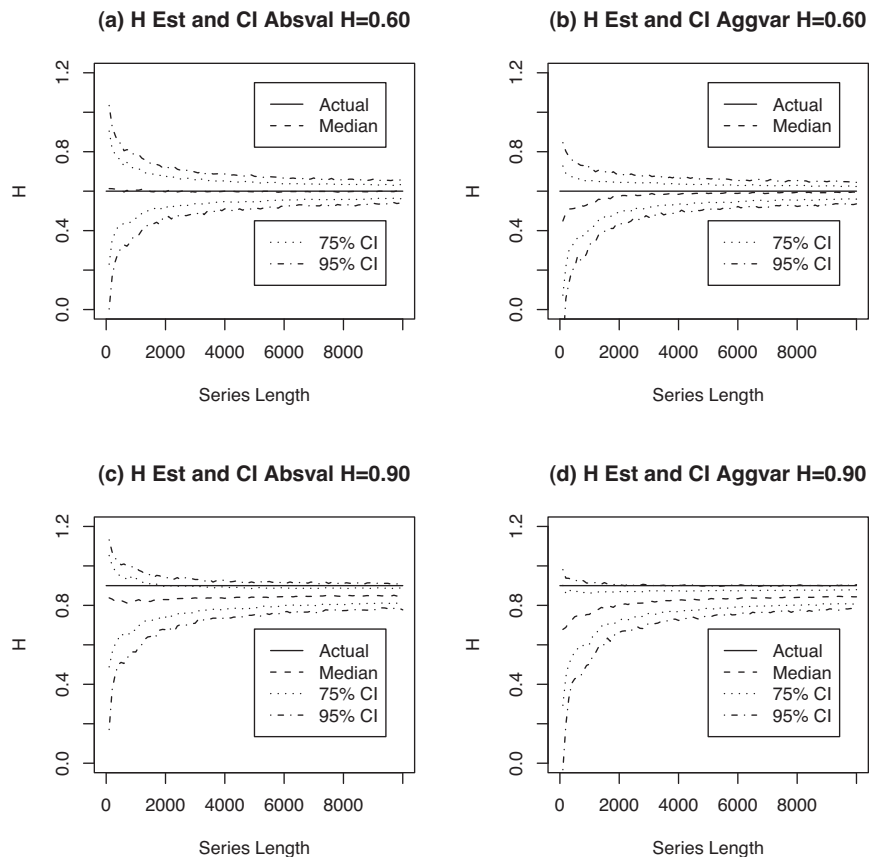
Fig. 1. Empirical confidence intervals for the *H* estimates with *H* = 0.60 and *H* = 0.90; (a) and (c) absolute value method, (b) and (d) aggregated variance estimator.

3. Method

Because the R [47] software and packages are freely available under the GNU Public License, can be downloaded from the web, and runs on a wide variety of computing platforms, we chose *H* and *d* estimators implemented in R for testing.

Ten *H* estimators are implemented in the contributed package fSeries of Wuertz [59]. They are the absolute value, aggregated variance, boxed periodogram, differenced variance, Higuchi, Peng, periodogram, rescaled range, wavelet, and the Whittle. The wavelet estimator is discussed in some detail by Abry and Veitch [1], Abry et al. [2], and Veitch and Abry [57] and the other nine are discussed by Taqqu et al. [55]. Further, the GPH [23] and Haslett–Raftery [31] estimators are implemented as estimators for *d* in the contributed package fracdiff of Fraley et al. [22].

Taqqu et al. [55] simulated FGNs and the corresponding discrete time fractionally integrated (FI(*d*)) series and found that each estimator performed similarly whether estimating *H* in simulated FGNs or *d* in simulated FI(*d*)s. For example, if an estimator was biased when estimating *H* it was also biased in a very similar manner when estimating *d*. Thus, with the exception of the GPH and Haslett–Raftery estimators, we only investigated each estimator's performance in estimating *H* for simulated FGNs. FGNs were generated using the function fgnSim in fSeries. We ran 1000 replications of simulated FGNs with 100 different lengths and eight different *H* values. The lengths were between 100 and 10,000 data points in steps of 100. The *H* values were between 0.55 and 0.90 in steps of 0.05. For each series *H* was estimated by each of these ten estimators. For each *H* value and series length we estimated the median, 75% and 95% confidence intervals empirically from the simulated data. The *H* or *d* estimates were sorted into ascending order and the median obtained by averaging the 500th and 501st values. Similar calculations were done for the upper and lower values of the 75% and 95% confidence intervals.
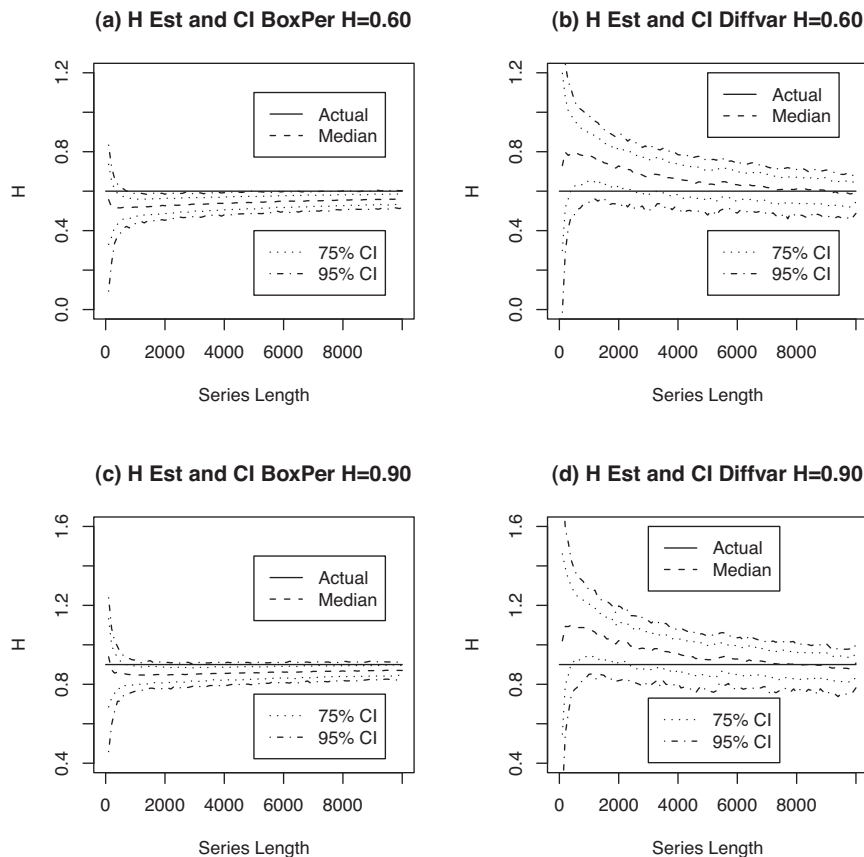
Fig. 2. Empirical confidence intervals for the *H* estimates with *H* = 0.60 and *H* = 0.90; (a) and (c) boxed periodogram method, (b) and (d) differenced variance estimator.

For the GPH and Haslett–Raftery estimators we generated FI(*d*) series with the function `farimaSim` in `fSeries` over the range 0.05 to 0.40 in steps of 0.05. The other details are the same as above. In the presentation of the results we converted the GPH and Haslett–Raftery *d* estimates to *H* equivalents to facilitate comparisons among the estimators.

The simulations and estimations were performed on a SunBlade 1000 with a 750 Mhz UltraSPARC-III CPU with 2 Gb of memory and a Sun Ultra 10 with a 440 Mhz UltraSPARC-IIi CPU and 1 Gb of memory.

4. Results

To present the results in tabular form would require a very large amount of space. Thus we present a selection of results in graphical form in Figs. 1–7. Figs. 1–6 are presented with the vertical axis with a range of 1.2 *H* units to facilitate comparisons among the estimators' standard deviation of their estimates. It should be noted that stationary long memory occurs in the parameter range $0.5 < H < 1.0$. Baillie [4] states that for $1.0 \le H < 1.5$ the series are non-stationary but mean reverting while for $0 \le H \le 0.5$ the series are anti-persistent. Fig. 7 presents the mean squared error (MSE) as a function of series length. We report MSE for series lengths greater than or equal to 500 data points. Again the vertical axes all have the same range to facilitate comparisons.

The results for the absolute value of the variance method are presented in Fig. 1(a) and (c). The absolute value of the variance method was unbiased at all series lengths when *H* was low (0.55 or 0.60) but became progressively biased and underestimated *H* as *H* increased.

The results for the aggregated variance method are presented in Fig. 1(b) and (d). The aggregated variance method exhibited bias and underestimated *H* in short series when *H* was low. As *H* increased the estimator became increasingly
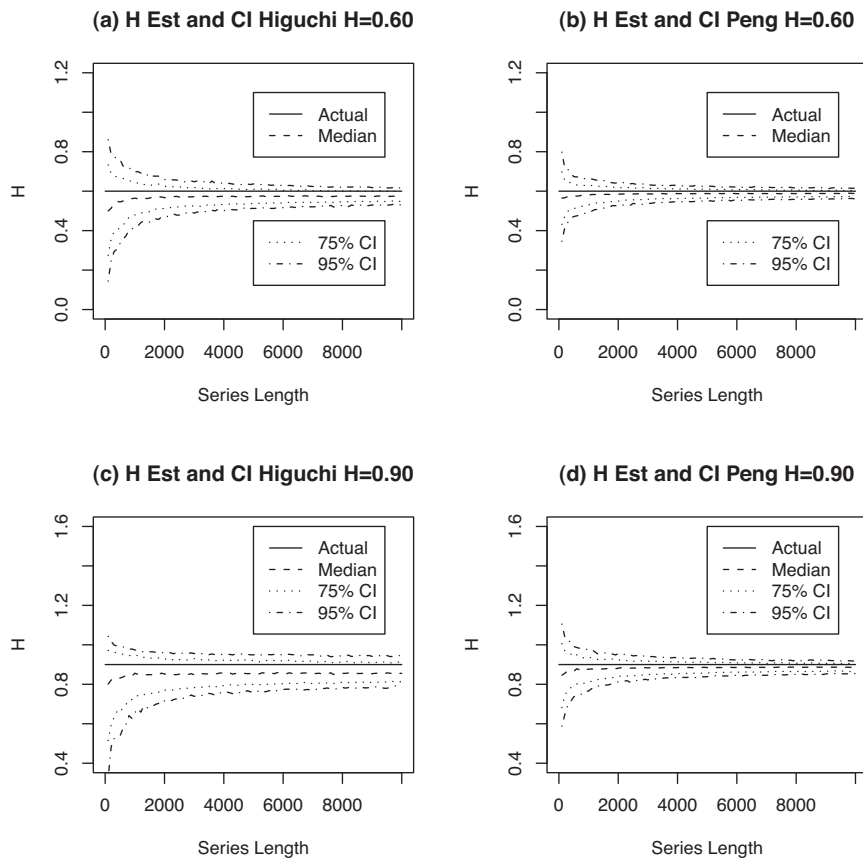
Fig. 3. Empirical confidence intervals for the *H* estimates with *H* = 0.60 and *H* = 0.90; (a) and (c) Higuchi estimator, (b) and (d) Peng estimator.

biased at all series lengths examined. With *H* = 0.90 the true value of *H* lay above the upper 95% empirical confidence interval for all but the shortest series lengths.

The results for the boxed periodogram method are presented in Fig. 2(a) and (c). The boxed periodogram method was developed specifically to deal with perceived problems with the periodogram estimator. Comparing the boxed periodogram with the unmodified periodogram method in Fig. 4(a) and (c) we can see that for FGNs where the series were short and *H* was high that the periodogram method was biased toward over estimating *H*. The boxed periodogram was biased toward underestimating *H* for almost all values of *H* and series lengths examined.

The results for the differenced variance method are presented in Fig. 2(b) and (d). The differenced variance method had one of largest confidence intervals of the estimators when the series were short but this slowly decreased as sample size increased. Only the GPH, periodogram and wavelet methods had a similarly wide confidence interval for short series. The differenced variance estimator exhibited bias toward over estimating *H* for any series with less than 7000 observations. The bias was very serious in the short series. For series longer than about 9000 observations the estimator exhibited a small amount of bias toward underestimating *H*.

The results for the Higuchi estimator [32] are presented in Fig. 3(a) and (c). The Higuchi was biased toward underestimating *H* but the magnitude of the bias appeared relatively independent of *H*. The width of the confidence interval of the estimate increased with increasing *H*.

The results for the Peng estimator [46] are presented in Fig. 3(b) and (d). The Peng estimator was biased toward under estimating *H* in the series lengths we investigated. This bias appeared to be independent of *H* but was very small though it appeared greater in short series.

The results for the periodogram estimator, presented in Fig. 4(a) and (c), were discussed above in conjunction with the boxed periodogram estimator.
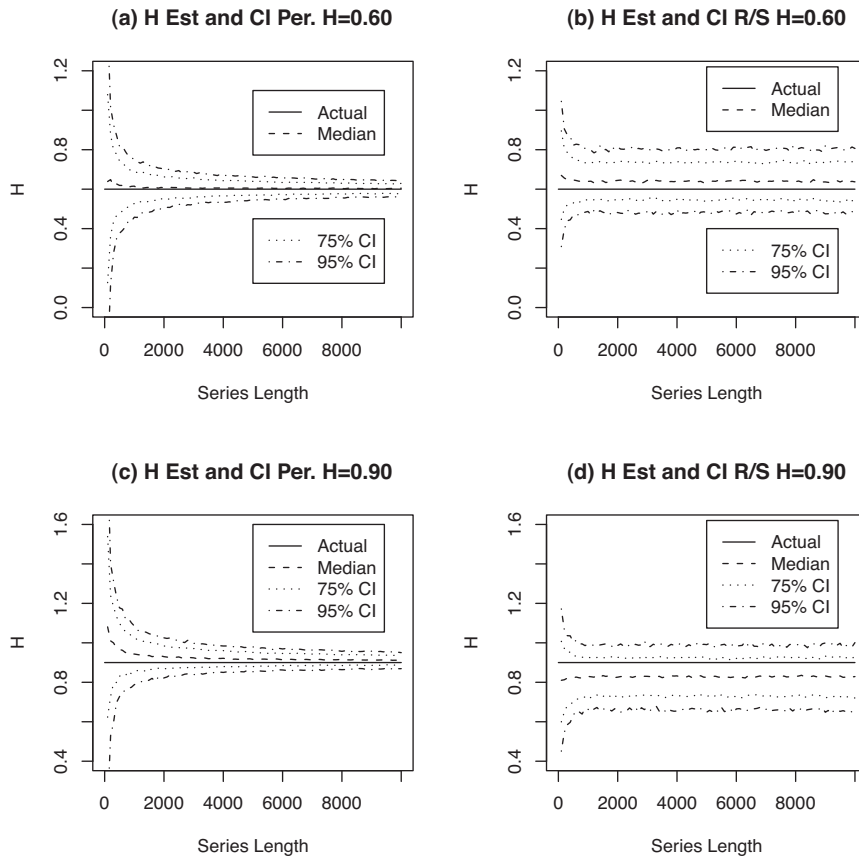
Fig. 4. Empirical confidence intervals for the *H* estimates with $H = 0.60$ and $H = 0.90$; (a) and (c) periodogram estimator, (b) and (d) R/S estimator.

The results for the R/S estimator are presented in Fig. 4(b) and (d). The R/S estimator is of considerable historical interest because it was first proposed by Hurst and was used extensively in early studies of long-memory processes and continues to be used by various authors for empirical work. However, as can be seen from Fig. 4(b) and (d) the R/S estimator exhibited three problems; it was biased upwards when *H* was low, it was biased downwards when *H* was high, and the confidence interval of the estimate did not decrease with increasing series length once the series reached about 1000 observations.

The results for the Whittle estimator are presented in Fig. 5(a) and (c). Compared to the other nine estimators implemented in fSeries the Whittle estimator was remarkable for its narrow confidence interval. It only displayed a small amount of downwards bias when the series were short and *H* was high. There was an implementation issue in the software we used. The Whittle estimator would sometimes terminate with an error when *H* was low and the series contained only a few hundred observations. Thus in Fig. 5(a) there was no data for series with less than 300 observations in the $H = 0.65$ results.

The results for the wavelet estimator are presented in Fig. 5(b) and (d). The wavelet estimator was unbiased for all *H* values at series lengths over 4100 data points. The bias present in series shorter than 4100 data points was very small. The availability of a new *octave* can be seen in Fig. 5(b) and (d) with each doubling of the series length. New *octaves* resulted in a series of steps in the reduction of the confidence interval of the estimate with increasing series length. The estimator had constant variance when the number of *octaves* was fixed.

The results for the GPH estimator are presented in Fig. 6(a) and (c). The GPH estimator exhibited a very small amount of bias toward overestimating *d* at all series lengths examined. It had a very wide confidence interval which narrowed slowly as the series length increased.

The results for the Haslett–Raftery estimator are presented in Fig. 6(b) and (d). The Haslett–Raftery did not report estimates of *d* less than zero ($H < 0.5$). Hence for low *d* and short series the distribution was truncated on the low side
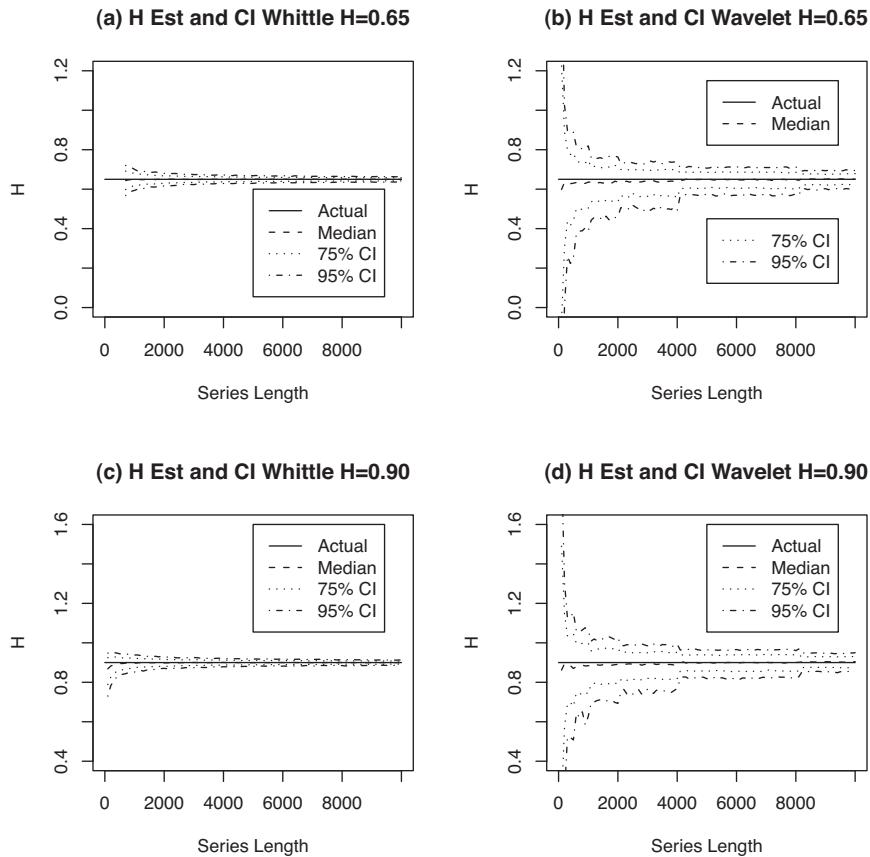
Fig. 5. Empirical confidence intervals for the *H* estimates with *H* = 0.65 and *H* = 0.90; (a) and (c) Whittle estimator, (b) and (d) wavelet estimator.

at *d* = 0 or *H* = 0.5 as in Fig. 6(a). The Haslett–Raftery estimator was an excellent estimator with only small amounts of bias in the short series and had a narrow confidence interval.

Fig. 7 presents the MSEs for the estimators for *H* = 0.9 or *d* = 0.4 as appropriate. This is an alternative way to look at the data from the simulations. We only report MSEs for series of 500 data points and longer because of the high MSEs for some estimators in the short series. The Whittle and Haslett–Raftery both had low MSEs in all series greater than 500 data points in length. The step reductions in the MSE for the wavelet estimator can be clearly seen each time a new *octave* became available.

5. Application: Campito Mountain data

The Campito Mountain bristlecone pine data is regarded as a standard example of a long memory time series. It is a 5405 year series of annual tree ring widths of bristlecone pines on Campito Mountain, California. It was studied by Baillie and Chung [6] who determined that an ARFIMA(0, 0.44, 0) model fitted the data best. The lack of additional short term correlation in the data means it is a good candidate for modeling with an FGN.

The Campito Mountain data is available in the R package `tseries` as the data set `camp`. We applied the 12 estimators to this series and estimated the goodness of fit to an FGN for all estimators, where possible, using the goodness-of-fit test of Beran [9] as implemented in the R package `longmemo` of Beran et al. [11]. The Beran test is more powerful against under estimation of *H* than over estimation. The Beran test was unable to be used for *H* values exceeding unity. It is important to note that Deo and Chen [17] showed that the asymptotic properties of the test as presented by Beran [9] were incorrect. We subjected the Beran test to a simulation study, the results of which will be presented at a later date. This study showed that for sample sizes of the order studied here the Beran test over rejects
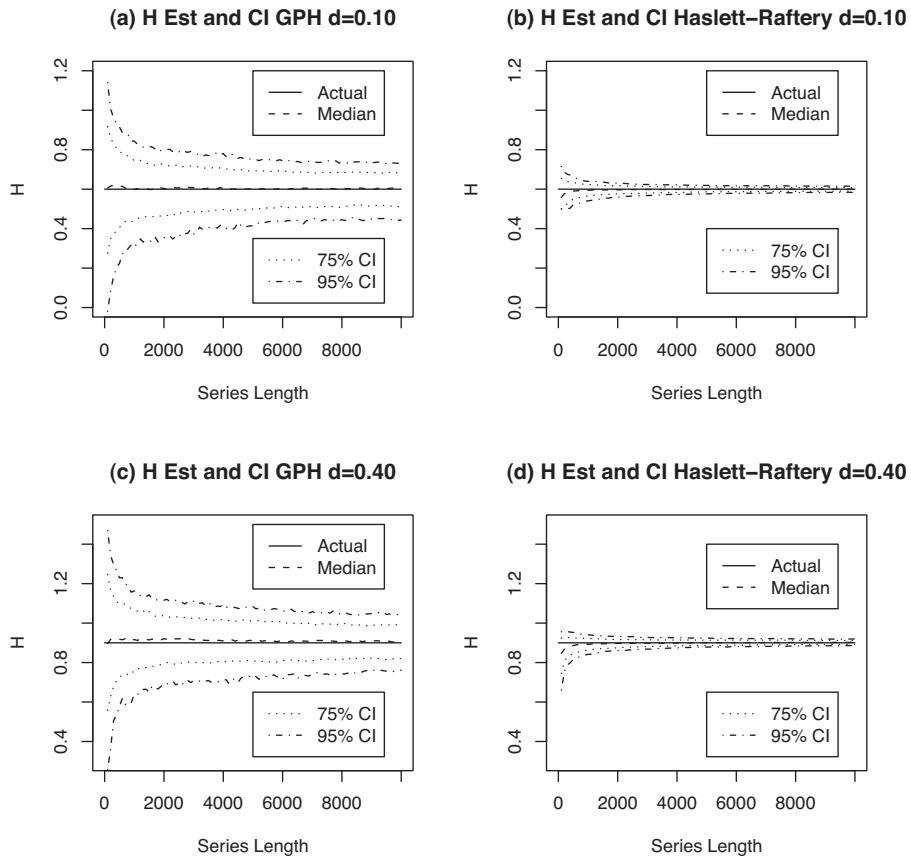
Fig. 6. Empirical confidence intervals for the *H* estimates with $d = 0.10$ ($H = 0.60$) and $d = 0.40$ ($H = 0.90$); (a) and (c) GPH estimator, (b) and (d) Haslett–Raftery estimator.

the null hypothesis by a small amount (e.g. typically 6% at the 5% level). Thus for pragmatic testing of goodness-of-fit, the Beran test can still be used with appropriate caution.

The results are presented in Table 1. The Beran test indicated an *H* value close to 0.89 fitted this data best. The maximum *p*-value was 0.577 for values of *H* estimated by the aggregated variance and rescaled range estimators. Nine of the 12 estimators reported *H* or *d* values which lie in an acceptable range on the basis of the Beran [9] test assuming we set our level of statistical significance at 0.05 to reject the null hypothesis of an FGN. The remaining three could not be tested.

Given the results from our simulated FGNs there were some unexpected *H* estimates for the Campito data. On the basis of the simulations we expected the aggregated variance, absolute value, boxed periodogram, Higuchi, and rescaled range to return a low estimate for *H*. None of these estimators did so. As the Beran test reported that $H = 0.89$ yielded the best fit we used the median value from the simulations with series length 5400 and $H = 0.90$, and adjusted for the difference of 0.01 *H* units, to estimate the value of *H* which would be reported by each estimator if the data was from an FGN. This value is reported in Table 1 as "Expected H". The sixth column reports the empirically determined *p*-value for the actual estimate again using the simulated data. It is interesting that six of the estimators reported *H* estimates which are statistically significantly higher than their expected values.

The estimator which had the least bias and narrowest confidence interval in the simulations with series length 5400 and $H = 0.90$, namely the Whittle, was marginally out performed by the aggregated variance and rescaled range judged on the basis of the Beran test.

The fourth column of Table 1 reports CPU times in seconds. With present day computer speeds estimation times on the Campito series are not an issue. Only four estimators required more than one second of CPU time on this 5405
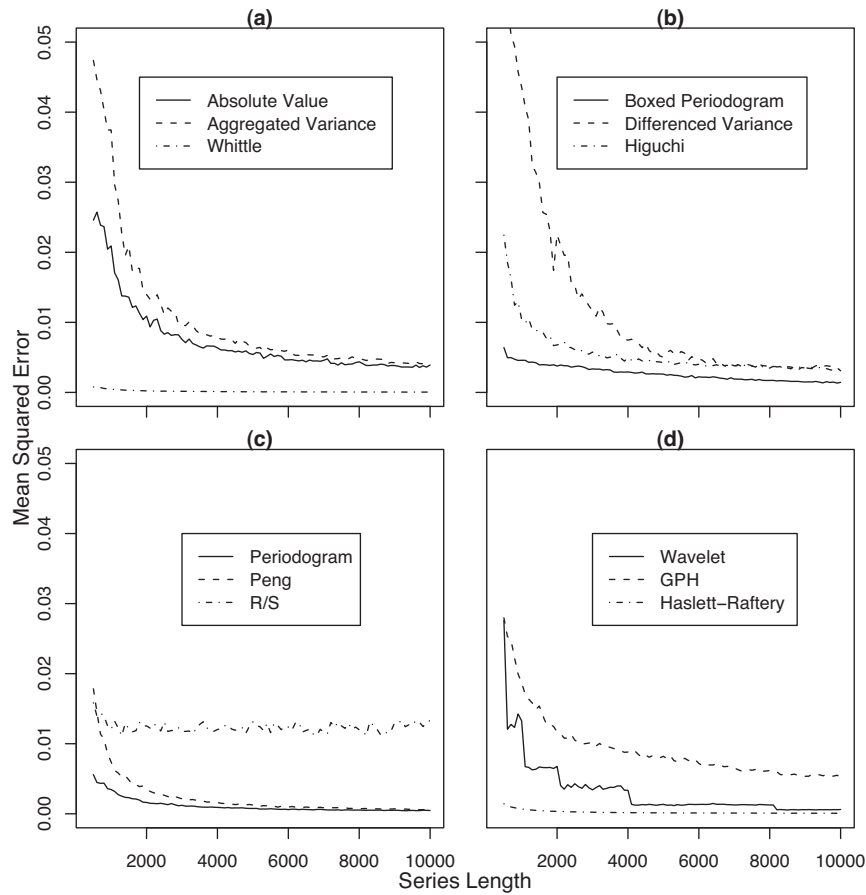
Fig. 7. Mean squared errors (MSE) as a function of series length for all 12 estimators with $d = 0.4$ for the GPH and Haslett–Raftery and $H = 0.9$ for the other ten. MSEs are reported starting at a series of 500 data points. (a) Absolute Value, Aggregated Variance and Whittle. (b) Boxed Periodogram, Difference Variance and Higuchi. (c) Peng, Periodogram and R/S. (d) Wavelet, GPH and Haslett–Raftery.

Table 1
The first two columns of results presents the H estimates and p-values returned by the Beran [9] test for the Campito Mountain data for each of the 10 estimators of H and the GPH and Haslett–Raftery estimators of d converted to H equivalent. CPU times are in seconds on the SunBlade described in the text. The expected H column is the expected value that the estimator would report the if the Campito data was an FGN with $H = 0.89$. The empirical p-value column is estimated empirically from the simulated data.

Method	H Est	Beran p-value	CPU seconds	Expected H	Empirical p-value
Absolute value	0.862	0.435	0.19	0.831	0.70
Aggregated variance	0.889	0.577	0.34	0.821	0.04[*]
Boxed periodogram	0.914	0.509	0.09	0.849	0.01[**]
Differenced variance	1.089	–	0.21	0.925	0.01[**]
GPH	1.037	–	1.43	0.897	0.16
Haslett–Raftery	0.947	0.241	0.17	0.911	0.45
Higuchi	0.966	0.102	19.65	0.845	<0.001[***]
Peng	0.936	0.344	18.46	0.875	<0.001[***]
Periodogram	1.007	–	0.06	0.908	<0.001[***]
Rescaled range	0.892	0.577	0.04	0.816	0.36
Wavelet	0.927	0.421	0.07	0.889	0.25
Whittle	0.876	0.540	1.05	0.890	0.15

[*] Significant at the 5% level.
[**] Significant at the 1% level.
[***] Significant at the 0.1% level

observation series. It is evident that some estimators which require longer compute times, such as the Higuchi and Peng, did not necessarily yield a more accurate estimation of H for this data.

6. Discussion

It is clear from the simulations that not all estimators are created equal. Long memory occurs in the range parameter $0.5 < H < 1$. Thus any estimator used to estimate the strength of the long memory needs to be both accurate and have a low variance.

The boxed periodogram method was developed specifically to deal with the problem of having most of the points used to estimate H on the right-hand side of the graph. This was believed to, possibly, cause bias in the periodogram estimator. Beran [10] and Taqqu et al. [55] outline some of the reasons such a method could be expected to be biased. In the series we investigated here the box periodogram estimator is inferior to the periodogram estimator it was intended to improve upon.

The differenced variance method was developed to be robust to trends which were known to cause spurious long memory in the R/S estimator [12]. We did not test its robustness. Teverovsky and Taqqu [56] established that the differenced variance method had a higher variance than the aggregated variance method, a result supported by our simulation study. In fairness to the method it must be pointed out that Teverovsky and Taqqu [56] did not intend for it to be used alone but rather in conjunction with the aggregated variance method to test for the presence of shifting means or deterministic trends.

The Higuchi estimator [32] only indirectly estimates H. It estimates the fractal dimension, D, of a series by estimating its path length. As implemented it then converts the estimate of D to H by the simple relationship $H = 2 - D$. This should be borne in mind if a researcher wishes to estimate D rather than H as it is a simple matter to recover D from the H estimate reported by this implementation.

Taqqu et al. [55] give a detailed proof that the Peng estimator of Peng et al. [46] is asymptotically unbiased. In our simulations the bias of the Peng estimator was never large but even at a sample size of 10,000 observations the estimator cannot be considered unbiased. However, its MSE approaches that of the Periodogram method as the series length increases which, in turn, is better than several others.

The wavelet estimator is asymptotically unbiased. In the simulations the bias was always small and was unbiased for series with longer than or equal to 4100 observations.

7. Conclusions and future research

Of the twelve estimators examined here the Whittle and Haslett–Raftery estimators performed the best on simulated series. If we require an estimator to be close to unbiased across the full range of H values for which long memory occurs and have a 95% confidence interval width of less than 0.1 H or d units (that is 20% of the range for H or d values in which long memory is observed), then for series with less than 4000 data points the Whittle and Haslett–Raftery estimators were the only two estimators worth considering. It should be noted that these estimators did not meet these criteria until the series lengths exceeded 700 and 1000 data points respectively. For series with 4000 or more data points, the Peng estimator gave acceptable performance. For series with more than 7000 data points the periodogram estimator was a worthwhile choice. For series with more than 8200 data points the wavelet became a viable estimator. The remaining seven estimators did not give acceptable performance at any series lengths examined and are not recommended.

The Higuchi estimator is useful if the researcher wishes to recover the fractal dimension of the time series. In contrast to the other estimators it provides useful information about a time series if the series is not an FGN (or FI(d)) series. As an estimator of H it is inferior to several others.

The boxed periodogram method is clearly inferior to the periodogram method it was intended to improve upon for FGNs. Further research would be needed to test if it is more robust than the periodogram method in series with departures from a pure FGN. This could be accomplished, for example, by simulating ARFIMA series with non-zero AR and MA components or series with structural breaks.

The R/S estimator is of considerable historical interest but had a major deficiency in that its MSE plateaued while all other estimators' MSEs decreased with increasing series length. Against this we must note that it was one of the two best performing estimators when applied to the Campito data when judged by the Beran [9] test.

The differenced variance estimator was the worst of the twelve estimators in short series. For series longer than 6000 data points its MSE was better than the R/S and on a par with the absolute value, aggregated variance and Higuchi methods. As noted above, Teverovsky and Taqqu [56] do not recommend its use in isolation as it is part of a test for shifting means or deterministic trends. Teverovsky and Taqqu [56] also recommend the aggregated and differenced variance plots always be examined visually. We agree with these recommendations. We did not test its robustness to shifting means or deterministic trends. Some numerical results on its performance in these two situations can be found in Teverovsky and Taqqu [56].

In the application to the Campito data six of the estimators reported statistically significantly different H estimates than expected based on the evidence from the simulated series. Although the fit of the Campito data to an FGN is good ($p = 0.577$), these six estimators do not seem to be robust to whatever specific departures from an FGN that are present in the data. This suggests that a researcher should not rely on a single estimator when estimating H and that the Beran [9] test should always be applied to test the goodness of fit of the data to an FGN, while being aware that its asymptotic properties are currently unknown.

Because of the apparent lack of robustness, a useful avenue of future research would be to quantify the sensitivity of these estimators to various types of departures from an FGN, e.g. FGN series with a small number of shifts in mean or a small number of outlier data points. Also, it may be worthwhile investigating the problem of model misspecification where times series generated by ARFIMA models with non-zero AR and MA orders are estimated as if they were FI(d) series. Some literature on this work already exists such as Smith et al. [52] and Gil-Alana [25].

Acknowledgements

References

[1] P. Abry, D. Veitch, Wavelet analysis of long-range-dependent traffic, IEEE Transactions on Information Theory 44 (1) (1998) 2–15.

[2] P. Abry, D. Veitch, P. Flandrin, Long-range dependence: revisiting aggregation with wavelets, Journal of Time Series Analysis 19 (3) (1998) 253–266.

[3] C. Agiakloglou, P. Newbold, M. Wohar, Bias in an estimator of the fractional difference parameter, Journal of Time Series Analysis 14 (3) (1991) 235–246.

[4] R.T. Baillie, Long memory processes and fractional integration in economics, Journal of Economics 73 (1996) 5–59.

[5] R.T. Baillie, C.F. Chung, M.A. Tieslau, Analyzing inflation by the factionally integrated ARFIMA-GARCH Model, Journal of Applied Econometrics 11 (1996) 23–40.

[6] R.T. Baillie, S.-K. Chung, Modeling and forecasting from trend-stationary long memory models with applications to climatology, International Journal of Forecasting 18 (2002) 215–226.

[7] J.T. Barkoulas, C.F. Baum, Long memory in forecasting of US monetary indices, Journal of Forecasting 25 (1997) 291–302.

[8] C.F. Baum, J.T. Barkoulas, M. Caglayan, Long memory or structural breaks: can either explain nonstationary real exchanges rates under the current float, Journal of International Markets, Finance and Money 9 (1999) 359–376.

[9] J. Beran, A goodness-of-fit test for time series with long range dependence, Journal of the Royal Statistical Society Series B 54 (3) (1992) 749–760.

[10] J. Beran, Statistics for Long Memory Processes, Chapman & Hall/CRC Press, 1994.

[11] J. Beran, B. Whitcher, M. Maechler, longmemo: Statistics for Long-Memory Processes (Jan Beran) – Data and Functions, R Package Version 0.9-3, 2006.

[12] R. Bhattacharya, V.K. Gupta, E. Waymire, The Hurst effect under trends, Journal of Applied Probability 20 (1983) 649–662.

[13] P. Bloomfield, Trends in global temperature, Climatic Change 21 (1992) 1–16.

[14] J.M. Box-Staffensmeier, R.M. Smith, Investigating political dynamics using fractional integration methods, American Journal of Political Science 42 (2) (1998) 661–689.

[15] R. Cioczek-Georges, B.B. Mandelbrot, A class of micropulses and antipresistent fractional Brownian motion, Stochastic Processes and their Applications 60 (1995) 1–18.

[16] R. Dahlhaus, Efficient parameter estimation for self-similar processes, The Annals of Statistics 17 (4) (1989) 1749–1766.

[17] R.S. Deo, W.W. Chen, On the integral of the squared periodogram, Stochastic Processes and Their Applications 85 (2000) 159–176.

[18] F.X. Diebold, A. Inoue, Long memory and regime switching, Journal of Econometrics 105 (2001) 131–159.

[19] P. Doukhan, G. Oppenheim, M. Taqqu, Theory and applications of long-range dependence, Birkhaüser, 2003.

[20] P. Embrechts, M. Maejima, Selfsimilar Processes, Princeton University Press, 2002.

[21] R. Fox, M.S. Taqqu, Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, The Annals of Statistics 14 (2) (1986) 517–532.

[22] C. Fraley, F. Leisch, M. Maechler, V. Reisen, A. Lemonte, fracdiff: Fractionally Differenced ARIMA aka ARFIMA(p, d, q) Models, R Package Version 1.3-0, 2006.

[23] J. Geweke, S. Porter-Hudak, The estimation and application of long memory time series models, Journal of Time Series Analysis 4 (1983) 221–237.

[24] L.A. Gil-Alana, Fractional integration in daily stock market returns, Review of Financial Economics 15 (2006) 28–48.

[25] L.A. Gil-Alana, Fractional integration and structural breaks at unknown periods of time, Journal of Time Series Analysis 29 (1) (2008) 163–185.

[26] L. Giraitis, P.M. Robinson, D. Surgailis, Variance-type estimation of long memory, Stochastic Processes and their Applications 80 (1999) 1–24.

[27] C.W.J. Granger, Long memory relationships and the aggregation of dynamic models, Journal of Econometrics 14 (1980) 227–238.

[28] C.W.J. Granger, R. Joyeux, An introduction to long-range time series models and fractional differencing, Journal of Time Series Analysis 1 (1980) 15–30.

[29] C.W.J. Granger, M.J. Morris, Time series modelling and interpretation, Journal of the Royal Statistical Society A 139 (2) (1976) 246–257.

[30] S.R. Hare, N.J. Mantua, Empirical evidence for North Pacific regime shifts in 1977 and 1989, Progress in Oceanography 47 (2000) 103–145.

[31] J. Haslett, A.E. Raftery, Space-time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion), Applied Statistics 38 (1) (1989) 1–50.

[32] T. Higuchi, Approach to an irregular time series on the basis of fractal theory, Physica D 31 (1988) 277–283.

[33] L. Horvath, Q.-M. Shao, Limit theorems for quadratic forms with applications to Whittle's estimate, The Annals of Statistics 9 (1) (1999) 146–187.

[34] J.R.M. Hosking, Fractional differencing, Biometrika 68 (1) (1981) 165–176.

[35] H.E. Hurst, Long-term storage capacity of reservoirs, Transactions of the American Society of Civil Engineers 116 (1951) 770–808.

[36] M.J. Jensen, Using wavelets to obtain a consistent ordinary least square estimator of the long-memory parameter, Journal of Forecasting 18 (1999) 17–32.

[37] H.-D.J. Jeong, J.-S.R. Lee, D. McNickle, K. Pawlikowski, Comparison of various estimators in simulated FGN, Simulation and Modelling Practice and Theory 15 (2007) 1173–1191.

[38] V. Klemes, The Hurst phenomenon – a puzzle? Water Resources Research 10 (4) (1974) 675–688.

[39] I. Lobato, P.M. Robinson, Averaged periodogram estimation of long memory, Journal of Econometrics 73 (1996) 303–324.

[40] B.B. Mandelbrot, Limit theorems of the self-normalized range for weakly and strongly dependant processes, Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete 31 (1) (1975) 271–285.

[41] B.B. Mandelbrot, M.S. Taqqu, Robust R/S analysis of long-run serial correlation, in: Proceedings of the 42nd Session of the International Statistical Institute, vol. 48 of Bulletin of the International Statistical Institute, 1979, pp. 69–104.

[42] B.B. Mandelbrot, J.W. van Ness, Fractional Brownian motions, fractional noises and applications, SIAM Review 10 (4) (1968) 422–437.

[43] E.J. McCoy, A.T. Walden, Wavelet analysis and synthesis of stationary long-memory processes, Journal of Computational and Graphical Statistics 5 (1) (1996) 26–56.

[44] J.E. Overland, D.B. Percival, H.O. Mofjeld, Regime shifts and red noise in the North Pacific, Deep-Sea Research I 53 (2006) 582–588.

[45] W. Palma, Long-Memory Time Series Theory and Methods, Wiley-Interscience, 2007.

[46] C.K. Peng, S.V. Buldyrev, M. Simons, H.E. Stanley, A.L. Goldberger, Mosaic organization of DNA nucleotides, Physical Review E 49 (1994) 1685–1689.

[47] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2005, ISBN 3-900051-07-0, http://www.R-project.org

[48] P.M. Robinson, Semiparametric analysis of long-memory time series, The Annals of Statistics 22 (1) (1994) 515–539.

[49] P.M. Robinson, Time Series with Long Memory, Oxford University Press, 2003.

[50] D.L. Rudnick, R.E. Davis, Red noise and regime shifts, Deep-Sea Research I 50 (2003) 691–699.

[51] K. Shimotsu, P.C.B. Phillips, Exact local Whittle estimation of fractional integration, The Annals of Statistics 33 (4) (2005) 1890–1933.

[52] J. Smith, N. Taylor, S. Yadav, Comparing the bias and misspecification in ARFIMA models, Journal of Time Series Analysis 18 (5) (1997) 507–527.

[53] F. Sowell, Maximum likelihood estimation of stationary univariate fractionally integrated time series models, Journal of Econometrics 53 (1992) 165–188.

[54] D.B. Stephenson, V. Pavan, R. Bojariu, Is the North Atlantic oscillation a random walk? International Journal of Climatology 20 (2000) 1–18.

[55] M. Taqqu, V. Teverovsky, W. Willinger, Estimators for long-range dependence: an empirical study, Fractals 3 (4) (1995) 785–798.

[56] V. Teverovsky, M. Taqqu, Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator, Journal of Time Series Analysis 18 (3) (1999) 279–304.

[57] D. Veitch, P. Abry, A wavelet-based joint estimator of the parameters of long-range dependence, IEEE Transactions on Information Theory 45 (3) (1999) 878–897.

[58] W. Willinger, V. Paxson, R.H. Riedi, M.S. Taqqu, Long-range dependence and data network traffic, in: P. Doukhan, G. Oppenheim, M. Taqqu (Eds.), Theory and Applications of Long-Range Dependence, Birkhaüser, 2003.

[59] D. Wuertz, fSeries: Financial Software Collection, R Package Version 220.10063, 2005, http://www.itp.phys.ethz.ch/econophysics/R/2.1.