PROD. TYPE: COM
PP: 1-16 (col.fig.: nil)
COMSTA3369 MODC+
ED: Vijay
PAGN: Vidya – SCAN: Global

ELSEVIER
Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Estimation of Hurst exponent revisited

J. Mielniczuk[a,b,*], P. Wojdyłło[a,c]

[a] Institute of Computer Science, Polish Academy of Sciences, Ordona 21, 01-237 Warsaw, Poland
[b] Warsaw University of Technology, Plac Politechniki 1, 00-601 Warsaw, Poland
[c] Institute of Mathematics, Polish Academy of Sciences, Śniadeckich 8, 00-950 Warsaw, Poland

Received 1 November 2004; received in revised form 17 July 2006; accepted 17 July 2006

Abstract

In order to estimate the Hurst exponent of long-range dependent time series numerous estimators such as based e.g. on rescaled range statistic (R/S) or detrended fluctuation analysis (DFA) are traditionally employed. Motivated by empirical behaviour of the bias of R/S estimator, its bias-corrected version is proposed. It has smaller mean squared error than DFA and behaves comparably to wavelet estimator for traces of size as large as 2^{15} drawn from some commonly considered long-range dependent processes. It is also shown that several variants of R/S and DFA estimators are possible depending on the way they are defined and that they differ greatly in their performance.

MSC: 60F05; 60F17

Keywords: Detrended fluctuation analysis (DFA) estimator; Hurst exponent; Long-range dependence; Rescaled adjusted range statistic (R/S); Scaling property; Wavelet estimator

1. Introduction

Let $(X_t)_{t=1}^{\infty}$ be a real-valued stationary time series such that its covariance function $\gamma(k) := \mathrm{Cov}\,(X_t, X_{i+k}) \sim c_\gamma |k|^{-\gamma}$ for $|k| \to \infty$, where $0 < \gamma < 1$ and $\sim$ denotes asymptotic equivalence. This is the most important case of long-range dependence (LRD), which in a general situation is defined by the condition $\sum_{k=0}^{\infty} |\gamma(k)| = \infty$, and it encompasses two frequently studied processes having this property, namely a fractional Gaussian noise (fGn) and a fractional autoregressive integrated moving average (FARIMA). The phenomenon of long-range dependence is a topic of active research in statistics as well as in many areas of applied sciences e.g. in economics, geophysics and meteorology. We refer to Beran (1994) for a book-length treatment of this subject.

Assuming the above condition of the hyperbolic decay of covariance, it is easily seen that, with $Y_n = \sum_{t=1}^{n} X_t$, we have

$$\mathrm{Var}\,(Y_n) \sim \frac{2c_\gamma}{(1-\gamma)(2-\gamma)} n^{2-\gamma} =: C_\gamma n^{2H}, \tag{1}$$

* Corresponding author. Institute of Computer Science, Polish Academy of Sciences, Ordona 21, 01-237 Warsaw, Poland. Tel.: +48 22 8362841; fax: +48 22 8376564.

E-mail address: miel@ipipan.waw.pl (J. Mielniczuk).

0167-9473/$ - see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.csda.2006.07.033

where $H := 1 - \gamma/2$ is traditionally called the Hurst exponent (cf. Hurst (1951)). As the Hurst exponent describes the strength of dependence, its estimation is of a great interest. A process with a larger value of H is more regular and less erratic than a process with a smaller one. Most estimators are based on scaling properties similar to a property following from (1), namely that $\mathrm{Var}\,(Y_{nk}) \sim k^{2H}\mathrm{Var}\,(Y_n)$, for any $k \in \mathbb{N}$. Thus e.g. regressing a logarithm of some estimator of the variance of the partial sums against the logarithm of its size yields an estimator of $2H$. However, as in the case described above, this does not always yield a reliable estimator of H. For a discussion of such approach and a review of long-range dependence see Robinson (1994). A Monte Carlo experiment comparing performance of some such estimators is discussed in Taqqu et al. (1995).

We consider below three of the most popular estimators based on scaling property: R/S estimator based on rescaled adjusted range statistic, DFA estimator pertaining to detrended fluctuation analysis and a wavelet estimator. Important competitors include, among others, the log periodogram estimator (Geweke and Porter-Hudak, 1983) based on an approximate scaling of spectral density and the Whittle (1953) estimator defined as the maximizer of an approximate version of loglikelihood. A semiparametric version of the last proposal called the local Whittle (LW) or Gaussian semiparametric estimator (Robinson, 1995) is proved to be more efficient than the log periodogram regression estimator. Modifications of the LW estimator intended to reduce its bias when the underlying LRD process is contaminated by noise have been proposed and studied by Andrews and Sun (2004), Arteche (2004) and Hurvich et al. (2005). They are shown to have a smaller mean squared error (MSE) than the previous proposals when the noise to signal ratio is considerable (see e.g. Hurvich and Ray, 2003). For recent contributions concerning estimation of Hurst exponent see also Lai (2004) and Stoev et al. (2006).

In Sections 2 and 3 we show that the performance of the studied estimators depends crucially on a way they are constructed. Moreover, as the main contribution of the paper, we indicate that R/S estimator, which is immensely popular among practitioners, but is widely known to be suboptimal can be enhanced by a bias correction to the effect that it outperforms the DFA estimator with respect to the MSE and performs on par with the wavelet estimator. In Section 2 we describe possible variants of R/S estimator, choose one of them and show how its performance can be improved by a bias correction. The bias correction method is based on approximate linearity of the bias of considered R/S estimator and takes advantage of the exact value of the slope of the approximating line.

In Section 3 careful analysis of the DFA estimator is provided. In Section 4 we briefly describe the wavelet estimator and in Section 5 we compare the performance of all three semiparametric estimators and the Whittle estimator for FARIMA and fGn processes with provided sample sizes ranging from 2^9 to 2^{15}. An application to analysis of daily exchange rates listed by the National Bank of Poland is discussed. Section 6 concludes the paper.

2. R/S method

For a trace $(X_t)_{t=1}^n$ of a time series, consider a partial sum $Y_k = \sum_{t=1}^k X_t,\ 1 \leqslant k \leqslant n$, and a sample variance $S_n^2 = (n-1)^{-1}\sum_{t=1}^n \left(X_t - \bar{X}_n\right)^2$, where $\bar{X}_n = n^{-1}\sum_{t=1}^n X_t$ is a sample mean. Rescaled adjusted range statistic $R/S(n)$ introduced by Hurst (1951) is defined as

$$R/S(n) = \frac{1}{S_n}\left\{\max_{1\leqslant k\leqslant n}\left(Y_k - \frac{k}{n}Y_n\right) - \min_{1\leqslant k\leqslant n}\left(Y_k - \frac{k}{n}Y_n\right)\right\}. \tag{2}$$

Observe that the numerator R_n in (2) can be viewed as a range of partial sums of $X_t - \bar{X}_n,\ t = 1, \ldots, n$, or, equivalently, as the sum of the maximal and the minimal distance of the partial sums $Y_k,\ k = 1, \ldots, n$ from a line passing through $Y_0 = 0$ and Y_n. Thus $R/S(n)$ is a measure of fluctuations of the partial sums of $(X_t)_{t=1}^n$ scaled by the standard deviation of observations. Obviously, $R/S \geqslant 0$ as the range of a collection of random variables is nonnegative; actually, $\max_{1\leqslant k\leqslant n}(Y_k - (k/n)Y_n) \geqslant 0$ and $\min_{1\leqslant k\leqslant n}(Y_k - (k/n)Y_n) \leqslant 0$. Its relevance in estimation of the Hurst coefficient H is demonstrated by the following theorem which is a slightly restated result in Mandelbrot (1975); see also Giraitis et al. (2003). By $\xrightarrow{\mathscr{D}}$ we denote convergence in distribution and $\xrightarrow{\mathscr{P}}$ stands for convergence in probability.

Theorem 1. *If $(X_t)_{t=1}^\infty$ is a stationary ergodic process such that $\left(X_t^2\right)$ is ergodic and for some process $B_H(\cdot)$*

$$n^{-H}\left(Y_{[nt]} - [nt]\mathbb{E}X_1\right) \xrightarrow{\mathscr{D}} \sigma_X B_H(t) \quad in \quad \mathscr{D}[0,1],$$

where $\sigma_X^2 = \mathrm{Var}X_1 < \infty$, *then*

$$Z_n = n^{-H} R/S(n) \xrightarrow{\mathscr{D}} Z := \sup_{t\in[0,1]} \left(B_H(t) - tB_H(1)\right) - \inf_{t\in[0,1]} \left(B_H(t) - tB_H(1)\right).$$

The result follows easily from the observation that $\max_{1\leqslant k\leqslant n}\sum_{t=1}^{k}\left(X_t - \bar{X}_n\right) = \max_{1\leqslant k\leqslant n}\left(\sum_{t=1}^{k}\left(X_t - \mathbb{E}X_t\right) - k/n\sum_{t=1}^{n}\left(X_t - \mathbb{E}X_t\right)\right)$ with an analogous equality holding for $\min_{1\leqslant k\leqslant n}\sum_{t=1}^{k}\left(X_t - \bar{X}_n\right)$. Thus, continuous mapping theorem implies that $n^{-H}R_n \xrightarrow{\mathscr{D}} \sigma_X Z$. Moreover, by ergodicity of (X_t) and $\left(X_t^2\right)$, $S_n \xrightarrow{\mathscr{P}} \sigma_X$.

For i.i.d. sample process $B_H(\cdot)$ is a Brownian motion on $[0, 1]$ and the asymptotic limit Z is the range of the associated Brownian bridge, the distribution of which is given e.g. in Kennedy (1976). For a long-range dependent Gaussian process or a linear process the limiting process $B_H(\cdot)$ is a fractional Brownian motion with Hurst exponent $1/2 < H < 1$.

Observe that $\log R/S(n) = \mathbb{E}Z_n + H\log n + (\log Z_n - \mathbb{E}Z_n)$, where in view of the Theorem 1 a distribution of Z_n resembles that of Z for large n. This gives rise to a regression type estimator of H. Two versions of such an estimator are considered in the literature differing in the choice of blocks and response variable in regression.

2.1. Variants of R/S estimator

1. *Estimator* $\hat{H}_{o,p}$ (Beran, 1994; Taqqu et al., 1995): Let $R/S(i,k)$ be R/S statistic calculated for the block of data $X_i, \ldots, X_{i+k-1}$ of size k starting at a time i such that $i + k - 1 \leqslant n$. For each value of k, statistic R/S is calculated for a number of possibly overlapping blocks of size k starting at i_j and a least squares (LS) line is fitted to a pox-plot $\left(\log k, \log R/S\left(i_j, k\right)\right)_{i_j\in I, k\in K_n}$, where I and K_n are some sets of indices to be specified later. The resulting estimator will be called $H_{o,p}$ with the index o, p standing for "overlapping" and "pox-plot".

2. *Estimator* $\hat{H}_{d,a}$ (e.g. Peters, 1994): Statistic R/S is calculated for $n_k = [n/k]$ disjoint blocks for each k and the resulting values are averaged yielding

$$\overline{R/S}(k) = \frac{1}{n_k}\sum_{j=1}^{n_k} R/S\left(i_j, k\right),$$

where $i_j = (j-1)k + 1$ for $j = 1, \ldots, n_k$. Then the LS line is fitted to points $\left(\log k, \log \overline{R/S}(k)\right)$. This estimator will be called $\hat{H}_{d,a}$ with the subscript d, a standing for "disjoint" and "averaged".

The differences between the two estimators consist in the method of choosing blocks and the way a scatterplot is constructed (consisting either of all points $\log R/S\left(i_j, k\right)$ or just logged averages $\log \overline{R/S}(k)$). One would expect greater dependence among values of R/S statistic calculated for overlapping than for disjoint blocks. We show that these two apparently minor differences in implementation are the source of major differences in behaviour of the estimators.

A test based on R/S statistic tends to reject independence in favour of long-range dependence in the case when the underlying process is in fact short-range dependent (cf. Davies and Harte, 1987). Lo (1991) suggested alternative scaling of R/S statistic to account for possible short-range dependence of X_n which consisted in replacing S_n^2 with an estimator of $n^{-1}\mathrm{Var}\left(Y_n\right)$. Taqqu et al. (1999) discussed difficulties connected with such approach and showed that the pertaining test has a low power for long-range dependent alternatives if a tuning parameter for the estimator is large. Giraitis et al. (2003) studied modified Lo's statistic with a standard deviation of the partial sums of $X_t - \bar{X}_n$ replacing numerator R_n in (2) and investigated properties of a pertaining test under short-memory hypotheses against long-memory alternatives.

A result of Hall et al. (2001) for an underlying process obtained as a result of sampling from continuous time process with a frequency $1/n$ and for a fixed number of k implies that the asymptotic distribution of $\hat{H}_{d,a}$ depends on whether $0 < H < \frac{3}{4}$ or $\frac{3}{4} \leqslant H < 1$. In the first case the distribution is normal and in the second case it is a mixture of Rosenblatt distributions.

To get a deeper insight into the behaviour of R/S estimator for small and moderate sample sizes, we also consider its two natural variants, namely:

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.

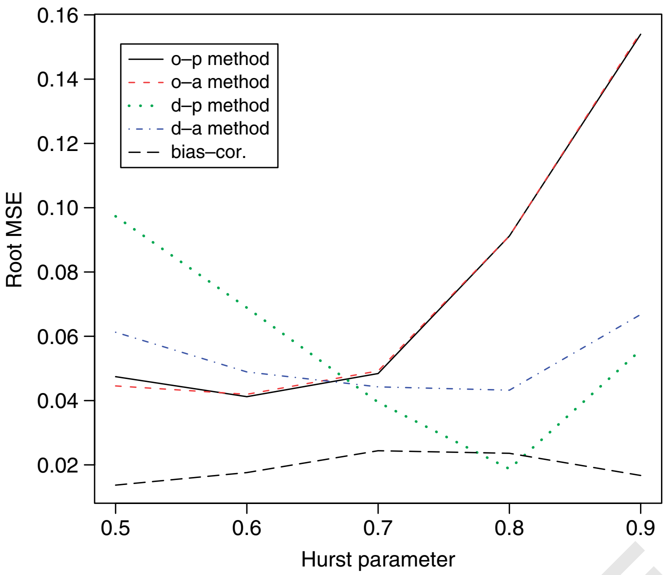

Fig. 1. Root MSE of four variants of R/S method for the fGn with $n = 2048$ together with bias-corrected $\hat{H}_{d,p}$.

3. *Estimator* $\hat{H}_{o,a}$ which pertains to averaged values of R/S statistic based on overlapping blocks.

4. *Estimator* $\hat{H}_{d,p}$, for which values of R/S statistic are calculated for disjoint blocks and the estimator is based on the pertaining pox-plot. This method does not use averaged values of R/S statistic.

Remark 1. Observe that the LS estimator for a homoscedastic regression model with n_i replications at ith level of x:

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i,$$

where ε_{ij} are i.i.d. with mean 0 and variance σ^2 coincides with a weighted least squares (WLS) estimator in the heteroscedastic regression model $\bar{y}_i = \beta_0 + \beta_1 x_i + \bar{\varepsilon}_i$, $i = 1, \ldots, k$, where $\bar{y}_i$ (resp. $\bar{\varepsilon}_i$) are averages of responses y_{ij} (resp. ε_{ij}) pertaining to x_i. Corresponding weights satisfy $w_i^{-1} = \sigma^2/n_i$. It follows easily from comparison of normal equations in both cases. Thus the estimator $\hat{H}_{d,p}$ is equivalent to WLS estimator for a response $y_k = n_k^{-1} \sum_{j=1}^{n_k} \log R/S\left(i_j, k\right)$ and the same is true for $\hat{H}_{o,p}$. However, $\hat{H}_{d,a}$ and $\hat{H}_{o,a}$ are constructed differently and that is why we considered it worthwhile to compare performance of averaged and pox-plot estimators. Precise settings are as follows. We considered only dyadic values of $n = 2^N$, where $N = 9, 10, \ldots, 15$. Throughout the paper N stands for $\log_2 n$. Overlapping blocks start at $i_j = 30 \times 2^{N-9} j + 1$, $j = 0, 1, \ldots, 13$ and have lengths $k = 10, 20, \ldots, 100 \times 2^{N-9}$. This setup is a direct generalization of a setting used by Beran (1994, p. 81). For the estimators based on disjoint blocks we used dyadic block sizes from $k = 2^1$ to $k = 2^N$.

2.2. Comparison of performance

Choice of estimator: We discuss now results on performance of all four estimators for the fGn and FARIMA(0, d, 0) processes with $H = 0.5, 0.6, \ldots, 0.9$ based on 500 simulations. For definition of these processes we refer to Sections 2.4 and 2.5 in Beran (1994) and note that the parameter of fractional differencing d equals $H - \frac{1}{2}$. Results behave in a stable manner for this number of replications, the absolute change of the root mean squared error (RMSE) in replications of the whole experiment was less than 0.002 for any choice of n and H. Fig. 1 shows the behaviour of the empirical RMSE for four R/S estimators of H for traces of the fGn having 2048 observations. The results for remaining sample sizes n show the same pattern of behaviour. Namely, the estimators based on overlapping blocks perform better than the ones based on disjoint blocks for smaller H ($\hat{H}_{o,a}$ is the best for $H = 0.5$ and $\hat{H}_{o,p}$ for $H = 0.6$) whereas $\hat{H}_{d,p}$ is a clear winner for $H \geqslant 0.7$ yielding in particular for $H = 0.8$ more than fourfold decrease of RMSE with respect to estimators based on overlapping blocks and twofold decrease when compared to its averaged counterpart $\hat{H}_{d,a}$.

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.

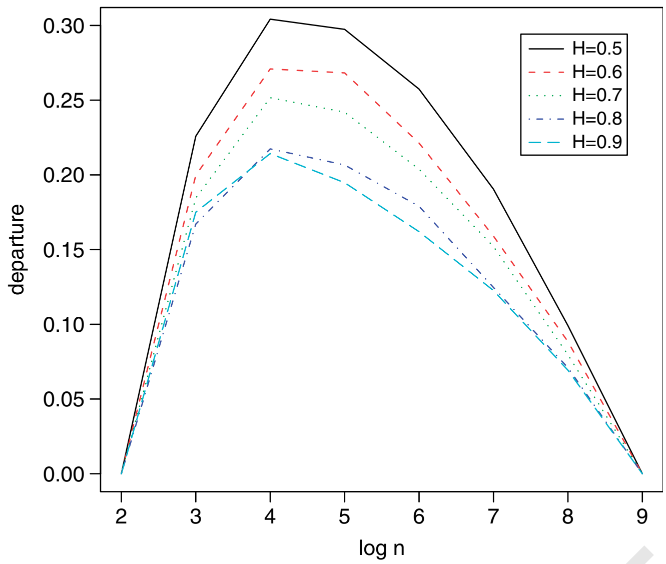

Fig. 2. Departure of $\log \mathbb{E}R/S\left(2^i\right)$, $i=2,\ldots,9$ from the line joining the first and the last point for fGn process.

Performance of both estimators based on overlapping blocks is very similar and better than the remaining ones for smaller H. Thus it seems that for such H strengthening the dependence of R/S statistic calculated for consecutive blocks is advantageous. Because of its strong performance for large H we have chosen $\hat{H}_{d,p}$ as a representative of the class of the R/S estimators for further considerations. Fig. 1 also shows RMSE of a bias-corrected $\hat{H}_{d,p}$described in Section 2.3. Its RMSE behaves in a very stable manner for $H \in [0.5, 0.9]$ and is significantly smaller than for other estimators based on R/S statistic except for a neigbourhood of 0.8, where it is slightly larger than that of its uncorrected version $\hat{H}_{d,p}$. In comparison with $\hat{H}_{d,p}$ the gain is the most significant for H close to 0.5, where the uncorrected estimator is heavily biased.

Minimal octave: An additional tuning parameter of $\hat{H}_{d,p}$ is a minimal length of block k taken into consideration. Let $i_1 \in \mathbb{N}$ be such that $k=2^{i_1}$ is the size of the minimal block referred to as the minimal octave of R/S estimator. We considered $i_1 = 1, 2, \ldots, 6$ as possible values of the index. The standard deviation of $\hat{H}_{d,p}$ is empirically established to be an increasing function and its bias is a decreasing function of i_1 for all considered cases. The minimal value of MSE is attained for i_1^* corresponding to approximate equality of the bias and the standard deviation. Moreover, simulations indicate that i_1^* is a *decreasing* function of H. This is valid for all considered sample sizes in the case of the fGn and FARIMA processes. This can be understood by noting that the bias of $\hat{H}_{d,p}$ is much larger than the standard deviation for the independent case when $i_1 = 1$ and decreases as a function of the minimal octave in such a way that an approximate equality of those two quantities is attained only for $i_1^* \geqslant 5$. However, for $H = 0.9$ approximate equality holds for $i_1^* = 2$. Thus the wrong choice of the minimal octave, especially for small H may lead to heavily biased estimator. An intermediate value of $i_1 = 3$ corresponding to the minimal block size equal 8 yields reasonable values of RMSE for $\hat{H}_{d,p}$ in the whole range of H and we propose to use it as the cut-off point of this particular version of R/S estimator.

Let us mention that difficulty in choosing the cut-off point i_1 was noted e.g. by Beran (1994, p. 84). In order to illustrate this, Fig. 2 displays the departure of $\log \mathbb{E}R/S\left(2^i\right)$, $i = 2, \ldots, 9$, from a line joining the first and the last values.

Observe that $\log \mathbb{E}R/S\left(2^i\right)$ is for $n \to \infty$ the limit of $\log \overline{R/S}\left(2^i\right)$ from which values of $\hat{H}_{d,a}$ are constructed. The plot indicates that the departures are concave functions and they diminish with increasing H. The discrepancy among the curves corresponding to $H = 0.8$ and 0.9 is much smaller than among the remaining ones.

Two alternative measures of spread of partial sums of $X_t - \bar{X}_n$, namely an interquartile range IQR and $2\times$ maximal absolute deviation, were tried instead of the range. Both show improved performance for H equal 0.5 and 0.6 and perform worse than $\hat{H}_{d,p}$ for $H \geqslant 0.8$. It is worth noting that a replacement of the range by IQR yields four times decrease of

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.

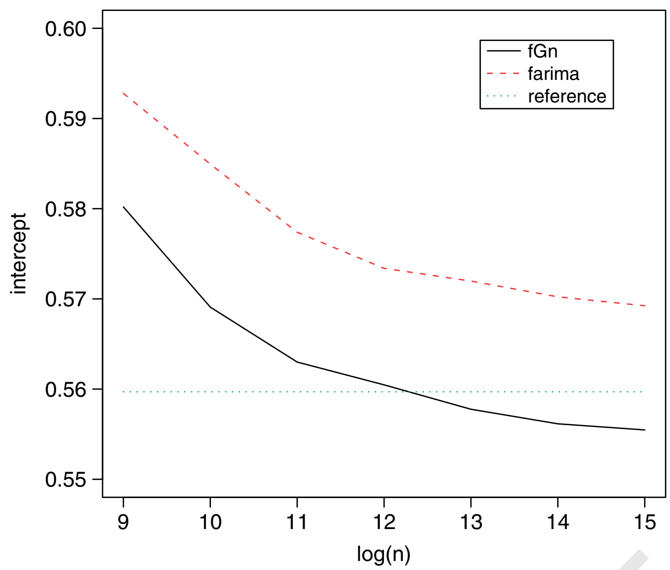

Fig. 3. Intercepts of the LS line approximating bias of $\hat{H}_{d,a}$ for fGn and FARIMA processes.

RMSE for fGn with $n = 2^9$ and $H = 0.5$ and six times decrease of bias of $\hat{H}_{d,p}$ indicating that the performance of R/S, especially for smaller H, is hindered by its considerable bias. An alternative statistic introduced in Kwiatkowski et al. (1992), whenceforth its name KPSS, for which the range in the numerator of (2) is replaced by a square root of a second moment of $Y_k - (k/n)Y_n$ was found to perform even worse than the alternative estimators for $H \geqslant 0.7$.

2.3. *The bias-corrected R/S estimator*

Observing that the bias is the main contribution to the MSE of R/S estimator leads immediately to a problem of accounting for it. Asymptotic form of the bias of $R/S(k)$ is known for $H = 0.5$ and Gaussian X_t only (cf. e.g. Peters, 1994, p. 69) but even in this case this does not translate to the expression for the bias of the pertaining estimator of H. However, results of simulation experiments for the fGn processes show that the bias of $\hat{H}_{d,p}$ decreases approximately linearly as the function of H for all considered sample sizes n and the slope of LS line varies negligibly with n around the value -0.618. Moreover, the LS line remains practically unchanged when the fGn process is replaced by the FARIMA$(0, d, 0)$ process. In Figs. 3 and 4 the slopes and the intercepts of LS approximation to the bias are given for both processes for the cut-off point $i_1 = 1$. The value of R^2 is around 0.99 for both processes in the considered range of n. For the larger cut-off points, the slopes of the LS line become less negative. Observe that the existence of limiting position of the LS lines when $n \to \infty$ suggested by Figs. 3 and 4 can be intuitively understood by noting that $\hat{H}_{d,p}$ for the sample size $2n$ is based on a cloud of values of R/S statistic consisting of two clouds pertaining to the sample size n each and a single value pertaining to the largest block of the size $2n$ which has diminishing influence on the position of the line. Approximately linear behaviour of the bias of $\hat{H}_{d,p}$ motivates the method of its correction described below.

Let $f(H) = aH + b$ be an equation of some line approximating the bias for a particular choice of the cut-off point i_1. Consider a bias-corrected estimator $\hat{H}_1 = \hat{H} - f\left(\hat{H}\right)$ with $\hat{H}$ denoting R/S estimator. Note that $\hat{H}_1$ is still a biased estimator as f is only an approximation to the bias and more importantly, because bias correction should have consisted in subtraction of unknown quantity $f(H)$ instead of $f\left(\hat{H}\right)$. Moreover, the standard deviation changes by the factor $|1 - a|$ in comparison with the initial estimator and thus it increases for negative values of a. The actual gain of a bias correction procedure depends on whether the decrease of the bias outweighs the increase of the standard deviation.

The proposed procedure uses the empirical observation discussed above that the bias of $\hat{H}_{d,p}$ is approximately linear with a slope around the value -0.618. We show in Theorem 2 that if the bias were exactly linear with this slope, a twofold bias correction of $\hat{H}_{d,p}$ would give unbiased estimator of H. Motivated by this simple result, we investigate the performance of twice bias-corrected $\hat{H}_{d,p}$. Note that the number of iterations of bias correction is linked to the value

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.

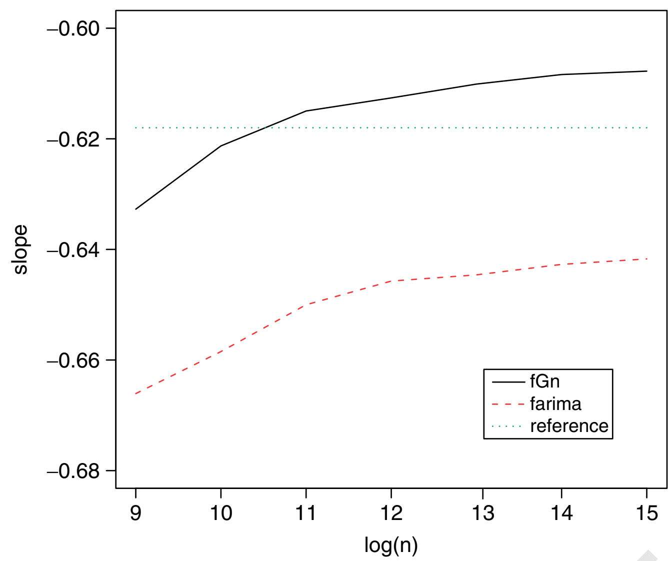

Fig. 4. Slopes of the LS line approximating bias of $\hat{H}_{d,a}$ for fGn and FARIMA processes.

of the slope of the bias. This modification yields a substantial reduction of the RMSE due to a tight concentration of the values of the bias around a line with a slope −0.618 and the fact that the bias is the main contribution to the RMSE. Corollary 3 states values of the bias and the variance of twice corrected $\hat{H}_{d,p}$.

Theorem 2. *Assume that $\hat{H}$ is an estimator of H such that its bias at H, $f(H) = \mathbb{E}\hat{H} - H$ equals $aH + b$ for $\frac{1}{2} \leqslant H < 1$ when a, b are such that $ab \neq 0$. Let T be a bias correcting transformation $T\left(\hat{H}\right) = \hat{H} - f\left(\hat{H}\right)$. Then $\hat{H}_c = T^2\left(\hat{H}\right)$ is unbiased if and only if $a = \left(1 \pm \sqrt{5}\right)/2$, i.e. if $a \approx 1.618$ or $a \approx -0.618$.*

Proof. As $T = I - f$ with I denoting the identity transform, unbiasedness of $\hat{H}_c$ is equivalent to $\mathbb{E}(I - f)^2\hat{H} = H$. Since f is linear

$$\mathbb{E}(I - f)^2\hat{H} = (I - f)^2\mathbb{E}\hat{H} = (I - f)^2(I + f)(H) = H.$$

It is easy to check that $(I - f)^2(I + f)(H) = \left(1 - a\left(a^2 - a - 1\right)\right)H + b\left(a^2 - a - 1\right)$. From this the result follows. □

By the same token it may be checked that a has to be a root of a kth degree polynomial $(1 - a)^k(1 + a) - 1$ for $T^k\left(\hat{H}\right)$ to be unbiased. For $k = 1$ the only solution is $a = 0$ meaning that the one time correction works only for a constant bias. For $k = 3$ the only real solution is approximately −0.839 and it decreases when $k \geqslant 3$. As for the larger cut-off points, the slopes of the LS line become less negative and it follows that iterating the bias-correction procedure beyond $k = 2$ will not improve the bias of the resulting estimator. The proof of Theorem 2 implies also the following corollary in which linearity of the bias is *not* assumed.

Corollary 3. *Let $g(H) = -0.618H + b$ and $\hat{H}_c$ be defined as in Theorem 2 with $f(H)$ replaced by $g(H)$. Then*

$$RMSE\left(\hat{H}_c\right) = (1.618)^2\left\{\mathrm{res}\left(\hat{H}\right)^2 + \mathrm{Var}\left(\hat{H}\right)\right\}^{1/2},$$

where res $\left(\hat{H}\right)$ is the residual value of the bias of $\hat{H}$ with respect to $g(H)$, i.e. $\mathrm{res}\left(\hat{H}\right) = \mathbb{E}\hat{H} - H - (-0.618H + b)$.

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.

Table 1
Intercepts and maximal absolute deviations of the bias of $\hat{H}_{d,p}$ from the line with slope −0.618 for fGn and FARIMA processes $\left(n = 2^N\right)$

N	fGn b	fGn max $\lvert e\rvert$	FARIMA b	FARIMA max $\lvert e\rvert$	Average b	fGn max $\lvert e\rvert$	FARIMA max $\lvert e\rvert$
9	0.5695	0.0085	0.5579	0.0132	0.5637	0.0097	0.0190
10	0.5667	0.0064	0.5556	0.0112	0.5612	0.0086	0.0168
11	0.5652	0.0052	0.5542	0.0092	0.5597	0.0088	0.0147

Observe that in view of the corollary it is very easy to assess the effect of double bias correcting scheme when the "ideal" slope −0.618 of the correcting line is used. Namely, the bias of the doubly corrected estimator depends solely on the residual value of the bias of the original estimator with respect to this line. Table 1 gives values of the intercepts $\hat{b}$ for the LS line with the fixed slope equal to −0.618 fitted to the bias curve for the fGn and FARIMA processes together with the corresponding value of the maximal absolute residual. Also the analogous values are given when the averages of the intercepts for these processes are used for the reference line. We see that the approximation is very good even though the LS line was changed to the LS line with the fixed slope −0.618.

In view of that the following correction of $\hat{H}_{d,p}$ is proposed. It consists of a double bias correction of this estimator with respect to the reference line

$$g(H) := -0.618H + 0.5597,$$

where the intercept 0.5597 is the average of the intercepts for the fGn process (equal to 0.5652) and the FARIMA process (equal to 0.5542) for $N = 11$. This estimator will be referred to as $\hat{H}_{R/S}$ from now on. We stress that the number of iterations of the bias correction procedure is motivated by empirically established behaviour of the bias.

Fig. 1 displays the effect of bias correction for $\hat{H}_{d,p}$ estimator and $n = 2048$; the same effect occurs for all sample sizes in the range from 2^9 to 2^{15} and for both considered processes. We stress that it follows from Corollary 1 that $\hat{H}_{R/S}$ is not asymptotically unbiased for a particular value of H unless asymptotic value of $res\left(\hat{H}\right)$ is 0. This obviously can not be hoped for in a case of all H. However, simulation experiments discussed in Section 5 indicate that due to the bias correction the bias of $\hat{H}_{R/S}$ is negligible in the range of $9 \leqslant N \leqslant 15$.

2.4. Filter interpretation of R/S estimator

It is known that the wavelet filters are represented in the form of filter banks (cf. e.g. Meyer, 1993, Chapter 3). Statistic R_n defined as the numerator in (2) can be represented in the similar form. Let us fix our attention on blocks of length n. Deviations of the partial sum process Y_k from the line connecting Y_0 and Y_n are given by

$$X_1 - \overline{X},\ X_1 + X_2 - 2\overline{X},\ X_1 + X_2 + X_3 - 3\overline{X}, \ldots$$

or by the system of linear filters acting on the vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$

$$\mathbf{v}^k = (1 - k/n, \ldots, 1 - k/n, -k/n, \ldots, -k/n), \quad k = 1, \ldots, n,$$

where the term $1 - k/n$ (respectively, $-k/n$) occurs k times (respectively, $n - k$ times). A similar reasoning can be found in Anis and Lloyd (1976). The output of filters $\mathbf{v}^k$ is n-decimated, i.e. we retain only every nth value omitting the remaining ones.

Let us define two seminorms m and s in $\mathbb{R}^n$ by

$$m(y) := \max_i (y_i) - \min_i (y_i) = \max_{i \neq j} \left(\lvert y_i - y_j\rvert\right) \quad \text{and} \quad s(y) := \sqrt{\sum (y_i - \overline{y})^2}.$$

The R/S statistic corresponding to the block $\mathbf{X}$ equals $m(\mathbf{Z})/s(\mathbf{X})$, where $\mathbf{Z} = \left(\mathbf{v}^k \mathbf{X}^{\mathrm{T}}\right)_{k=1}^n$ and T denotes transposition.

Note that the process $\left(Z_j^k\right)_{j=1}^{\infty}$, where $Z_j^k = \mathbf{v}^k \left(X_{jn+1}, X_{jn+2}, \ldots, X_{jn+n-1}\right)^{\mathrm{T}}$ is weakly dependent for each $k = 1, \ldots, n$. Indeed, the power function of $\mathbf{v}^k$ equals zero at $\lambda = 0$ since $\sum_i \mathbf{v}_i^k = 0$.

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.

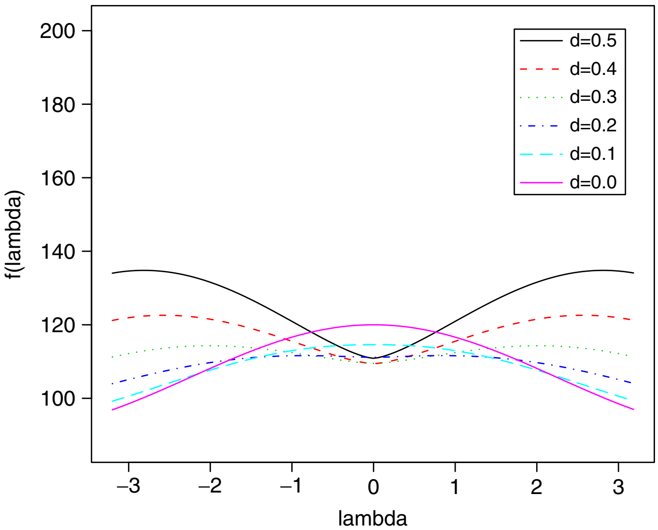

Fig. 5. Spectral densities of a process with a spectral density $\sim t^{-d}$ filtered by the filter $\mathbf{v}^6$ for $n = 16$.

The process with spectral density f passing through the filter with power function V and then n-decimated has the spectral density (cf. Percival and Walden, 2000)

$$f^1(t) = \sum_{l=0}^{n-1} f\left(\frac{t}{n} + \frac{2\pi l}{n}\right) V\left(\frac{t}{n} + \frac{2\pi l}{n}\right).$$

Fig. 5 shows spectral densities f_d^1 corresponding to an input process with $f(t) = C|t|^{-d}$ and $\mathbf{v}^6$ with $n = 16$, $d = 0, 0.1, \ldots, 0.5$.

We note that sum of squared errors (SSE) statistic appearing in a definition of DFA estimator discussed below can be also expressed in terms of a value of a linear operator acting on $\mathbf{X}$. The detailed comparison of the linear operators pertaining to R_n defined as the numerator in (2) and the DFA estimator may provide the background necessary to explain the behaviour of the considered estimators of the Hurst exponent.

3. Detrended fluctuation analysis

3.1. Variants of DFA estimator

The second method to assess the strength of long-range dependence is based on a different measure of fluctuations of partial sums $Y_i = \sum_{t=1}^{i} X_t$ (Peng et al., 1994). Namely, instead of measuring the maximal deviation in both directions of (Y_k) from the line joining $Y_0 = 0$ and Y_n one considers an average of squared vertical distances of (Y_i) from the LS line

$$\overline{SSE}_k = k^{-1} SSE_k = \frac{1}{k} \sum_{i=1}^{k} (Y_i - a_k i - b_k)^2,$$

where $a_k i + b_k$ is the LS line fitted to points (i, Y_i), $i = 1, \ldots, k$. Taqqu et al. (1995) proved that $\mathbb{E}\overline{SSE}_k \sim Ck^{2H}$ for the fGn process and to the best of our knowledge this remains the only theoretical result supporting the DFA analysis.

The DFA method consists in calculating $\overline{SSE}_k$ for all $[n/k]$ disjoint blocks of size k, averaging the outcomes to get $\overline{\overline{SSE}}_k = [n/k]^{-1} \sum_{\text{blocks}} \overline{SSE}_k$ and fitting the LS line to points $\left(\log k, \log\left(\overline{\overline{SSE}}_k^{1/2}\right)\right)$. We call this estimator $\hat{H}_{s,m}$, where a sequence s, m corresponds to an order in which the mean and the square root are calculated. Observe that one can reverse the order of averaging and taking a square root in the above procedure, namely one averages $\left(\overline{SSE}_k\right)^{1/2}$

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.

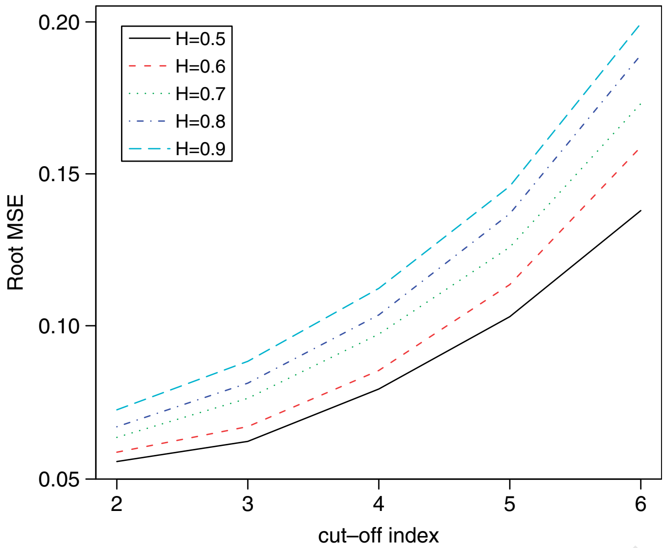

Fig. 6. Dependence of Root MSE for $\hat{H}_{m,s}$ on the cut-off point i_1 for fGn processes with $n = 512$.

over blocks of size k and treats the logarithm of the average as the response variable in regression. We call this estimator $\hat{H}_{m,s}$. This was proposed without explicitly being stated the deviation from the original DFA method by Weron (2001). Observe that in this case, similarly as for $\hat{H}_{d,a}$, a response variable in regression is a logged averaged characteristic of disjoint blocks. Below we provide some heuristic justification for $\hat{H}_{m,s}$.

We first present results of simulation experiments comparing $\hat{H}_{m,s}$ and $\hat{H}_{s,m}$ for different H and different sizes i_1 of the minimal block. It is empirically established that in contrast to R/S estimator for both DFA estimators, $RMSE$ is a decreasing function of the size of the minimal block, i.e. inclusion of smaller blocks in an estimation scheme leads to smaller values of $RMSE$. The same monotone pattern of behaviour is exhibited by the standard deviation of the estimators. Let us note that different dependence of RMSE on the cut-off point for the DFA and R/S estimators is due to the bias which in the second case is a nonnegligible part of RMSE.

Fig. 6 shows the dependence of RMSE on the minimal size of the block. Observe that the smallest dyadic size of the block is 4 as size 2 leads to 0 value of SSE. $RMSE\,(i_1)$ changes considerably when i_1 varies, e.g. its ratio for $i_1 = 6$ to $i_1 = 1$ is around 2.8 for $H = 0.5$ and 2.7 for $H = 0.9$ for the fGn traces of size $n = 512$. The estimator $\hat{H}_{s,m}$ performs equally comparable to $\hat{H}_{m,s}$ working slightly better for larger H in the case of the fGn processes. In the case of the FARIMA processes the last estimator yields smaller values of RMSE for all H but $H = 0.5$.

Consideration of overlap and pox-plot versions of $\hat{H}_{m,s}$ does not lead to substantial improvements, in particular the pox plot version of $\hat{H}_{m,s}$ based on disjoint blocks yields slightly smaller values of RMSE for the largest considered value of $H = 0.9$ only. In view of this $\hat{H}_{m,s}$ with $i_1 = 2$ was chosen as a representative of DFA estimators and will be called $\hat{H}_{dfa}$ from now on.

3.2. *Heuristic justification of* $\hat{H}_{m,s}$

We will argue that the property $\mathbb{E}\left(\overline{SSE}_k^{1/2}\right) \sim Ck^H$ holds for the fGn process. Denote by C a generic positive constant the exact value of which may change from line to line. Observe that Taylor expansion yields

$$\left(\overline{SSE}_k\right)^{1/2} - \left(\mathbb{E}\overline{SSE}_k\right)^{1/2} = \frac{1}{2\left(\mathbb{E}\overline{SSE}_k\right)^{1/2}}\left(\overline{SSE}_k - \mathbb{E}\overline{SSE}_k\right) - \frac{1}{4\left(\mathbb{E}\overline{SSE}_k\right)^{3/2}}\left(\overline{SSE}_k - \mathbb{E}\overline{SSE}_k\right)^2 + \text{higher order terms.}$$

Thus disregarding higher order terms the property $\mathbb{E}\left(\overline{SSE}_k^{1/2}\right) \sim Ck^H$ will follow from $\mathrm{Var}\overline{SSE}_k \sim Ck^{4H}$ as for the fGn process $\mathbb{E}\overline{SSE}_k \sim Ck^{2H}$ (Taqqu et al., 1995). This by the same argument is implied by $\mathbb{E}\left(\overline{SSE}_k^2\right) \sim Ck^{4H}$. The last property is plausible in view of the following theorem.

Theorem 4. *For the fGn we have* $C_1 k^{4H} \leqslant \mathbb{E}\left(\overline{SSE}_k^2\right) \leqslant C_2 k^{4H}$ *for all* $k \in \mathbb{N}$ *and some positive constants* C_1, C_2.

Proof. The left hand side inequality follows from $\mathbb{E}\left(\overline{SSE}_k^2\right) \geqslant \left(\mathbb{E}\overline{SSE}_k\right)^2 \sim Ck^{4H}$. For the second inequality observe that $\overline{SSE}_k \leqslant W_k := k^{-1}\sum_{t=1}^k Y_i^2$. Moreover, $\mathbb{E}\left(W_k^2\right) = \mathrm{Var}\left(W_k\right) + \left(\mathbb{E}W_k\right)^2$ and both terms in the last expression have order k^{4H}. Indeed, noting that Y_k is the fractional Brownian motion, from the Mercer equality it follows that with $\sigma^2 = \mathrm{Var}\left(Y_1\right)$

$$\begin{aligned}\mathrm{Var}\left(W_k\right) &= \frac{1}{k^2}\sum_{1\leqslant s,t\leqslant k}\mathrm{Cov}\left(Y_s^2, Y_t^2\right) = \frac{1}{k^2}\sum_{1\leqslant s,t\leqslant k} 2\,\mathrm{Cov}^2\left(Y_s, Y_t\right) \\ &= \frac{\sigma^2}{k^2}\sum_{1\leqslant s,t\leqslant k}\left(|s|^{2H} + |t|^{2H} - |t-s|^{2H}\right)^2 \sim Ck^{4H}.\end{aligned}$$

Moreover,

$$\mathbb{E}W_k = \frac{1}{k}\sum_{t=1}^{k}\mathbb{E}Y_i^2 = \frac{1}{k}\sum_{t=1}^{k} i^{2H} \sim \frac{1}{2H+1}k^{2H}. \qquad \square$$

4. Wavelet estimator of H

We briefly describe Abry–Veitch estimator of H and refer to Veitch and Abry (1999) for a detailed treatment. As before, let $X_1, \ldots, X_n$ denote an observable part of a trace of a discrete long-range dependent process such that its covariance function satisfies $\gamma(k) \sim c_\gamma k^{2H-2}$. We assume additionally that $(X_t)_{t=1}^{\infty}$ is a Gaussian process. Denote by $\hat{d}_{ik}$ a wavelet coefficient pertaining to a wavelet ψ at resolution i and location k calculated through a pyramidal algorithm initiated by scaling coefficients at level 0, $a_{o,k} = I * X(k)$, where $*$ denotes a discrete convolution and I is a filter used to account for a discrete structure of X_k (see Veitch et al., 2000). Let $\hat{\mu}_i = n_i^{-1}\sum_{k=1}^{n_i}\hat{d}_{i,k}^2$, where n_i is a number of wavelet coefficients at resolution i which can be calculated from available data without extrapolating past or future values. It follows from Veitch and Abry (1999) that $\log_2\hat{\mu}_i$ is approximately distributed as

$$\log_2\left(\frac{Z_{n_i}}{n_i}\right) + i(2H-1) + \log_2 C, \tag{3}$$

where Z_{n_i} is χ^2-distributed with n_i degrees of freedom and C is an absolute constant depending on ψ and c_γ. Thus accounting for the expected value and the variance of Z_{n_i}/n_i we obtain the following heteroscedastic regression model

$$\log_2\hat{\mu}_i + \left(n_i \ln 2\right)^{-1} = i(2H-1) + \log_2 C + \varepsilon_i,$$

where $\mathbb{E}\varepsilon_i \approx 0$ and $\mathrm{Var}\left(\varepsilon_i\right) \approx \left(2n_i \ln^2 2\right)^{-1}$. In order to obtain a wavelet estimator $\hat{H}_d$ of H a WLS regression line is fitted to points $\left(i, \log_2\hat{\mu}_i + \left(n_i \ln 2\right)^{-1}\right)$ for $i = i_1, \ldots, i_2$ with weights $w_i = n_i$. Then $\hat{H}_d$ is defined as $\hat{H}_d = \left(1 + \hat{\gamma}_d\right)/2$, where $\hat{\gamma}_d$ is a slope of the fitted WLS line. It turns out that a judicious choice of the smallest octave i_1 is crucial for performance of $\hat{H}_d$ as relation (3) holds asymptotically for large j only. In experiments discussed below we use as a benchmark estimator $\hat{H}_d$ with i_1 chosen optimally for a given family of processes (fGn or FARIMA) and $i_2 = \log_2 n - 4$. Optimal values of i_1 depend on the process as well as on a sample size and are equal to 2 or 3 in the case of the fGn and to 1 or 2 in the case of the FARIMA process in the considered range of n. The Daubechies wavelet D(7) was employed in the simulations. For more discussion on $\hat{H}_d$ and a parallel estimator based on scaling coefficients see Mielniczuk and Wojdyłło (2005).

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.

Table 2
RMSE of Hurst exponent estimators for 2^N observations of fGn with prescribed H

	$H=0.5$	$H=0.6$	$H=0.7$	$H=0.8$	$H=0.9$
$N=9$, $\widehat{H}_{R/S}$	0.0426*	0.0505*	0.0494*	0.0470*	0.0405*
$N=9$, $\widehat{H}_{dfa}$	0.0553	0.0584	0.0633	0.0668	0.0723
$N=9$, $\widehat{H}_{d}$	0.0980	0.1006	0.1045	0.1041	0.0992
$N=10$, $\widehat{H}_{R/S}$	0.0294*	0.0337*	0.0387*	0.0356*	0.0281*
$N=10$, $\widehat{H}_{dfa}$	0.0473	0.0492	0.0510	0.0516	0.0574
$N=10$, $\widehat{H}_{d}$	0.0555	0.0574	0.0561	0.0612	0.0591
$N=11$, $\widehat{H}_{R/S}$	0.0184*	0.0234*	0.0300*	0.0292*	0.0209*
$N=11$, $\widehat{H}_{dfa}$	0.0388	0.0387	0.0440	0.0451	0.0498
$N=11$, $\widehat{H}_{d}$	0.0321	0.035	0.0361	0.0395	0.0399
$N=12$, $\widehat{H}_{R/S}$	0.0136*	0.0175*	0.0243*	0.0235*	0.0166*
$N=12$, $\widehat{H}_{dfa}$	0.0318	0.0343	0.0379	0.0413	0.0413
$N=12$, $\widehat{H}_{d}$	0.0195	0.022	0.0251	0.0267	0.0298
$N=13$, $\widehat{H}_{R/S}$	0.0114*	0.0122*	0.0201	0.0207*	0.0211*
$N=13$, $\widehat{H}_{dfa}$	0.0280	0.0300	0.0324	0.0351	0.0363
$N=13$, $\widehat{H}_{d}$	0.0129	0.0154	0.0187	0.0217	0.0226
$N=14$, $\widehat{H}_{R/S}$	0.0103	0.0099*	0.0172	0.0195	0.0110*
$N=14$, $\widehat{H}_{dfa}$	0.0247	0.0264	0.0287	0.0297	0.0324
$N=14$, $\widehat{H}_{d}$	0.0088	0.0125	0.0152	0.0181	0.0198
$N=15$, $\widehat{H}_{R/S}$	0.0104	0.0076*	0.0164	0.0183	0.0103*
$N=15$, $\widehat{H}_{dfa}$	0.0208	0.0238	0.0234	0.0255	0.0282
$N=15$, $\widehat{H}_{d}$	0.0063	0.0094	0.0136	0.0165	0.0176

5. Comparison of performance of estimators

5.1. Simulation results

In Tables 2 and 3 values of RMSE for $\hat{H}_{R/S}$, $\hat{H}_{dfa}$ and $\hat{H}_d$ are displayed. $\hat{H}_{R/S}$ stands for the bias-corrected $\hat{H}_{d,p}$ estimator with $i_1=1$, $\hat{H}_{dfa}$ denotes the DFA estimator $\hat{H}_{m,s}$ with $i_1=2$ and the choice of the minimal block length of the wavelet estimator $\hat{H}_d$ is described in the previous section. It turns out that contrary to a common belief R/S estimator, after suitable correction, is the strong competitor to other estimators of the Hurst exponent. Namely, it has consistently lower RMSE than $\hat{H}_{dfa}$ for both considered processes and $\log_2 n =: N = 9, 10, \ldots, 15$. Compared with the wavelet estimator $\hat{H}_d$, $\hat{H}_{R/S}$ outperforms it in the case of the fGn process for $9 \leqslant N \leqslant 13$ (excluding $N=13$, $H=0.7$) and for $9 \leqslant N \leqslant 12$ in the case of the FARIMA process excluding $N=11, H \geqslant 0.7$.

The cases when RMSE of $\hat{H}_{R/S}$ is the smallest are indicated by the asterisk. It should be stressed $\hat{H}_d$ is not even an estimator in the strict sense as it involves truncation at the optimal unknown minimal octave i_1 (for a recent proposal of choosing i_1 see Veitch et al., 2003). When $\hat{H}_{R/S}$ is compared to $\hat{H}_d$ with $i_1=1$ for the fGn with $N=12$ the ratio $RMSE\left(\hat{H}_d\right) \Big/ RMSE\left(\hat{H}_{R/S}\right)$ equals 2.4, 3.3, 2.2, 2.5, 3.8 for $H=0.5, 0.6, \ldots, 0.9$, respectively.

We also compared the behaviour of $\hat{H}_{R/S}$ with that of the Whittle estimator $\hat{H}_w$ defined as the maximizer of an approximate version of a loglikelihood function in a case when a parametric form of an underlying long-range dependent process is known. A version of an algorithm to calculate $\hat{H}_w$ described in Beran (1994, pp. 223–233) was implemented. Table 4 lists ratios $R_1 = RMSE\left(\hat{H}_{R/S}\right) \Big/ RMSE\left(\hat{H}_w\right)$ for considered values of H and n in a case of the FARIMA process when a parametric model is correctly specified and analogously defined ratios R_2 when the observed fGn process is treated as the FARIMA process.

The Whittle estimator performs very well when the assumed parametric model is correct and, remarkably, its RMSE changes little with a change of H (by less than 10% for all considered values of H and n when compared with its RMSE for $H=0.5$). The bias-corrected R/S estimator performs worse in overall than the Whittle estimator but performance

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.

Table 3
RMSE of Hurst exponent estimators for 2^N observations of FARIMA process with prescribed H

	$H=0.5$	$H=0.6$	$H=0.7$	$H=0.8$	$H=0.9$
$N=9$, $\widehat{H}_{R/S}$	0.0444*	0.0394*	0.0382*	0.0405*	0.0461*
$N=9$, $\widehat{H}_{dfa}$	0.0566	0.0554	0.0611	0.0642	0.0733
$N=9$, $\widehat{H}_d$	0.0712	0.0602	0.0578	0.0566	0.0556
$N=10$, $\widehat{H}_{R/S}$	0.0302*	0.0280*	0.0283*	0.0295*	0.0338*
$N=10$, $\widehat{H}_{dfa}$	0.0469	0.0463	0.0542	0.0513	0.0601
$N=10$, $\widehat{H}_d$	0.0506	0.0416	0.0352	0.0328	0.0355
$N=11$, $\widehat{H}_{R/S}$	0.0198*	0.0195*	0.0224	0.0245	0.0355
$N=11$, $\widehat{H}_{dfa}$	0.0380	0.0371	0.0386	0.0456	0.0490
$N=11$, $\widehat{H}_d$	0.0321	0.0298	0.0222	0.0213	0.0291
$N=12$, $\widehat{H}_{R/S}$	0.0135*	0.0158*	0.0179*	0.0226*	0.0336*
$N=12$, $\widehat{H}_{dfa}$	0.0292	0.0302	0.0352	0.0343	0.0415
$N=12$, $\widehat{H}_d$	0.0195	0.0189	0.0221	0.0289	0.0359
$N=13$, $\widehat{H}_{R/S}$	0.0113*	0.0142	0.0174	0.0222	0.0299
$N=13$, $\widehat{H}_{dfa}$	0.0251	0.0247	0.0296	0.0346	0.0391
$N=13$, $\widehat{H}_d$	0.0129	0.0138	0.0117	0.0104	0.0159
$N=14$, $\widehat{H}_{R/S}$	0.0100	0.0136	0.0168	0.0213	0.0323
$N=14$, $\widehat{H}_{dfa}$	0.0249	0.0254	0.0265	0.0300	0.0338
$N=14$, $\widehat{H}_d$	0.0086	0.009	0.0092	0.0072	0.0014
$N=15$, $\widehat{H}_{R/S}$	0.0108	0.0135	0.0161	0.0213	0.0322
$N=15$, $\widehat{H}_{dfa}$	0.0217	0.0208	0.0259	0.0251	0.0290
$N=15$, $\widehat{H}_d$	0.0067	0.0066	0.0078	0.0057	0.0094

Table 4
Ratios of RMSE $\left(\hat{H}_{R/S}\right)$ /RMSE $\left(\hat{H}_w\right)$ for FARIMA and fGn with prescribed H

	$H=0.5$	$H=0.6$	$H=0.7$	$H=0.8$	$H=0.9$
$N=9$, R_1	1.1619	1.0448	0.9708	1.1056	1.2311
$N=9$, R_2	1.1242	1.3009	0.9693	0.6636	0.4844
$N=10$, R_1	1.1317	1.0391	1.0997	1.1162	1.5040
$N=10$, R_2	1.1261	1.0975	0.8230	0.5399	0.3499
$N=11$, R_1	1.1908	1.0673	1.2467	1.3616	1.9456
$N=11$, R_2	1.0687	0.8502	0.6616	0.4645	0.2640
$N=12$, R_1	1.0576	1.2598	1.4626	1.9033	2.8896
$N=12$, R_2	1.1027	0.6843	0.5537	0.3855	0.2113
$N=13$, R_1	1.3712	1.6047	1.9424	2.5548	3.4335
$N=13$, R_2	1.3301	0.5230	0.4655	0.3366	0.2682
$N=14$, R_1	1.6464	2.1229	2.7303	3.5972	5.0960
$N=14$, R_2	1.6564	0.4305	0.3996	0.3163	0.1410
$N=15$, R_1	2.4530	2.9786	3.7121	5.0984	7.5705
$N=15$, R_2	2.4823	0.3322	0.3844	0.2993	0.1326

of both estimators is comparable for $n \leqslant 2^{12}$ and $H \leqslant 0.7$. However, when the parametric model is misspecified and fGn traces are taken for FARIMA traces the performance of the Whittle estimator, indicated by ratios R_2, is visibly inferior to that of $\hat{H}_{R/S}$, especially for larger n and H. Simulations also indicate that the same observation holds true for a maximum likelihood estimator of the Hurst exponent.

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.

Table 5
Mean computing times in seconds for $\hat{H}_{R/S}$, $\hat{H}_{dfa}$ and $\hat{H}_d$ when $H = 0.7$

	$N = 9$	$N = 10$	$N = 11$	$N = 12$	$N = 13$	$N = 14$	$N = 15$
$\hat{H}_{R/S}$	0.0465	0.0943	0.1764	0.3611	0.7421	1.5776	3.3671
$\hat{H}_{dfa}$	0.9259	1.9053	3.7704	7.8050	16.0158	33.7666	74.0151
$\hat{H}_d$	0.0139	0.0172	0.0247	0.0359	0.0517	0.0849	0.1323

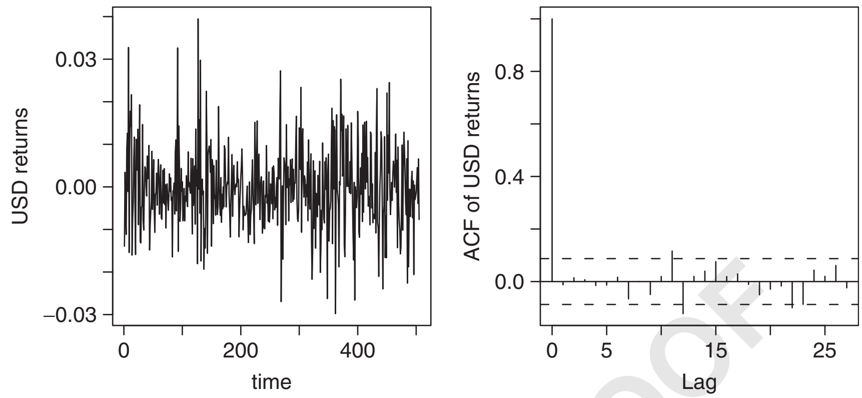

Fig. 7. Returns of USD exchange rate for the period 2002–2003 and corresponding ACF.

The behaviour of $\hat{H}_{R/S}$ can also be compared with the local Whittle estimator LWE when the number of Fourier components, which is a parameter of the method, is chosen by a direct approach described in Henry and Robinson (1996). Comparison of the results listed in Table 2 for $N = 10$ with Table 1.6 in their paper shows that for the fGn process the RMSE of $\hat{H}_{R/S}$ is slightly smaller than RMSE of the LW estimator, which equals 0.030, 0.034, 0.047, 0.056 and 0.067 for $H = 0.5(0.1)0.9$.

Mean computing times for considered semiparametric estimators are shown in Table 5. As they vary little with a change of the Hurst exponent we report them only for an intermediate value of $H = 0.7$. Computations were performed on PC with Pentium 4, 3.2 GHz and 1 GB RAM. The wavelet estimator is the least time consuming on average and the DFA estimator the most. The DFA estimator requires around 20 times more computing time than R/S estimator for all considered sample sizes.

A possible way to improve the bias correction of $\hat{H}_{R/S}$ estimator would be to choose an intercept of a reference line depending on a sample size. On the other hand, the possible drawback of the presented bias correction procedure is that the reference line is chosen with respect to combined families of FARIMA and fGn processes and at this point it is unclear how the method will work for other strongly dependent processes which are distant, in some sense, from the considered template processes. Let us finally note that the construction of Monte Carlo confidence intervals for H based on the bias-corrected $\hat{H}_{d,p}$ is possible by adapting a method described in Hall et al. (2001). The approach is used in the example discussed below.

5.2. *An illustrative example*

As an example we consider logarithmic returns $\log_2 (P_t/P_{t-1})$ of a daily exchange rate P_t of dollar, Swiss frank and euro to Polish złoty listed daily by the National Bank of Poland for a period January, 2, 2002–December, 31, 2003 consisting of $n = 506$ observations. Fig. 7 depicts a trace of logarithmic returns of the dollar exchange rate for that period and corresponding values of an empirical autocorrelation function (ACF) for lags $0, 1, \ldots, 27$. The maximal considered lag equals $\lfloor 10 \log_{10} n \rfloor$ and coincides with the default value used e.g. in S-PLUS. As 3 out of 27 ACFs fall outside $\pm 1.96/\sqrt{n}$ strip, a usual rule-of-thumb procedure suggests that the logarithmic returns are dependent.

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.

Table 6
Estimators of H and confidence intervals for logarithmic returns of exchange rate

Currency	$\hat{H}^0_{R/S}$	$\hat{H}_{R/S}$	$\hat{H}_{dfa}$	$\hat{H}_d$
USD	0.6219	0.5818	0.5188	0.5347
	(0.6105, 0.7392)	(0.5131, 0.6675)	(0.4663, 0.6643)	(0.3050, 0.7603)
CHF	0.5641	0.5148	0.4981	0.3164
	(0.5793, 0.7084)	(0.4653, 0.6085)	(0.4609, 0.6674)	(0.0844, 0.5310)
EURO	0.5687	0.4688	0.5153	0.3289
	(0.5847, 0.7032)	(0.4150, 0.5639)	(0.4566, 0.6412)	(0.0790, 0.5744)

Estimators of the Hurst coefficient together with pertaining Monte Carlo confidence intervals are listed in Table 6. $\hat{H}^0_{R/S}$ stands for the uncorrected $\hat{H}_{d,p}$ with $i_1 = 3$ and the minimal octave for $\hat{H}_d$ was 2. For the calculated value of $\hat{H}_{R/S} = 0.5818$, 500 traces of the FARIMA $\left(0,\ \hat{d},\ 0\right)$ process of length 506 were generated with $\hat{d} = \hat{H}_{R/S} - \frac{1}{2}$ and estimators $\hat{H}^1_{R/S}, \ldots, \hat{H}^{500}_{R/S}$ were computed based on them. The reported confidence interval is a Monte Carlo 95% percentile interval pertaining to these values. Other confidence intervals were calculated analogously.

Assuming, as it is indicated by simulations, that $\hat{H}_{R/S}$ performs more adequately for this sample size ($n = 506$) than $\hat{H}_{dfa}$ and $\hat{H}_d$, the fact that the confidence interval based on $\hat{H}_{R/S}$ lies to the right of the point 0.5 in case of the dollar returns suggests existence of weak long-range dependence. In contrast, results for Swiss frank and euro indicate short-range dependence. At the same time, the results for the uncorrected $\hat{H}^0_{R/S}$ would suggest long-range dependence for all considered currencies.

6. Conclusions

The goal of our paper was to construct a bias correction to the R/S estimator of the Hurst parameter H. The proposed method is based on an approximate linearity (as a function of H) of the bias of one of the possible variants of $\hat{H}_{R/S}$, namely $\hat{H}_{d,p}$ defined in Section 2.1. A specific form of the employed bias correction is justified by Theorem 2 and Corollary 1. It turns out that the bias-corrected R/S estimator performs much better than its uncorrected versions and it compares favourably with other semiparametric estimators of Hurst parameter considered in the paper, namely the DFA and the wavelet estimators. In particular, it has a lower RMSE than the DFA estimator for the FARIMA and the fGn processes with $2^9 \leqslant n \leqslant 2^{15}$ and $H = 0.5, 0.6, \ldots, 0.9$. Also, its performance is comparable with performance of the wavelet method.

As expected, performance of the bias-corrected R/S estimator is worse than that of the Whittle estimator when a parametric model is correctly specified. However, if a parametric form of a spectral density of a process is misspecified and e.g. the fGn process is observed instead of the assumed FARIMA process with the same value of H, the gain of R/S estimator over the Whittle estimator are significant with RMSE ratio exceeding 2.5 for $N \geqslant 12$ and $H \geqslant 0.8$ (cf. Table 4).

Moreover, we have discussed a fact that various scaling properties on which semiparametric estimators of H are usually based allow for natural variations in construction of specific estimators which can differ e.g. in the way a sample is divided into blocks, a size of the minimal block, etc. We have considered such different constructions in the case of R/S estimator and show that the variants differ greatly in performance. Also, in the case of the DFA we have given theoretical justification for one of the variants of the estimators used in practice.

7. Uncited references

Hurvich and Ray (2003); Teverovsky et al. (1999).

References

Andrews, D.W., Sun, Y., 2004. Adaptive local polynomial Whittle estimation of long-range dependence. Econometrica 72, 569–614.
Anis, A.A., Lloyd, E.H., 1976. The expected value of the adjusted rescaled Hurst range of independent normal summands. Biometrika 63, 111–116.

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.

Arteche, J., 2004. Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models. J. Econometrics 119, 31–154.
Beran, J., 1994. Statistics for Long Memory Processes. Chapman & Hall, New York.
Davies, R.B., Harte, D.S., 1987. Tests for Hurst effect. Biometrika 74, 95–101.
Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series model. J. Time Ser. Anal. 4, 221–228.
Giraitis, L., Kokoszka, P., Leipus, R., Teyssiere, G., 2003. Rescaled variance and related tests for long memory in volatility and levels. J. Econometrics 112, 2265–2294.
Hall, P., Härdle, W., Kleinow, T., Schmidt, P., 2001. Semiparametric bootstrap approach to confidence intervals for the Hurst coefficient. Statist. Inference for Stochastic Processes 3, 263–276.
Henry, M., Robinson, P.M., 1996. Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In: Robinson, P.M., Rosenblatt, M. (Eds.), Athens Conference on Applied Probability and Time Series, vol. 2, 220–232.
Hurst, H.R., 1951. Long-term storage in reservoirs. Trans. Amer. Soc. Civil Eng. 116, 770–799.
Hurvich, C.M., Ray, B.K., 2003. The local Whittle estimator of long-memory stochastic volatility. J. Financial Econometrics 1, 445–470.
Hurvich, C.M., Moulines, E., Soulier, P., 2005. Estimating long memory in volatility. Econometrica 73, 1283–1328.
Kennedy, D., 1976. The distribution of the maximum Brownian excursion. J. Appl. Probab. 13, 371–376.
Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? J. Econometrics 54, 159–178.
Lai, D., 2004. Estimating the Hurst effect and its applications in monitoring clinical trials. Comput. Statist. Data Anal. 45, 549–562.
Lo, A.W., 1991. Long-memory in stock-market prices. Econometrica 59, 1279–1313.
Mandelbrot, B., 1975. Limit theorems of the self-normalized range for weakly and strongly dependent processes. Z. Wahrscheinlichkeitstheorie und verwandene Gebiete 31, 271–285.
Meyer, I., 1993. Wavelets, Algorithms and Applications. SIAM, Philadelphia.
Mielniczuk, J., Wojdyłło, P., 2005. Wavelets for time series data—a survey and new results. Control & Cybernet. 34, 1093–1126.
Peng, C.K., Buldyrev, S.V., Simons, M., Stanley, H.E., Goldberger, A.L., 1994. Mosaic organization of DNA nucleotides. Phys. Rev. E 49, 1685–1689.
Percival, D.B., Walden, A.T., 2000. Wavelet Methods for Time Series Analysis. Cambridge University Press, Cambridge.
Peters, E.E., 1994. Fractal Market Analysis. Wiley, New York.
Robinson, P.M., 1994. Time series with strong dependence. In: Sims, C.A. (Ed.), Advances of Econometrics. Sixth World Congress. Cambridge University Press, Cambridge, pp. 147–195.
Robinson, P.M., 1995. Gaussian semiparametric estimation of time series with long-range dependence. Ann. Statist. 23, 1630–1661.
Stoev, S., Taqqu, M., Park, C., Michailidis, G., Marron, J.S., 2006. LASS: a tool for the local analysis of self-similarity. Comput. Statist. Data Anal. 50, 2447–2471.
Taqqu, M., Teverovsky, V., Willinger, W., 1995. Estimators for long-range dependence: empirical study. Fractals 3, 785–798.
Teverovsky, V., Taqqu, M., Willinger, W., 1999. A critical look at Lo's modified R/S statistic. J. Statist. Plann. Inference 80, 211–227.
Veitch, D., Abry, P., 1999. A wavelet-based joint estimator of the parameters of long-range dependence. IEEE Trans. Inform. Theory 45, 878–897.
Veitch, D., Taqqu, M.S., Abry, P., 2000. Meaningful MRA initialisation for discrete time series. Signal Process. 80, 1971–1983.
Veitch, D., Taqqu, M.S., Abry, P., 2003. On automatic selection of the onset of scaling. Fractals 11, 377–390.
Weron, R., 2001. Estimating long-range dependence: finite sample properties and confidence intervals. Phys. A 31, 285–299.
Whittle, P., 1953. Estimation and information in stationary time series. Ark. Mat. 2, 423–434.

Please cite this article as: J. Mielniczuk, P. Wojdyłło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), doi: 10.1016/j.csda.2006.07.033.