

SFB
823

Estimation methods for the LRD parameter under a change in the mean

Aeneas Rooch, Ieva Zelo, Roland Fried

Nr. 32/2016

Discussion Paper



Estimation methods for the LRD parameter under a change in the mean

Aeneas Rooch · Ieva Zelo · Roland Fried

Received: date / Accepted: date

Abstract When analyzing time series which are supposed to exhibit long-range dependence (LRD), a basic issue is the estimation of the LRD parameter, for example the Hurst parameter $H \in (1/2, 1)$. Conventional estimators of H easily lead to spurious detection of long memory if the time series includes a shift in the mean. This defect has fatal consequences in change-point problems: Tests for a level shift rely on H , which needs to be estimated before, but this estimation is distorted by the level shift.

We investigate two blocks approaches to adapt estimators of H to the case that the time series includes a jump and compare them with other natural techniques as well as with estimators based on the trimming idea via simulations. These techniques improve the estimation of H if there is indeed a change in the mean. In the absence of such a change, the methods little affect the usual estimation. As adaption, we recommend an overlapping blocks approach: If one uses a consistent estimator, the adaption will preserve this property and it performs well in simulations.

Keywords Hurst parameter; estimation; jump; long-range dependence; long memory; change-point problems

1 Introduction

Long-range dependence is an issue in several fields of time series analysis, ranging from climate sciences [45] and network data traffic [13], [31], [27] to economics and finance. In the latter context, volatilities may exhibit long memory [6], and there are discussions whether long-range dependence can be found in stock market prices [54], [7], [2], [9], [32]; for a survey on long-range dependence in economics, see [1].

Aeneas Rooch

Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany

E-mail: aeneas.rooch@rub.de

Ieva Zelo

Technische Universität Dortmund, Fakultät für Statistik

E-mail: zelo@statistik.tu-dortmund.de

Roland Fried

Technische Universität Dortmund, Fakultät für Statistik

E-mail: fried@statistik.tu-dortmund.de

To make ideas precise, we consider a stochastic process $(X_i)_{i \geq 1}$ of the type

$$X_i = \mu_i + \epsilon_i, \quad (1)$$

where $(\mu_i)_{i \geq 1}$ is a sequence of unknown means and

$$\epsilon_i = G(\xi_i), \quad i \geq 1, \quad (2)$$

is a functional of a stationary Gaussian process $(\xi_i)_{i \geq 1}$ with $E[\xi_i] = 0$ and $E[\xi_i^2] = 1$ and autocovariance function

$$\gamma(k) = k^{-D} L(k), \quad k \geq 1, \quad (3)$$

where $0 < D < 1$ and $L(k)$ is a slowly varying function. $G : \mathbb{R} \rightarrow \mathbb{R}$ is a measurable function satisfying $E[G(\xi_i)] = 0$. The process $(\xi_i)_{i \geq 1}$ exhibits long-range dependence (LRD) due to the non-summability of the autocovariances, and depending on the transformation G , it passes the LRD on to the process $(X_i)_{i \geq 1}$. That means, we consider a stochastic process $(X_i)_{i \geq 1}$ which is an instantaneous functional of a stationary Gaussian LRD process, which has unknown means and may exhibit LRD.

The exponent $D \in (0, 1)$ in (3) determines the intensity of the LRD. It is more common to use the fractional differencing parameter $d = (1 - D)/2 \in (0, 1/2)$ or the Hurst parameter $H = 1 - D/2 \in (1/2, 1)$ in the context of LRD and stationary time series. The higher H is, the slower is the decay of the autocovariance and the stronger is the dependence of the random variables. The limiting value $H = 1/2$ implies short range dependence.

There are various approaches to estimate the Hurst parameter. A parametric Whittle approach involving the approximate log-Likelihood was introduced by [53]. Based on this estimation technique [30] proposed the semiparametric local Whittle estimator, while consistency and normality of this estimator was shown by [39]. Prefiltering and tapering were considered by [42] where they studied the performance of the local Whittle estimator with and without both preprocessing procedures. Consistency and the distribution of the local Whittle estimator in the nonstationary case were considered by [48] and [49] involving data tapers. Dependence of the convergence rate and the limit distribution on H was introduced in [42]. Cases in which the local Whittle estimator is not consistent are discussed in [36]. Another semiparametric estimator of the Hurst parameter based on the log-periodogram regression was proposed by [14]. In contrast to the local Whittle estimator it has a closed form. While [14] used all frequencies and the corresponding periodogram values in the formula of the estimator, [40] introduced a trimmed version of this estimation procedure by truncating the lowest and the highest frequencies to improve the performance of the estimator.

If one wants to test whether a sample X_1, \dots, X_n includes a change-point, i.e., a point at which the structure of the data changes (examples of such changes are given below), knowledge about H is essential since it is involved in the scaling of test statistics. So is the function L which appears in (3), but we do not concentrate on this issue in this article.

Change-point problems for LRD time series have been in the focus of research for many years. They may concern a possible change in the mean; here, one wants to test the null hypothesis

$$H_0 : \mu_1 = \dots = \mu_n \quad (4)$$

of a constant mean against the alternative

$$A : \mu_1 = \dots = \mu_k \neq \mu_{k+1} = \dots = \mu_n \text{ for some } k \in \{1, \dots, n-1\} \quad (5)$$

that there is a change-point at which the level shifts. We will refer to this test problem as (H_0, A) . CUSUM-type tests for the test problem (H_0, A) have been studied e.g. by [20], [28] or [50]. In [51] a test procedure based on Wilcoxon rank statistics is analysed, while [10] proposed a test which is based on the Wilcoxon two-sample test statistic:

$$W_{k,n} := \frac{1}{n d_n} \sum_{i=1}^k \sum_{j=k+1}^n \left(1_{\{X_i \leq X_j\}} - \frac{1}{2} \right) \quad (6)$$

where

$$d_n^2 \sim c_r n^{2-r(2-2H)} L^r(n) \quad (7)$$

with the Hermite rank $r \in \mathbb{N}$ of the class of functions $\{1_{G(\xi_t) \leq x} - F(x), x \in \mathbb{R}\}$ and $c_r = [2r!/(1 - (2 - 2H)r)(2 - (2 - 2H)r)]$ being the right scaling for $W_{k,n}$ to have a non-degenerate limit distribution under H_0 if the X_i have a continuous distribution function F . The proposed change-point test for the test problem (H_0, A) rejects the null hypothesis H_0 of no change-point for large values of the test statistic

$$W_n := \max_{1 \leq k \leq n} |W_{k,n}|. \quad (8)$$

Under H_0 [10] proved that W_n with $r = 1$ converges in distribution to

$$\frac{1}{2\sqrt{\pi}} \sup_{0 \leq \nu \leq 1} |Z_1(\nu) - \nu Z_1(1)|,$$

where $(Z_1(\nu))_{\nu \geq 0}$ denotes the standard fractional Brownian motion with Hurst parameter H if the observations are strictly monotone functions of a stationary Gaussian process $(\xi_i)_{i \geq 1}$ with mean zero, variance 1 and autocovariance function as in (3). Upper quantiles of this asymptotic distribution can be used as critical values for the test. For details and critical values, see [10].

Since the scaling function d_n^2 in (7) depends on the Hurst parameter H which is not known in practice, one has to use a suitable consistent estimator \hat{H} of H to apply the test. The modified test statistic is then defined as follows:

$$\widetilde{W}_n = \sqrt{\frac{n^{2-r(2-2H)}}{n^{2-r(2-2\hat{H})}}} \max_{1 \leq k \leq n} |W_{k,n}| = n^{r(H-\hat{H})} \max_{1 \leq k \leq n} |W_{k,n}| \quad (9)$$

Taking the logarithm of \widetilde{W}_n yields

$$\log \widetilde{W}_n = r(H - \hat{H}) \cdot \log n + \log \max_{1 \leq k \leq n} |W_{k,n}| \quad (10)$$

Combining the continuous mapping theorem with Slutsky's theorem it is obvious from (10) that \widetilde{W}_n has the same limit distribution as W_n as long as \hat{H} has a convergence rate faster than $\log n$. This is e.g. satisfied by the Geweke and Porter-Hudak (GPH) and the local Whittle (LW) estimator, see Section 5 in [4] or [22].

Other issues under a structural change occur in the context of testing for long range dependence, see e.g. [12] where the confusion of long memory with structural changes is stated. This theory is supported by [17] who consider the relation between the presence of long memory and the number of detected level shifts by the Schwarz or Bayesian Information criterion: spurious change-points may be found if long range dependence is not taken into account.

In this article, we will treat the situation where a change in the mean occurs. Note that there are other change-point problems: changes in the marginal distribution of the observations [16], in the coefficients of linear regression models where the errors are LRD [29] or in the dependence structure [19], [37]. All these methods have in common that they require knowledge of the LRD parameter of the data, be it D as in

(3), the Hurst exponent $H = 1 - D/2$ or other related quantities to describe the intensity of LRD. For an overview about change-point analysis, see the book of [8] or the recent article by [25].

For an overview on estimators of the LRD parameter see [46] or [47]. Unfortunately, when one estimates H in order to apply a change-point test, one may have to deal with a time series that includes a jump, and such a jump intrinsically gets in the way of estimating the LRD parameter, since the usual estimators do not consider a change-point and likely misspecify the structure of the data as a more intense LRD. As a consequence, the power of change-point tests decreases, because the jump may cause an overestimation of H which may bring the test to interpret the jump as LRD behavior. In [44] it is reported that tests for a break in persistence, i.e., a break in the long range-dependence structure, are not robust against a shift in the mean and must be adapted in this case. For an overview about how structural breaks and trends lead to misspecification of LRD and about methods for distinguishing both effects, see the survey of [43], who also points out that change-point estimators cannot distinguish between LRD and change-points.

It is thus a challenging problem to distinguish between behavior that originates from long memory and behavior that originates from change-points. CUSUM-type tests which discriminate between an LRD time series and a short-range dependent time series with changes in the mean have been proposed by [5] and by [41], which both contain a lot of references concerning LRD and structural breaks. We focus to some extent on the situation where a LRD time series may possess one change-point, but also consider situations with multiple breaks in our simulations. In this situation, especially for change-point tests which usually require the true, but unknown LRD parameter, there is a strong need for methods which estimate the LRD parameter without being affected strongly by a level shift. Our goal is thus to develop estimators for H which are on the one hand not heavily biased by a jump in the mean of the data, and which on the other hand do not lose much efficiency if there is no jump. To achieve this, [21] proposed usage of a local Whittle estimator which requires the knowledge of the number of change-points. Such a method was applied to German stock market data by [18], whereas [33] used a trimmed version of the Geweke and Porter-Hudak estimator to remove the effect of the change in the mean by trimming the lower frequencies in the spectral domain. The estimation method of [24] is also based on trimming the lowest frequencies, applied to the local Whittle estimator, originally proposed by [30]. We will restrict ourselves on the latter two approaches, since the method proposed by [21] requires knowledge about the number of level shifts.

We investigate a broader approach based on the idea that estimation of the LRD parameter in the presence of a jump should not use the whole time series, but should first segregate the jump or concentrate on local environments in the sample. This general idea can be combined with any usual estimation technique. We examine the overlapping and the non-overlapping blocks approach along with two further variants of this idea for comparison and verify that they improve the respective standard estimators which ignore jumps. We compare our proposals with the adapted approaches of [33] and [24].

The rest of this paper is organized as follows: In Section 2 we present our adaption techniques together with the methods of [33] and [24] for estimating H under a jump in the mean. In Section 3 we prove the consistency of the blocks estimators, and in Section 4, we analyse the performance of our methods in a large set of Monte Carlo simulations. In Section 5, we apply the estimation methods to real data. In Section 6, we give a summary and an outlook.

2 Adaption techniques

As noted before, conventional estimators of H tend to be positively biased by level shifts. We will investigate several general techniques to adapt estimators for H to time series with a jump in the mean. These techniques can be applied in combination with any of the generic estimators reviewed in Subsection 2.1.

2.1 Generic estimation techniques

In what follows we will involve the two following generic techniques in the estimation of the Hurst parameter.

The regression based GPH estimator, which was first introduced in [14], uses the spectral representation of the long memory processes. The regression is based on the representation

$$f(\lambda) = L(\lambda)\lambda^{2H-1} \quad (11)$$

of the spectral density $f(\lambda)$ as the frequency λ decreases to zero, and the corresponding periodogram $I(\lambda_j)$ at frequency $\lambda_j = 2\pi j/n$, $j = 1, \dots, n$. The GPH estimator is defined as

$$\hat{H}_{\text{GPH}} = -0.5 \sum_{j=1}^b (Y_j - \bar{Y}) \log I(\lambda_j) / \sum_{j=1}^b (Y_j - \bar{Y})^2 + 0.5, \quad (12)$$

where $Y_j = \log |1 - \exp(-i\lambda_j)|$, $\bar{Y} = b^{-1} \sum_{j=1}^b Y_j$ and $b = [n^u]$, $u > 0$, is the bandwidth parameter which satisfies $b \log(b)/n = o(1)$ for the consistency of the parameter estimator (see [22]).

The local Whittle (LW) estimator \hat{H}_{LW} proposed in [30] minimizes the following approximate frequency domain likelihood function:

$$R(H) = \log \left(\frac{1}{b-l+1} \sum_{j=1}^b \lambda_j^{2H-1} I(\lambda_j) \right) - (2H-1) \frac{1}{b} \sum_{j=1}^b \log(\lambda_j) \quad (13)$$

where the relationship

$$f(\lambda) = G\lambda^{2H-1} \quad (14)$$

as the frequency λ decreases to zero with $G \in (0, \infty)$ is assumed.

2.2 Pre-estimating the jump

The most intuitive approach is removal of a jump before application of any of the previous methods. For this purpose, we estimate the position and the height of a jump, ignoring that we do not know the LRD parameter H yet, and then either eliminate the jump and estimate H on the whole (now) approximately jump-free time series, or we estimate H on the observations before and after the jump separately and take the average as an estimate for H on the overall sample.

We get down to details: Given the observations X_1, \dots, X_n , we apply a change-point test. In principle, we may take any change-point test we like. Here, we choose the Wilcoxon change-point test from [10] (see section 1) which rejects the null hypothesis H_0 (that there is no change in the mean) for large values of the test statistic

$$W_n = \max_{1 \leq k \leq n} |W_{k,n}|,$$

where

$$W_{k,n} := \frac{1}{n d_n} \sum_{i=1}^k \sum_{j=k+1}^n \left(1_{\{X_i \leq X_j\}} - \frac{1}{2} \right). \quad (15)$$

For more details see [10]. The change-point is supposed to take place after the observation X_{k^*} where $k^* \in \{1, \dots, n-1\}$ is the value of k for which $|W_{k,n}|$, as defined in (15), takes its maximum value.

Note that in W_n , H is only included in the scaling factor $(n d_n)^{-1}$. Thus, knowledge about the LRD parameter H is only essential for a test decision *if* there is a change-point; the possible location k^* is unaffected. This allows us to obtain a candidate for the jump location without knowing H as

$$k^* = \arg \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^k \sum_{j=k+1}^n \left(1_{\{X_i \leq X_j\}} - \frac{1}{2} \right) \right|. \quad (16)$$

The estimator k^*/n is a weakly consistent estimator for the true jump location $\tau \in (0, 1)$ as can be proven using the arguments of [20]. If there is a jump after $X_{[\tau n]}$ for a $\tau \in (0, 1)$, k^* estimates its location $[\tau n]$; if there is no jump in the time series, k^* is some more or less meaningless index in $\{1, \dots, n\}$, but it will not affect the following estimation procedure.

Given the observations X_1, \dots, X_n and an estimate k^* for the jump location as in (16), we divide the observations into two parts

$$X_1, \dots, X_{k^*} \quad \text{and} \quad X_{k^*+1}, \dots, X_n.$$

The jump height can be estimated by

$$\hat{h} = \frac{1}{n - k^*} (X_{k^*+1} + \dots + X_n) - \frac{1}{k^*} (X_1 + \dots + X_{k^*}),$$

and so we can remove the most plausible jump by considering the time series

$$X_1, \dots, X_{k^*}, X_{k^*+1} - \hat{h}, \dots, X_n - \hat{h}.$$

On this new time series, which is regarded as jump-free, we can estimate H ; we denote this estimate, which is based on pre-estimating and one long time series, by $\hat{H}_{\text{pre},1}$.

Alternatively, we can estimate H on both arising blocks separately, which yields two estimates $\hat{H}_{k^*}^{(1)}$ and $\hat{H}_{k^*}^{(2)}$, and take their average as an estimate for H on the whole sample:

$$\hat{H}_{\text{pre},2} = \frac{\hat{H}_{k^*}^{(1)} + \hat{H}_{k^*}^{(2)}}{2}.$$

Remark 1 (i) Other change-point tests which can be used to pre-detect the jump (instead of the Wilcoxon test based on (16) above) may need the LRD parameter H – which is just our aim when applying the change-point tests here. A way out of this vicious circle could be an iterative procedure: Starting with an estimate of H on the whole sample, one could apply the change-point test. This yields, by the just described pre-estimating, a new estimate for H , which can then be used again to execute the change-point test, and so on. One can cherish hopes that such a procedure converges or yields, when stopped by a certain rule, a useful estimate for H . This bears, of course, the risk that the initial estimation of H is so heavily affected by the jump that the jump is not detected.

(ii) Such an iterative procedure cannot be used to refine the estimator $\hat{H}_{\text{pre},1}$: The removal of the jump is error prone since both the location and the height of the jump are estimated, so after removing the jump, one could again apply the change-point test, remove the resulting jump and estimate H on this new (double)

jump-free sample, and so on; one could stop the iterations when the estimated value of H does not change significantly (for example when the difference between the new estimate and the estimate from the previous iteration is ≤ 0.01). But this does not lead to a useful estimate since the technique removes too much of the structure of the data and H is heavily underestimated.

(iii) It is not possible to refine $\hat{H}_{\text{pre},2}$ with such an iterative procedure: As described, the estimation of the jump location by k^* is not affected by the value of H – we would obtain the same estimate $\hat{H}_{\text{pre},2}$ in any iteration step.

2.3 Estimation from two blocks

Instead of application of a change-point test statistic for splitting the sample into two subsets, we can consider any splitting of the data into two blocks of reasonable size,

$$X_1, \dots, X_k \quad \text{and} \quad X_{k+1}, \dots, X_n$$

and estimate H on each block separately. This is done for all possible cutting points k . Because the estimation of LRD intrinsically needs many observations, it is not sensible if one allows early or late cutting points which produce a small block on which the estimation of H is useless. We consider cutting points k in the set

$$K = \{k_{\text{low}}, k_{\text{low}} + 1, \dots, k_{\text{up}}\}$$

with

$$k_{\text{low}} = \max\{\lceil n/10 \rceil, 10\} + 1$$

$$k_{\text{up}} = n - k_{\text{low}}$$

leading to $|K| = k_{\text{up}} - k_{\text{low}} + 1$ pairs of estimations $(\hat{H}_k^{(1)}, \hat{H}_k^{(2)})$ of H , derived from the first and the second block.

As an estimate for the Hurst parameter of the whole sample, we consider two functions of $\hat{H}_k^{(1)}, \hat{H}_k^{(2)}, k \in K$: the mean value of the estimate on the first and the second block, averaged over all cutting points,

$$\hat{H}_{\text{mean}} = \frac{1}{|K|} \sum_{k \in K} \frac{\hat{H}_k^{(1)} + \hat{H}_k^{(2)}}{2}$$

and the mean value of the estimate on the first and the second block

$$\hat{H}_{\text{mindiff}} = \frac{\hat{H}_{k^*}^{(1)} + \hat{H}_{k^*}^{(2)}}{2}$$

divided by that cutting point k^* where both estimates differ least:

$$k^* = \arg \min_{k \in K} \left| \hat{H}_k^{(2)} - \hat{H}_k^{(1)} \right|.$$

2.4 Estimation from many blocks

The techniques presented in the previous subsections rely on the existence of only one jump and finding an adequate splitting into two blocks. Our main proposal extends this idea and splits the data into many blocks

$$X_{c-w}, X_{c-w+1}, \dots, X_c, \dots, X_{c+w}$$

of length $\bar{w} = 2w + 1$ around the center X_c through the time series X_1, \dots, X_n . For each $c \in M = \{w + 1, w + 2, \dots, n - w\}$, we estimate H on the window around the center X_c , leading to $|M| = n - 2w$ estimates $\check{H}_{\bar{w},c}$ for H . Since estimates $\check{H}_{\bar{w},c}$, $c \in M$, derived from moderately large windows will have high variability, we average all estimations and use

$$\hat{H}_{O,\bar{w}} = \frac{1}{|M|} \sum_{m \in M} \check{H}_{\bar{w},m} \quad (17)$$

to estimate the Hurst parameter H .

As usually, the choice of the length of the block or window is an important issue. We choose the flank length $w = w(n)$ (which results in a window of size $\bar{w} = 2w + 1$) as a function of the overall sample size n ,

$$w = w(n) = \max \{ \lfloor \sqrt{n} \rfloor, 10 \} \quad (\bar{w} = 2w + 1),$$

since the square root of n has proven to be a good choice in many contexts and since we do not want the window to be too small as the estimation of an LRD parameter gets bad for a short sample of observations. For the reason of comparison, we also analyse w as the total window size.

Whenever data are separated into blocks, it is much discussed if the blocks should be overlapping or non-overlapping. In our context, the latter approach has the advantage that each jump affects only a single block and thus results only in one most likely erroneous estimate. So it is interesting (and a contribution to the above discussion) to analyse also separation into non-overlapping blocks

$$X_1, \dots, X_{\bar{w}}, \quad X_{\bar{w}+1}, \dots, X_{2\bar{w}}, \quad X_{2\bar{w}+1}, \dots, X_{3\bar{w}}, \quad \dots \quad X_{([n/\bar{w}]-1)\bar{w}+1}, \dots, X_{[n/\bar{w}]\bar{w}}$$

of length $\bar{w} = 2w + 1$, resulting in $[n/\bar{w}]$ estimates $\check{H}_{\bar{w},k}$, $k = 1, \dots, [n/\bar{w}]$, for H . Again it seems reasonable to average and consider

$$\hat{H}_{NO,\bar{w}} = \frac{1}{[n/\bar{w}]} \sum_{k=1}^{[n/\bar{w}]} \check{H}_{\bar{w},k} \quad (18)$$

as final estimate for the Hurst parameter H . As block sizes, it is plausible to consider the same choices w and \bar{w} as above, since \sqrt{n} is widely used in the context of moving windows.

2.5 Trimmed estimators

A trimmed version of the regression based GPH estimator was investigated by [33] where the lowest frequencies are trimmed, as these are affected by changes in the mean. Let $I(\lambda_j)$ be the value of the periodogram at frequency $\lambda_j = 2\pi j/n$. The trimmed GPH estimator is

$$\hat{H}_{\text{GPH}}^{\text{tr}} = -0.5 \sum_{j=l}^b (Y_j - \bar{Y}) \log I(\lambda_j) / \sum_{j=l}^b (Y_j - \bar{Y})^2 + 0.5,$$

where $Y_j = \log |1 - \exp(-i\lambda_j)|$ and $\bar{Y} = (b - l + 1)^{-1} \sum_{j=l}^b Y_k$. The trimming parameter is denoted by $l = \lceil n^{1/2+\epsilon} \rceil$, $\epsilon > 0$, while $b = \lceil n^u \rceil$, $u > 0$, is the bandwidth parameter. Moreover, the adapted GPH estimation technique was proposed, where the trimming parameter is computed recursively. Let $l_i = \lceil n^{(2-2\hat{H}_i)/(3-2\hat{H}_i+\epsilon)} \rceil$, where \hat{H}_1 is computed using $l_1 = \lceil n^{1/2+\epsilon} \rceil$. The value \hat{H}_i denotes the estimate computed using the trimming parameter l_i . Then the Hurst parameter is estimated by \hat{H}_i if $|\hat{H}_i - \hat{H}_{i-1}| < 0.01$ or $i > 9$. The authors suggest values $\epsilon = 0.05$ and $u = 0.8$. We will use the adapted trimmed estimator for comparison in this paper.

Similarly, [24] proposed the trimmed version of the local Whittle estimator, where the computation of the estimate is also based only on the higher frequencies – this shall take into account a change in the mean. The trimmed local Whittle estimator \hat{H}_{LW}^{tr} minimizes the function

$$R(H) = \log \left(\frac{1}{b-l+1} \sum_{j=l}^b \lambda_j^{2H-1} I(\lambda_j) \right) - (2H-1) \frac{1}{b-l+1} \sum_{j=l}^b \log(\lambda_j),$$

where l is the trimming parameter and b is the bandwidth parameter. Suggested values are $l = \lceil 1 + 0.2n^{0.62} \rceil$ and $b = \lceil 0.8n^{0.79} \rceil$.

3 Inheriting asymptotic consistency

In this section we investigate conditions for the asymptotic consistency of our main proposal to split the data into many blocks and estimate H separately. It turns out that the estimators of H constructed using overlapping or non-overlapping blocks inherit asymptotic consistency in the absence or presence of one or even several jumps, as is demonstrated in the following. This is e.g. satisfied by the GPH-estimator under appropriate assumptions (see remark 2, part (i)).

Theorem 1 *Let $(\xi_i)_{i \geq 1}$ be a stationary Gaussian LRD process, with mean 0, variance 1 and autocovariances satisfying (3) and consider observations X_1, \dots, X_n of the type $X_i = G(\xi_i) + \mu_i$ for a measurable function $G : \mathbb{R} \rightarrow \mathbb{R}$ satisfying $E[G(\xi_i)] = 0$ and $\mu_i \in \mathbb{R}$. Consider the null hypothesis*

$$H_0 : \mu_1 = \dots = \mu_n.$$

and the alternative that at least one change in the mean occurs. Consider the overlapping blocks estimator (17) with a monotone increasing window length $\bar{w} = o(n)$. As generic estimation method $\check{H}_{\bar{w},j}$ on the j -th window choose any estimator \hat{H} for H which is asymptotically consistent under H_0 . Under the modified alternative A (see (5)) where we are confronted with $m = o(n/\bar{w})$ structural breaks we assume that the variance in the perturbed block grows sufficiently slowly with the sample size. Let \tilde{K} denote the index set of the windows which include a jump, and that on these windows

$$\text{Var} \left(\check{H}_{\bar{w},\tilde{k}} \right) = o \left(\frac{n^2}{(m\bar{w})^2} \right), \quad (19)$$

$$E \left(\check{H}_{\bar{w},\tilde{k}} - H \right) = o \left(\frac{n}{m\bar{w}} \right), \quad \tilde{k} \in \tilde{K}. \quad (20)$$

Then the non-overlapping blocks estimator (18) is asymptotically unbiased and MSE-consistent.

Proof Due to the stationarity of the X_i under the null hypothesis H_0 we have

$$E\left(\hat{H}_{NO,\bar{w}} - H\right) = \frac{1}{n - \bar{w} + 1} \sum_{j=1}^{n-\bar{w}+1} E(\check{H}_{\bar{w},j} - H) = \frac{n - \bar{w} + 1}{n - \bar{w} + 1} E(\check{H}_{\bar{w},1} - H) \rightarrow 0.$$

In the presence of level shifts we obtain by (20)

$$\begin{aligned} E\left(\hat{H}_{NO,\bar{w}} - H\right) &= \frac{1}{n - \bar{w} + 1} \sum_{j=1, j \notin \tilde{K}}^{n-\bar{w}+1} E(\check{H}_{\bar{w},j} - H) + \frac{1}{n - \bar{w} + 1} \sum_{\tilde{k} \in \tilde{K}} E(\check{H}_{\bar{w},\tilde{k}} - H) \\ &= \frac{n - \bar{w} + 1 - m\bar{w}}{n - \bar{w} + 1} E(\check{H}_{\bar{w},1} - H) + \frac{m\bar{w}}{n - \bar{w} + 1} o\left(\frac{n}{m\bar{w}}\right) \rightarrow 0. \end{aligned} \quad (21)$$

This proves the asymptotic unbiasedness. Moreover, we obtain for the variance

$$\begin{aligned} Var(\hat{H}_{NO,\bar{w}}) &= \frac{1}{(n - \bar{w} + 1)^2} \left(\sum_{k=1}^{n-\bar{w}+1} Var(\check{H}_{\bar{w},k}) + \sum_{k=1, j \neq k}^{n-\bar{w}+1} Cov(\check{H}_{\bar{w},k}, \check{H}_{\bar{w},j}) \right) \\ &\leq \frac{1}{(n - \bar{w} + 1)^2} \left(\sum_{k=1}^{n-\bar{w}+1} Var(\check{H}_{\bar{w},k}) + \sum_{k=1, j \neq k}^{n-\bar{w}+1} \sqrt{Var(\check{H}_{\bar{w},k})} \sqrt{Var(\check{H}_{\bar{w},j})} \right) \\ &= \frac{m\bar{w}}{(n - \bar{w} + 1)^2} Var(\check{H}_{\bar{w},\tilde{k}}) \end{aligned} \quad (22)$$

$$+ \frac{n - \bar{w} + 1 - m\bar{w}}{(n - \bar{w} + 1)^2} Var(\check{H}_{\bar{w},1}) \quad (23)$$

$$+ \frac{2m\bar{w}(n - \bar{w} + 1 - m\bar{w})}{(n - \bar{w} + 1)^2} \sqrt{Var(\check{H}_{\bar{w},\tilde{k}})} \sqrt{Var(\check{H}_{\bar{w},1})} \quad (24)$$

$$+ \frac{m\bar{w}(m\bar{w} - 1)}{(n - \bar{w} + 1)^2} Var(\check{H}_{\bar{w},\tilde{k}}) \quad (25)$$

$$+ \frac{(n - \bar{w} + 1 - m\bar{w})(n - \bar{w} - m\bar{w})}{(n - \bar{w} + 1)^2} Var(\check{H}_{\bar{w},1}) \quad (26)$$

$$\rightarrow 0$$

□

Remark 2 (i) We assume that we are dealing with $m\bar{w}$ jump-contaminated overlapping windows which is the maximum number and therefore represents the worst case scenario. The terms in (23) and (26) tend to zero as $n \rightarrow \infty$, since we are dealing with a consistent estimator of the Hurst parameter.

The terms (22), (24) and (25) represent the decomposition of the covariance part where it has to be ensured that the growth rate of $Var(\check{H}_{\bar{w},\tilde{k}})$ is not faster than the decay rate of the remaining factors. This is e.g. satisfied by the ordinary GPH estimator if the window length \bar{w} , bandwidth $b = [\bar{w}^u]$ and the number of jumps m satisfy

$$\frac{\bar{w}^2}{nb} = O(1), \quad (27)$$

$$\frac{\log^3(b)m\bar{w}}{n} \left(m + \frac{\bar{w}^{2-2H}}{b} \right) = o(1). \quad (28)$$

Condition (27) requires that the window length does not grow faster than the number of blocks. To be on the safe side we assume that we are dealing with a sample contaminated by structural changes so the interest in considering only few windows does not arise. Condition (28) guarantees that the number of level shifts grows sufficiently slowly. Moreover, this condition ensures that the test statistic \widetilde{W}_n in (9) has the same limiting distribution as W_n .

Proof of Remark 2 (i) The variance of the ordinary GPH estimator is given by [22]:

$$\text{Var}(\check{H}_{\bar{w}, \bar{k}}) = \frac{1}{4S_{YY}^2} \sum_{j=1}^b a_j^2 \text{Var}(e_j) + \frac{1}{2S_{YY}^2} \sum_{k=1}^b \sum_{j=k+1}^b a_k a_j \text{Cov}(e_k, e_j) \quad (29)$$

where $e_j = \log(I(\lambda_j)/f(\lambda_j)) + C$, C is Euler's constant, $a_j = Y_j - \bar{Y}$ and $S_{YY} = \sum_{k=1}^b a_k^2$. We will use the approximations of a_j and S_{YY} by [23]:

$$a_j = O(\log(b)) \quad (30)$$

$$S_{YY} = b + o(b) \quad (31)$$

and the approximation of $\text{Var}(e_j)$ by [33]:

$$\text{Var}(e_j) = \frac{\pi^2}{6} + O\left(\frac{\log(j)}{j}\right) + O\left(\frac{\bar{w}^{2-2H}}{j^{3-2H}}\right) \quad (32)$$

to express the variance $\text{Var}(\check{H}_{\bar{w}, \bar{k}})$ of the GPH estimator in a perturbed block in terms of the Landau notation:

$$\begin{aligned} \text{Var}(\check{H}_{\bar{w}, \bar{k}}) &\leq \frac{1}{4S_{YY}^2} \sum_{j=1}^b a_j^2 \text{Var}(e_j) + \frac{1}{2S_{YY}^2} \sum_{k=1}^b \sum_{j=k+1}^b |a_k a_j| \sqrt{\text{Var}(e_j) \text{Var}(e_k)} \\ &= \frac{1}{4S_{YY}^2} \sum_{j=1}^b a_j^2 \text{Var}(e_j) + \frac{1}{2S_{YY}^2} \sum_{k=1}^b \sum_{j=k+1}^b |a_k a_j| O(\text{Var}(e_k)) \end{aligned} \quad (33)$$

$$= \frac{1}{4(b + o(b))^2} \left(\frac{\pi^2 b}{6} + o(b) + O(b \log^2(b)) + O(\log^2(b) \bar{w}^{2-2H}) \right) \quad (34)$$

$$+ \frac{1}{2(b + o(b))^2} O(\log^2(b)) \left(O\left(\frac{b^2 \pi^2}{6}\right) + O(b^2) + O(b \bar{w}^{2-2H}) \right). \quad (35)$$

In (33) we used the fact that the terms $\log(j)/j$ and \bar{w}^{2-2H}/j^{3-2H} are monotonously decreasing in j . Obviously the dominating terms of $\text{Var}(\check{H}_{\bar{w}, \bar{k}})$ are $\log^2(b) b^2 / (b + o(b))^2$ and $\log^2(b) \bar{w}^{2-2H} b / (b + o(b))^2$. Considering (22), (25) and the square of (24) together with the relationship $\text{Var}(\check{H}_{\bar{w}, 1}) = O(\pi^2 / (24b) + o(1/b))$ (see e.g. [22]) we observe that the following term is dominating:

$$\begin{aligned} \text{Var}(\check{H}_{\bar{w}, \bar{k}}) \frac{(m\bar{w})^2}{(n - \bar{w} + 1)^2} &= \log^2(b) \frac{b^2 + \bar{w}^{2-2H} b}{(b + o(b))^2} \frac{(m\bar{w})^2}{(n - \bar{w} + 1)^2} \\ &= \frac{\log^2(b) m^2 \bar{w}}{n - \bar{w} + 1} \frac{(b^2 + \bar{w}^{2-2H} b) \bar{w}}{(b + o(b))^2 (n - \bar{w} + 1)}. \end{aligned} \quad (36)$$

Combining (36) with conditions (27) and (28) we conclude that the term (25) tends to zero. It is easy to see that the remaining terms (22) and (26) also converge to zero.

The asymptotic unbiasedness (21) of the non-overlapping blocks estimator follows from condition (28) and the following relationship (see [33]):

$$E(\check{H}_{\bar{w}, \bar{k}} - H) = O\left(\frac{b^2}{\bar{w}^2}\right) + O\left(\frac{\log^3(b)}{b}\right) + O\left(\frac{\bar{w}^{2-2H} \log^2(b)}{b}\right). \quad (37)$$

□

Thus the bias and the variance of the averaged estimator $\hat{H}_{O,\bar{w}}$ tend to zero regardless of whether there is a level shift or not.

- (ii) The proof for the above Theorem does not use the specific window length $[\sqrt{n}]$. It remains valid for any monotone sequence which is increasing in n and satisfies condition (27).
- (iii) The consistency of the non-overlapping blocks estimator from Subsection 2.4 is proven analogously.
- (iv) The derivation of the distribution of the blocks estimators involves asymptotic theory of triangular arrays in the case of LRD processes and is beyond the scope of this paper.

4 Simulations

In this section we analyse the performance of the adaption techniques described in Section 2.

4.1 Simulation scenarios

As generic (not adapted) estimation methods for the Hurst parameter H we choose the local Whittle estimator and the GPH estimator in order to compare the results with the trimmed counterparts. We simulate fGn time series X_1, \dots, X_n with $H = 0.5, 0.7, 0.9$ of length $n = 500, 1000, 1500$, by generating ξ_1, \dots, ξ_n from fGn with the corresponding Hurst parameter H , using the `fArma` package in R, and choosing G as the identity $G(t) = t$ or the quantile transform

$$G(t) = (3/4)^{-(1/2)} \left((\Phi(t))^{-1/3} - 3/2 \right),$$

where Φ denotes the standard Gaussian c.d.f., so that the observations are Pareto(3,1) distributed. This corresponds to two extreme cases: Well-behaved Gaussian data on the one hand and heavy tailed data on the other hand.

We add a jump of height $h = 0.5, 1, 2$ after a proportion of $\tau = 0.1, 0.5$ of the data (i.e., after $0.1 \cdot n$ or after $0.5 \cdot n$ observations) to these time series. To the resulting seven types of time series (one without a jump, six with a jump at different positions and of different heights) for each n we apply the adapted estimators. For comparison, we also apply the two estimation techniques proposed by [33] and [24], $\hat{H}_{\text{GPH}}^{\text{tr}}$ and $\hat{H}_{\text{LW}}^{\text{tr}}$ described in Section 2.5, to the generated time series.

We repeat these simulations 1000 times for each scenario, leading to 1000 estimates $\tilde{H}_1, \dots, \tilde{H}_{1000}$ for each combination of the generic estimation method, adaption technique \tilde{H} and data situation. We use the root mean-squared error (RMSE) in order to compare the quality of the estimates. RMSE is widely used as evaluation criterion also for the memory parameter (see e.g. the simulation studies of [33] and [24]) as it takes bias and variance of the estimators into account, while being on the same scale as the parameter to be estimated.

For the case of multiple breaks we will add three jumps after proportions $\tau = 0.1, 0.5, 0.75$. Moreover, we simulate from a fractionally integrated ARMA model, i.e., ARFIMA(p, d, q) model

$$(1 - \varphi B)x_t = (1 - \theta B)\varepsilon_t \tag{38}$$

with $p = 0, q = 1$ (i.e. ARFIMA(0, $d, 1$)) and $\theta = -0.6$, using the `arfima` package in R and following the simulation study of [33], where $d = H - 0.5$ is the differencing parameter. Finally, we consider the short

range dependent series from the ARMA model. Since fractional Gaussian noise and ARFIMA(0, d , 0) are equivalent models (see e.g. [14]) the latter model is included implicitly in this simulation study.

4.2 Simulation results for $H = 0.7$

We will discuss in detail the results for $H = 0.7$ in the following and comment on the results for $H = 0.5$ and $H = 0.9$ thereafter.

Tables 1 and 2 report the RMSE when applying our adaption techniques with the local Whittle estimator. We see that there is no uniformly best estimator; instead we observe the following:

- For fGn time series
 - if there is no jump, all estimation procedures yield similar results.
 - the overlapping blocks technique $\hat{H}_{O,\bar{w}}$ with block length $\bar{w} = 2\sqrt{n} + 1$ performs better than the other estimators when dealing with a small number of observations and low jumps.
 - the overlapping blocks estimator $\hat{H}_{O,w}$ with window length $w = \sqrt{n}$ yields the best results in all other cases, followed by the non-overlapping blocks estimator $\hat{H}_{NO,w}$. The performance of the blocks estimators depends on the sample size rather than the jump height.
- For the Pareto(3,1)-transformed fGn
 - with the estimation on two blocks \hat{H}_{mean} we obtain the best results in the case of low jumps in the mean.
 - the overlapping blocks estimator $\hat{H}_{O,\bar{w}}$ dominates if the jump in the mean is high and in the jump-free scenario.
 - most of the estimators yield better results when dealing with high jumps rather than low jumps. Figure 3 suggests that the Hurst parameter is being underestimated under the heavy-tailed distribution, which is compensated by the overestimation, originating from a high jump in the mean.

4.3 Simulation results for $H = 0.5$ and $H = 0.9$

For the case of $H = 0.5$ and $H = 0.9$ we report the results in the Figures 4 and 5 in the appendix. Here we restrict the graphical illustration of the boxplots to the case of fGn and $n = 1000$. The results are similar to the case $H = 0.7$, see Section 4.2. The overlapping blocks estimator $\hat{H}_{O,\bar{w}}$ and the non-overlapping blocks approach $\hat{H}_{NO,\bar{w}}$ exhibit small variability and yield rather small bias in the case of short memory, i.e., $H = 0.5$. The estimators $\hat{H}_{\text{pre},1}$ and $\hat{H}_{\text{pre},2}$ perform well for jumps in the middle. The trimmed version of the local Whittle estimator $\hat{H}_{\text{LW},\text{tr}}$ overestimates the parameter if the jump is high. Both $\hat{H}_{\text{LW},\text{tr}}$ and $\hat{H}_{\text{GPH},\text{tr}}$ exhibit rather high variability in comparison to the blocks estimators. For $H = 0.9$, i.e., very strong dependence, again the blocks and the pre-estimation approaches yield good results and outperform the trimmed versions of the local Whittle and the GPH estimators.

In the case of the GPH estimator as a generic estimation technique we present the boxplots of the estimates for $H = 0.7$ in Figure 6 in the appendix. We can observe that the pre-estimation techniques yield the least bias in most cases, while having rather high variability. When dealing with early and high jumps the overlapping blocks estimator $\hat{H}_{O,\bar{w}}$ exhibits the least bias and variability. In the case of $H = 0.5$ and $H = 0.9$ the corresponding figures show similar results and are omitted in this paper.

Table 1: Estimated RMSE of different estimators for the Hurst parameter $H = 0.7$ in time series without ($h = 0$) and with change-point (jump of height h after a proportion of τ of the data), each based on 1000 simulation runs with $n = 500$, $n = 1000$ and $n = 1500$ values from fGn, based on the local Whittle estimator. The best results for every scenario are emphasized in bold.

h	\hat{H}_{LW}	\hat{H}_{mean}	\hat{H}_{mindiff}	$\hat{H}_{O,w}$	$\hat{H}_{O,\bar{w}}$	$\hat{H}_{NO,w}$	$\hat{H}_{NO,\bar{w}}$	$\hat{H}_{\text{pre},1}$	$\hat{H}_{\text{pre},2}$	\hat{H}_{LW}^{tr}	$\hat{H}_{\text{GPH}}^{\text{tr}}$
$n = 500$											
0	0.0552	0.0513	0.0810	0.0488	0.0479	0.0560	0.0539	0.0620	0.0614	0.0846	0.1093
$\tau = 0.1$											
0.5	0.0592	0.0554	0.0912	0.0482	0.0477	0.0553	0.0542	0.0581	0.0616	0.0848	0.1094
1	0.0844	0.0731	0.1143	0.0469	0.0485	0.0539	0.0547	0.0579	0.0715	0.0865	0.1158
2	0.1720	0.1223	0.1922	0.0447	0.0518	0.0514	0.0566	0.0872	0.1068	0.0992	0.1410
$\tau = 0.5$											
0.5	0.0705	0.0544	0.0833	0.0484	0.0480	0.0556	0.0541	0.0597	0.0552	0.0853	0.1092
1	0.1269	0.0708	0.1011	0.0472	0.0487	0.0544	0.0558	0.0568	0.0531	0.0879	0.1131
2	0.2394	0.1181	0.1367	0.0453	0.0513	0.0523	0.0593	0.0564	0.0538	0.1033	0.1299
$n = 1000$											
0	0.0419	0.0394	0.0648	0.0321	0.0356	0.0363	0.0385	0.0446	0.0442	0.0625	0.0716
$\tau = 0.1$											
0.5	0.0473	0.0446	0.0697	0.0320	0.0360	0.0361	0.0392	0.0426	0.0476	0.0633	0.0729
1	0.0759	0.0647	0.1107	0.0318	0.0374	0.0359	0.0411	0.0455	0.0578	0.0659	0.0769
2	0.1609	0.1149	0.1938	0.0317	0.0409	0.0359	0.0450	0.0774	0.0967	0.0806	0.0947
$\tau = 0.5$											
0.5	0.0601	0.0439	0.0659	0.0321	0.0361	0.0363	0.0388	0.0437	0.0406	0.0635	0.0729
1	0.1159	0.0627	0.0832	0.0319	0.0374	0.0361	0.0401	0.0428	0.0403	0.0661	0.0761
2	0.2186	0.1087	0.1179	0.0318	0.0408	0.0360	0.0438	0.0424	0.0407	0.0806	0.0922
$n = 1500$											
0	0.0341	0.0326	0.0559	0.0259	0.0329	0.0300	0.0350	0.0357	0.0358	0.0538	0.0569
$\tau = 0.1$											
0.5	0.0409	0.0384	0.0601	0.0258	0.0335	0.0299	0.0352	0.0343	0.0392	0.0545	0.0580
1	0.0716	0.0597	0.1114	0.0259	0.0353	0.0300	0.0358	0.0395	0.0532	0.0562	0.0616
2	0.1542	0.1092	0.2026	0.0264	0.0389	0.0303	0.0381	0.0691	0.0892	0.0660	0.0757
$\tau = 0.5$											
0.5	0.0548	0.0384	0.0591	0.0259	0.0337	0.0300	0.0363	0.0345	0.0325	0.0544	0.0578
1	0.1101	0.0581	0.0749	0.0261	0.0354	0.0302	0.0386	0.0340	0.0328	0.0562	0.0610
2	0.2074	0.1027	0.1061	0.0265	0.0389	0.0308	0.0423	0.0340	0.0332	0.0660	0.0745

4.4 Multiple breaks

It is also interesting to investigate how the proposed methods perform in the case of multiple breaks. For this purpose we generate time series ξ_1, \dots, ξ_n from fGn with the Hurst parameter $H = 0.7$ choosing G as the identity $G(t) = t$. We add three jumps of equal height $h = 0.5, 1, 2$ after a proportion of $\tau = 0.1, 0.5, 0.75$ of the data (i.e., after $0.1 \cdot n$, $0.5 \cdot n$ and $0.75 \cdot n$ observations) to these time series. To the resulting three types of time series we apply the adapted estimators. For comparison we also apply the two adapted estimation techniques \hat{H}_{LW}^{tr} and $\hat{H}_{\text{GPH}}^{\text{tr}}$ proposed by [24] and [33] to the generated time series. We repeat these simulations 1000 times for each scenario, leading to 1000 estimates $\tilde{H}_1 \dots, \tilde{H}_{1000}$ for each estimator \tilde{H} and each data situation.

Table 2: Estimated RMSE of different estimators for the Hurst parameter $H = 0.7$ in time series without ($h = 0$) and with change-point (jump of height h after a proportion of τ of the data), each based on 1000 simulation runs with $n = 500$, $n = 1000$ and $n = 1500$ values from Pareto(3,1)-transformed fGn, based on the local Whittle estimator. The best results for every scenario are emphasized in bold.

h	\hat{H}_{LW}	\hat{H}_{mean}	\hat{H}_{mindiff}	$\hat{H}_{O,w}$	$\hat{H}_{O,\bar{w}}$	$\hat{H}_{NO,w}$	$\hat{H}_{NO,\bar{w}}$	$\hat{H}_{\text{pre},1}$	$\hat{H}_{\text{pre},2}$	\hat{H}_{LW}^{tr}	$\hat{H}_{\text{GPH}}^{\text{tr}}$
$n = 500$											
0	0.0818	0.0760	0.1063	0.0863	0.0750	0.0905	0.0792	0.1045	0.0924	0.1086	0.1476
$\tau = 0.1$											
0.5	0.0661	0.0600	0.1244	0.0828	0.0692	0.0868	0.0765	0.0904	0.0866	0.1054	0.1474
1	0.0831	0.0609	0.1548	0.0790	0.0617	0.0822	0.0706	0.0798	0.0894	0.0992	0.1458
2	0.1888	0.1000	0.2073	0.0755	0.0544	0.0781	0.0617	0.1074	0.1081	0.0992	0.1551
$\tau = 0.5$											
0.5	0.0650	0.0560	0.0875	0.0831	0.0694	0.0876	0.0715	0.0934	0.0794	0.1057	0.1471
1	0.1339	0.0527	0.0811	0.0795	0.0622	0.0834	0.0629	0.0920	0.0783	0.1000	0.1446
2	0.2501	0.0925	0.1134	0.0761	0.0551	0.0792	0.0564	0.0914	0.0781	0.1027	0.1491
$n = 1000$											
0	0.0721	0.0672	0.0913	0.0712	0.0642	0.0732	0.0673	0.0873	0.0783	0.0908	0.1139
$\tau = 0.1$											
0.5	0.0531	0.0498	0.1024	0.0688	0.0600	0.0710	0.0615	0.0755	0.0703	0.0881	0.1125
1	0.0670	0.0483	0.1581	0.0661	0.0543	0.0681	0.0547	0.0669	0.0704	0.0816	0.1094
2	0.1674	0.0920	0.2159	0.0634	0.0483	0.0649	0.0494	0.0889	0.0959	0.0747	0.1078
$\tau = 0.5$											
0.5	0.0508	0.0473	0.0678	0.0688	0.0601	0.0710	0.0630	0.0811	0.0707	0.0880	0.1124
1	0.1136	0.0417	0.0579	0.0659	0.0545	0.0680	0.0571	0.0801	0.0703	0.0814	0.1106
2	0.2271	0.0833	0.0932	0.0633	0.0488	0.0647	0.0510	0.0795	0.0699	0.0741	0.1081
$n = 1500$											
0	0.0681	0.0643	0.0801	0.0655	0.0564	0.0668	0.0579	0.0814	0.0735	0.0847	0.1012
$\tau = 0.1$											
0.5	0.0467	0.0448	0.0964	0.0636	0.0531	0.0658	0.0566	0.0692	0.0650	0.0829	0.1011
1	0.0582	0.0398	0.1593	0.0613	0.0485	0.0639	0.0538	0.0603	0.0639	0.0777	0.0988
2	0.1548	0.0839	0.2224	0.0591	0.0436	0.0611	0.0484	0.0777	0.0896	0.0649	0.0946
$\tau = 0.5$											
0.5	0.0430	0.0434	0.0529	0.0636	0.0532	0.0643	0.0536	0.0748	0.0665	0.0826	0.1009
1	0.1029	0.0342	0.0523	0.0614	0.0486	0.0614	0.0483	0.0742	0.0661	0.0771	0.0990
2	0.2117	0.0746	0.0804	0.0592	0.0440	0.0592	0.0441	0.0739	0.0657	0.0645	0.0942

Table 3 reports the estimated RMSE for fGn time series. In the case of three change-points, the overlapping and non-overlapping blocks estimators with window length \sqrt{n} perform best. The overlapping blocks estimator $\hat{H}_{O,w}$ outperforms the other estimation methods. This can be explained by the fact that only a few blocks are contaminated with the jumps in the mean, so the biased estimates of the corresponding blocks do not influence the resulting estimates dramatically. The other methods suffer from a high number of change points and are highly biased then. The methods proposed by [24] and [33] are outperformed by our adapted blocks techniques in every case. Taking into account the fast computation and the good performance of the overlapping blocks estimator it can be recommended in our survey.

Table 3: Estimated RMSE of different estimators for the Hurst parameter $H = 0.7$ in time series with three change-points (jumps of height h after a proportion of $\tau = 0.25, 0.5$ and 0.75 of the data), each based on 1000 simulation runs with $n = 500$ and $n = 1000$ values from fGn, based on the local Whittle estimator. The best results for every scenario are emphasized in bold.

h	\hat{H}_{LW}	\hat{H}_{mean}	\hat{H}_{mindiff}	$\hat{H}_{O,w}$	$\hat{H}_{O,\bar{w}}$	$\hat{H}_{NO,w}$	$\hat{H}_{NO,\bar{w}}$	$\hat{H}_{\text{pre},1}$	$\hat{H}_{\text{pre},2}$	$\hat{H}_{\text{LW}}^{\text{tr}}$	$\hat{H}_{\text{GPH}}^{\text{tr}}$
$n = 500$											
0.5	0.1277	0.0757	0.1022	0.0475	0.0484	0.0547	0.0548	0.0573	0.0595	0.0894	0.1162
1	0.2463	0.1484	0.1686	0.0444	0.0547	0.0520	0.0617	0.1201	0.1208	0.1147	0.1590
2	0.2999	0.2463	0.2184	0.0420	0.0755	0.0501	0.0838	0.2547	0.2444	0.2176	0.3258
$n = 1000$											
0.5	0.1177	0.0682	0.0854	0.0319	0.0374	0.0360	0.0410	0.0485	0.0503	0.0681	0.0789
1	0.2272	0.1386	0.1499	0.0319	0.0438	0.0361	0.0490	0.1149	0.1117	0.0922	0.1094
2	0.2999	0.2416	0.2338	0.0343	0.0600	0.0393	0.0678	0.2375	0.2259	0.1893	0.2391

4.5 ARFIMA model

As already mentioned in [33], the estimators of the differencing parameter d tend to be positively biased in the ARFIMA(p, d, q) model. We simulated ARFIMA(0, $d, 1$) time series with $\theta = -0.6$ for $d = 0.2$ and $d = 0$ of length $n = 1000$. Again, jumps of various heights were added at the beginning or in the middle of the data in order to compare the performance of the estimators in absence and presence of level shifts. In Table 4 we can observe that the blocks procedures fail to estimate the parameter correctly when the block size is small. The best results are obtained by $\hat{H}_{\text{pre},1}$. In the case of short memory, i.e. ARFIMA(0, 0, 1), again, the pre-estimating procedure $\hat{H}_{\text{pre},1}$ performs best in the short memory scenario, since it involves all data points in the estimation procedure.

The two blocks approaches perform similarly and do not yield desirable results for the ARFIMA model when a small block size is chosen. In Table 7 in the appendix we show the results corresponding to further block sizes of the non-overlapping blocks estimator (the results for the overlapping blocks approach are similar throughout the simulation study, though the 1000 simulation runs require more computation time). We observe that its performance can be improved by choosing a larger block size \bar{w} . Moreover, in the fGn-scenario, selecting a higher value for \bar{w} does not worsen the performance of $\hat{H}_{O,\bar{w}}$ compared to the case of the block length \sqrt{n} considerably, see Table 7. Still, the results are not satisfying when dealing with both, long and short memory.

Therefore we followed the ARMA correction procedure, proposed by [38], where the differencing parameter d in the ARFIMA(p, d, q) model is estimated recursively in several steps:

1. Obtain an estimate \hat{d} of d in the ARFIMA(p, d, q) model.
2. Calculate $\hat{Y}_t = (1 - B)^{\hat{d}} X_t$.
3. Identify and estimate the order and the parameters in the ARMA(p, q) model $\Phi(B)\hat{Y}_t = \Theta(B)\varepsilon_t$.
4. Calculate $\hat{Z}_t = \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)} X_t$.
5. Obtain new estimate \hat{d} of d in the ARFIMA(0, $d, 0$) model $(1 - B)^{\hat{d}} \hat{Z}_t = \varepsilon_t$.

6. Repeat steps 2 to 5, until the difference between two consecutive estimates of d is less than 0.01 or the number of steps exceeds 10.

The RMSE values obtained using the above procedure are given in Table 8. We observe that in the case of the non-overlapping blocks approach the results could be improved considerably. For $d = 0$, i.e. short memory, higher window values are preferred, while for LRD with $d = 0.2$ ($H = 0.7$) the block length \sqrt{n} yields the lowest RMSE. Moreover, the results are nearly the same when dealing with fGn, i.e. when no short memory components are involved. In the case of the overlapping blocks approach we show the results for $n = 1000$, $\lambda = 0.5$ and $h = 2$ at the end of Table 8. We can observe that even lower RMSE values can be achieved when choosing overlapping instead of non-overlapping blocks, hence it can be recommended in our survey. To be on the safe side, we suggest choosing a block length \bar{w} equal to a multiple of \sqrt{n} , preferably $\bar{w} > 100$ in order to obtain desirable results.

Table 4: Estimated RMSE of different estimators for the differencing parameters $d = 0.2$ ($H = 0.7$) and $d = 0$ ($H = 0.5$) in time series without ($h = 0$) and with change-point (jump of height h after a proportion of τ of the data), each based on 1000 simulation runs with $n = 1000$ values from the ARFIMA(0, d, 1) model, based on the local Whittle estimator. The best results for every scenario are emphasized in bold.

ARFIMA(0, 0.2, 1)											
h	\hat{H}_{LW}	\hat{H}_{mean}	\hat{H}_{mindiff}	$\hat{H}_{O,w}$	$\hat{H}_{O,\bar{w}}$	$\hat{H}_{NO,w}$	$\hat{H}_{NO,\bar{w}}$	$\hat{H}_{\text{pre},1}$	$\hat{H}_{\text{pre},2}$	\hat{H}_{LW}^{tr}	$\hat{H}_{\text{GPH}}^{\text{tr}}$
$n = 1000$											
0	0.0638	0.0856	0.0915	0.2150	0.1768	0.2171	0.1782	0.0537	0.0743	0.1029	0.1274
$\tau = 0.1$											
0.5	0.0698	0.0897	0.0930	0.2152	0.1761	0.2165	0.1779	0.0552	0.0799	0.1016	0.1279
1	0.0824	0.1025	0.1032	0.2151	0.1773	0.2150	0.1803	0.0595	0.0942	0.1064	0.1347
2	0.1319	0.1351	0.1393	0.2162	0.1807	0.2179	0.1834	0.0809	0.1210	0.1133	0.1570
$\tau = 0.5$											
0.5	0.0767	0.0870	0.0960	0.2157	0.1770	0.2161	0.1764	0.0519	0.0751	0.1025	0.1297
1	0.1056	0.1016	0.1170	0.2162	0.1779	0.2150	0.1787	0.0558	0.0730	0.1053	0.1366
2	0.1753	0.1326	0.1554	0.2166	0.1820	0.2176	0.1821	0.0570	0.0762	0.1140	0.1530
ARFIMA(0, 0, 1)											
h	\hat{H}_{LW}	\hat{H}_{mean}	\hat{H}_{mindiff}	$\hat{H}_{O,w}$	$\hat{H}_{O,\bar{w}}$	$\hat{H}_{NO,w}$	$\hat{H}_{NO,\bar{w}}$	$\hat{H}_{\text{pre},1}$	$\hat{H}_{\text{pre},2}$	\hat{H}_{LW}^{tr}	$\hat{H}_{\text{GPH}}^{\text{tr}}$
$n = 1000$											
0	0.0560	0.0743	0.0874	0.2629	0.1717	0.2661	0.1733	0.0486	0.0728	0.0914	0.1469
$\tau = 0.1$											
0.5	0.0818	0.0925	0.0993	0.2650	0.1745	0.2644	0.1745	0.0535	0.0878	0.0899	0.1469
1	0.1361	0.1268	0.1449	0.2646	0.1777	0.2662	0.1802	0.0703	0.1124	0.0961	0.1511
2	0.2431	0.1872	0.2264	0.2673	0.1867	0.2691	0.1903	0.1018	0.1540	0.1165	0.1608
$\tau = 0.5$											
0.5	0.1201	0.0965	0.1181	0.2633	0.1738	0.2646	0.1738	0.0500	0.0649	0.0960	0.1501
1	0.2038	0.1298	0.1540	0.2654	0.1768	0.2667	0.1785	0.0538	0.0658	0.0968	0.1505
2	0.3234	0.1856	0.1916	0.2664	0.1864	0.2684	0.1883	0.0528	0.0675	0.1174	0.1635

5 Application to real data

We illustrate our improved estimation techniques with the help of two real data sets. We use the famous *Nile River minima*, yearly minimal water levels of the Nile River for the years 773 to 1281; the data is available e.g. in the R-package `wavelets`¹. It is a widely-used example for LRD time series. Log-periodogram values of the Nile data were considered in [3, p. 21] where the evidence of long range dependence was found. A broad analysis of the data was carried out by [34] where the authors applied several techniques to estimate the LRD parameter taking the seasonality of the river data into account. It was pointed out that the dataset may exhibit seasonal long memory rather than nonseasonal long range dependence.

We also apply the estimation techniques to the *global temperature for the northern hemisphere* for the years 1854–1989; the data is available in the R-package `longmemo`². This dataset was discussed in [3, p. 29] where a linear trend rather than a possible structural change in the data was considered. This was supported by [11], and [26] assumed that the temperature growth is rather due to impact of human activity or climate change. Evidence of long range dependence was found by [3, p. 173] and [15]. A possible change point in the data was detected by [52] using kernel and nearest neighbour regression methods.

5.1 Nile River minima

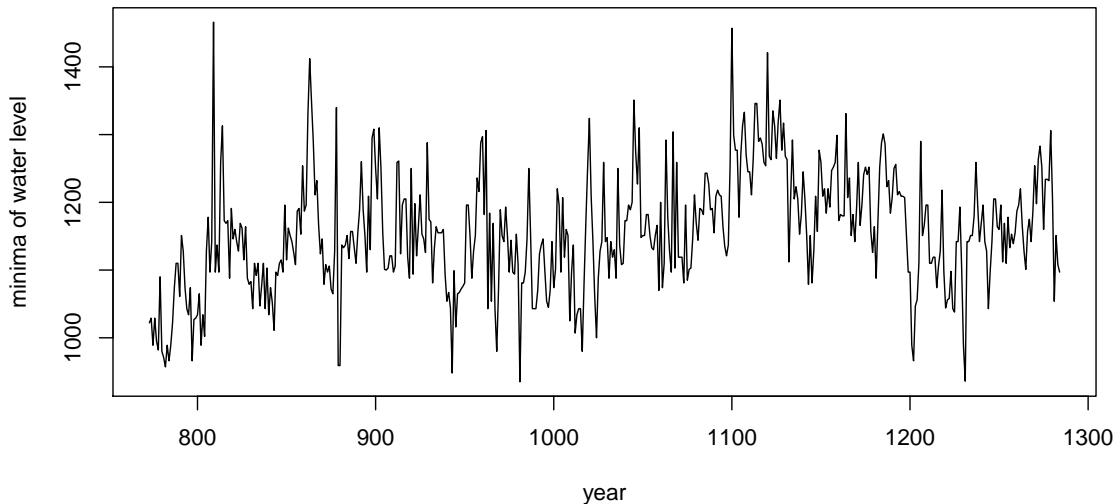


Fig. 1: Annual minima of the water level of the Nile River near Cairo.

The data is displayed in Figure 1. In order to estimate the LRD parameter H from this data, we have applied the local Whittle estimator to it – as well as the adaption methods investigated in this article. The results are given in Table 5. The blocks estimators with ARMA correction from Section 4.5 are denoted by $\hat{H}_{O,w}^{\text{ARMA}}$ and $\hat{H}_{NO,w}^{\text{ARMA}}$. For analysing the data [35, p. 386] use wavelets and the maximum-likelihood

¹ <http://artax.karlin.mff.cuni.cz/r-help/library/wavelets/html/nile.html>

² <http://cran.r-project.org/web/packages/longmemo/longmemo.pdf>

method yielding 0.9452 as an estimate for the Hurst parameter which is similar to the value $\hat{H}_{LW} = 0.9706$ obtained by the ordinary local Whittle estimator. As opposed to this the overlapping blocks approach with ARMA correction with the window length \sqrt{n} yields the smaller value $\hat{H}_{O,w}^{ARMA} = 0.7820$. The two estimation techniques \hat{H}_{GPH}^{tr} and \hat{H}_{LW}^{tr} described in Section 2.5 yield the higher values 0.9999 and 1.0071. The discrepancy between the estimates suggests that the time series might be contaminated by a level shift.

Table 5: LRD parameter H of the Nile River Minima data, estimated by different adaption methods, based on the local Whittle estimator.

\hat{H}_{LW}	\hat{H}_{mean}	$\hat{H}_{mindiff}$	$\hat{H}_{pre,1}$	$\hat{H}_{pre,2}$	$\hat{H}_{O,w}$	$\hat{H}_{NO,w}$	$\hat{H}_{O,w}^{ARMA}$	$\hat{H}_{NO,w}^{ARMA}$	\hat{H}_{LW}^{tr}	\hat{H}_{GPH}^{tr}
0.9706	0.9639	0.9607	0.9339	0.9486	0.8358	0.8749	0.7820	0.8294	0.9999	1.0071

5.2 Global temperature for the northern hemisphere

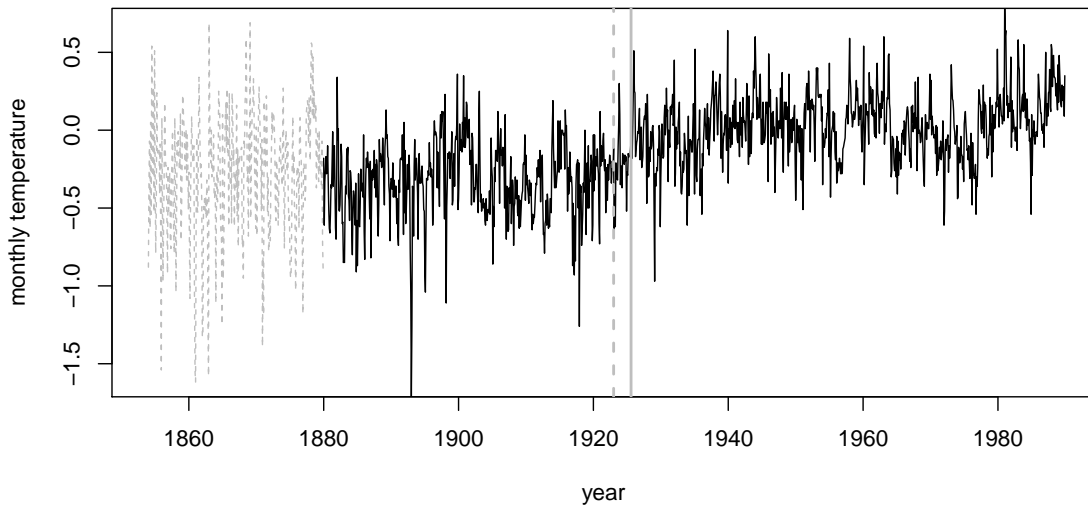


Fig. 2: Monthly global temperature for the northern hemisphere for the years 1854-1989. The vertical lines denote a change-point, suggested by [52] (dashed) and detected using the Wilcoxon change-point test (bold). The grey part of the time series was excluded from the estimation procedure.

Figure 2 shows the global temperature for the northern hemisphere. A presence of a level shift after the year 1923 was suggested by [52]. Moreover, we can observe a change in the scale after the year 1880. Since we are interested in changes in the mean only, we reduce the data to the years 1880-1989 yielding 1320 monthly observations.

We have applied the local Whittle estimator together with the adaption methods to this data. The results are given in Table 6. The ordinary local Whittle estimator yields $\hat{H}_{LW} = 0.9383$, while the other estimates obtained by the adapted methods are smaller. The overlapping blocks approach with ARMA correction and window length \sqrt{n} yields $\hat{H}_{O,w}^{ARMA} = 0.8378$ which is similar to the value obtained by $\hat{H}_{pre,1} = 0.8457$. It can be assumed that the pre-estimation technique was able to determine the location of a level shift

adequately and estimated the Hurst parameter precisely in the two blocks with considerably large number of observations. Hence, we assume that we are dealing with a change in the mean in the presented time series.

Table 6: LRD parameter H of the global temperature for the northern hemisphere, estimated by different adaption methods, based on the local Whittle estimator.

\hat{H}_{LW}	\hat{H}_{mean}	\hat{H}_{mindiff}	$\hat{H}_{\text{pre},1}$	$\hat{H}_{\text{pre},2}$	$\hat{H}_{O,w}$	$\hat{H}_{NO,w}$	$\hat{H}_{O,w}^{\text{ARMA}}$	$\hat{H}_{NO,w}^{\text{ARMA}}$	\hat{H}_{LW}^{tr}	$\hat{H}_{GPH}^{\text{tr}}$
0.9383	0.9109	0.9005	0.8457	0.8658	0.8541	0.8618	0.8378	0.8568	0.9052	0.9300

6 Conclusion and outlook

In time series which are supposed to exhibit LRD, it is important for statistical inference to estimate the LRD parameter like the Hurst parameter $H \in (0.5, 1)$, respectively $D = 2H - 2$ in (3). Different approaches for this problem exist, but most of them suffer from the defect that they are disturbed by structural changes in the data: If there is a shift in the mean, the usual estimation methods easily misjudge the structure of the data and may detect spurious or too heavy LRD. This drawback is a fatal double bind when it comes to testing for change-points in LRD time series: Change-point tests naturally require knowledge of the Hurst parameter H in order to discriminate between a change-point and behavior which originates from the long memory, but in practical situations, H must be estimated. Hence, a jump in the data may lead to overestimating H which may lead to overlooking the jump.

We have investigated techniques to adapt estimation procedures to time series with a change in the mean. In a simulation study, we have compared these proposals with different tuning parameters and in different jump-contaminated and also jump-free situations, using the local Whittle and the Geweke and Porter-Hudak estimators as generic estimation techniques. Our research shows that estimation of H can be considerably improved by such techniques which yield better estimates when there is a jump in the time series and which little affect the estimation when there is none. We compared the performance of the adapted estimation techniques with the trimmed versions of the local Whittle and the Geweke and Porter-Hudak estimators proposed by [24] and [33], respectively. We could observe that our proposals outperform the trimmed counterparts of the two generic estimation techniques used in this paper in most simulation scenarios.

We recommend estimation of H by averaging estimates obtained from overlapping blocks with length approximately equal to a multiple of \sqrt{n} . This adaption inherits asymptotic consistency from its underlying estimation method, while offering fast computation. A minimum of 100 observations per block should be available for the estimation when dealing with ARFIMA models. In this case the estimation should be carried out using the ARMA correction from Section 4.5.

Our approach opens up some ideas for future research:

- One could explore possible improvements by e.g. choosing other monotone increasing subsequences of n as window sizes for the moving window methods or by introducing weighted means for the methods which separate the sample into two blocks, taking the different block lengths into account.
- In the moving window or the non-overlapping blocks approach, one could use a trimmed mean in order to trim the estimations on blocks which are influenced by the jump.

Acknowledgements

The financial support of the Deutsche Forschungsgemeinschaft (SFB 823, “Statistical modelling of nonlinear dynamic processes”) is gratefully acknowledged. We would like to thank two referees for their constructive comments which improved the presentation of our work considerably.

References

1. R. T. Baillie, Long memory processes and fractional integration in econometrics, *Journal of Econometrics*, 73(1), 5–59 (1996)
2. J. T. Barkoulas, C. F. Baum, N. Travlos, Long memory in the Greek stock market, *Applied Financial Economics*, 10(2), 177–184 (2000)
3. J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall/CRC, Boca Raton (Florida) 1994, ISBN 0-412-04901-5 (1994)
4. J. Beran, Y. Feng, S. Ghosh, R. Kulik, *Long-memory processes: Probabilistic properties and statistical methods*, Springer (2013)
5. I. Berkes, L. Horváth, P. Kokoszka, Q.-M. Shao, On discriminating between long-range dependence and changes in the mean, *The Annals of Statistics*, 34, 1140–1165 (2006)
6. F. J. Breidt, N. Crato, P. de Lima, The detection and estimation of long memory in stochastic volatility, *Journal of Econometrics*, 83(1-2), 325–348 (1998)
7. Y.-W. Cheung, K. S. Lai, A search for long memory in international stock market returns, *Journal of International Money and Finance*, 14(4), 597–615 (1995)
8. M. Csörgő, L. Horváth, *Limit Theorems in Change-Point Analysis*, John Wiley & Sons, Chichester 1997, ISBN 0-471-95522-1 (1997)
9. N. J. Cutland, P. E. Kopp, W. Willinger, Stock price returns and the Joseph effect: a fractional version of the Black-Scholes model, in: E. Bolthausen, M. Dozzi, F. Russo (eds.) *Seminar on Stochastic Analysis, Random Fields and Applications*, Birkhäuser, Boston, 327–351 (1995)
10. H. Dehling, A. Rooch, M. S. Taqqu, Nonparametric Change-Point Tests for Long-Range Dependent Data. *Scandinavian Journal of Statistics*, 40, 153–173 (2013)
11. R. S. Deo, C. M. Hurvich, Linear trend with fractionally integrated errors. *Journal of Time Series Analysis*, 19(4), 379–397 (2013)
12. F. X. Diebold, A. Inoue, Long memory and regime switching. *Journal of econometrics*, 105(1), 131–159 (2001)
13. A. Erramilli, O. Narayan, W. Willinger, Experimental queueing analysis with long-range dependent packet traffic, *IEEE/ACM Transactions on Networking*, 4(2), 209–223 (1996)
14. J. Geweke, S. Porter-Hudak, The estimation and application of long-memory times series models *Journal of Time Series Analysis*, 4(4), 221–238 (1983)
15. L. A. Gil-Alana, Statistical modeling of the temperatures in the Northern Hemisphere using fractional integration techniques, *Journal of Climate*, 18(24), 5357–5369 (2005)
16. L. Giraitis, R. Leipus, D. Surgailis, The change-point problem for dependent observations, *Journal of Statistical Planning and Inference*, 53, 297–310 (1996)
17. C. W. Granger, N. Hyung, Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *Journal of empirical finance*, 11(3), 399–421 (2004)
18. U. Hassler, M. Olivares, Long memory and structural change: New evidence from German stock market returns, *Goethe University Frankfurt Discussion Paper* (2008)
19. U. Hassler, J. Scheithauer, Detecting Changes from Short to Long Memory. *Statistical Papers*, 52, 847–870 (2009)
20. L. Horváth, P. Kokoszka, The effect of long-range dependence on change-point estimators, *J. Statist. Plann. Inference*, 64, 57–81 (1997)
21. C.-C. Hsu, Long memory or structural changes: An empirical examination on inflation rates, *Economics Letters*, 88(2), 289–294 (2005)
22. C. M. Hurvich, R. Deo, J. Brodsky, The mean squared error of Geweke and Porter-Hudak’s estimator of the memory parameter of a long-memory time series, *Journal of Time Series Analysis*, 19(1), 19–46 (1998)
23. C. M. Hurvich, K. I. Beltrao, Asymptotics for the Low Frequency Ordinates of the Periodogram of a Long-Memory Time Series, *Journal of Time Series Analysis*, 14(5), 455–472 (1993)

24. F. Iacone, Local Whittle estimation of the memory parameter in presence of deterministic components, *Journal of Time Series Analysis*, 31(1), 37-49 (2010)
25. V. Jandhyala, S. Fotopoulos, I. MacNeill, P. Liu, Inference for single and multiple change-points in time series, *Journal of Time Series Analysis*, doi: 10.1111/jtsa.12035 (2013)
26. D. Jaruskova, Some problems with application of change-point detection methods to environmental data, *Environmetrics*, 8(5), 469–484 (1997)
27. T. Karagiannis, M. Faloutsos, R. H. Riedi, Long-Range Dependence: Now you see it, now you don't!, *Global Telecommunications Conference*, 3, 2165–2169 (2002)
28. P. Kokoszka, R. Leipus, Change-point in the mean of dependent observations, *Stat. & Prob. Letters*, 40, 385–393 (1998)
29. W. Krämer, P. Sibbertsen, Testing for Structural Changes in the Persistence of Long Memory. *International Journal of Business and Economics*, 1, 235–242 (2002)
30. H. R. Künsch, Statistical aspects of self-similar processes, in: Yu. A. Prohorov, V.V. Sazonov (eds.), *Proceedings of the First World Congress of the Bernoulli Society*, 1, 67–74, VNU Science Press, Utrecht (1987)
31. Q. Li, D. L. Mills, On the Long-range Dependence of Packet Round-trip Delays in Internet, *In Proceedings of IEEE ICC98*, 1185–1191 (1998)
32. A. W. Lo (1991), Long-term memory in stock market prices. *Econometrica*, 59, 1279–1313
33. A. McCloskey, P. Perron, Memory parameter estimation in the presence of level shifts and deterministic trends, *Econometric Theory*, 29(06), 1196-1237 (2013)
34. A. Montanari, R. Rosso, M. S. Taqqu, A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan, *Water Resources Research*, 36(5), 1249–1259 (2000)
35. D. B. Percival, A. T. Walden, *Wavelet methods for time series analysis (Vol. 4)*. Cambridge University Press. (2006)
36. P. C. Phillips, K. Shimotsu, Local Whittle estimation in nonstationary and unit root cases, *The Annals of Statistics*, 32(2), 656–692 (2004)
37. H. Rachinger, Multiple Breaks in Long Memory Time Series, *Job Market Paper, Universidad Carlos III de Madrid* (2011)
38. V. Reisen, B. Abraham, S. Lopes, Estimation of parameters in ARFIMA processes: A simulation study, *Communications in Statistics-Simulation and Computation*, 30(4), 787–803(2001)
39. P. M. Robinson, Gaussian semiparametric estimation of long range dependence, *The Annals of statistics*, 5, 1630–1661 (1995)
40. P. M. Robinson, Log-periodogram regression of time series with long range dependence, *The Annals of statistics*, 3, 1048–1072 (1995)
41. X. Shao, A simple test of changes in mean in the possible presence of long-range dependence, *Journal of Time Series Analysis*, 32(6), 598–606 (2011)
42. K. Shimotsu, P. C. Phillips, Exact local Whittle estimation of fractional integration, *The Annals of Statistics*, 33(4), 1890–1933 (2005)
43. P. Sibbertsen, Long Memory versus Structural Breaks: An Overview, *Statistical Papers*, 45, 465–515 (2004)
44. P. Sibbertsen, J. Willert, Testing for a Break in Persistence under Long-Range Dependencies and Mean Shifts, *Statistical Papers*, 53, 357–370 (2010)
45. R. L. Smith, Long-Range Dependence and Global Warming, in: R. A. Madden, R. W. Katz: *Applications of Statistics to Modeling the Earth's Climate System*, 89–92 (1993)
46. M. S. Taqqu, V. Teverovsky, On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series, in: R. Adler, R. Feldmann, M. S. Taqqu (eds.), *A Practical Guide To Heavy Tails: Statistical Techniques and Applications*, Birkhäuser, Boston 1998, 177–217 (1998)
47. M. S. Taqqu, V. Teverovsky, W. Willinger, Estimators for long-range dependence: an empirical study, *Fractals*, 3, 785–798 (1995)
48. C. Velasco, Gaussian Semiparametric Estimation of Non-stationary Time Series, *Journal of Time Series Analysis*, 20(1), 87–127 (1999)
49. C. Velasco, P. M. Robinson, Whittle pseudo-maximum likelihood estimation for nonstationary time series, *Journal of the American Statistical Association*, 95(452), 1229–1243 (2000)
50. L. Wang, Change-in-mean problem for long memory time series models with applications, *Journal of Statistical Computation and Simulation*, 78(7), 653–668 (2008)
51. L. Wang (2008), Change-point detection with rank statistics in long-memory time-series models, *Australian and New Zealand Journal of Statistics*, 50(3), 241–256

-
52. L. Wang, Change-point estimation in long memory nonparametric models with applications. *Communications in Statistics-Simulation and Computation*, 37(1), 48–61 (2008)
 53. P. Whittle, Estimation and information in stationary time series, *Arkiv för matematik*, 2(5), 423–434 (1953)
 54. W. Willinger, M. S. Taqqu, V. Teverovsky, Stock market prices and long-range dependence, *Finance and Stochastics*, 3(1), 1–13 (1999)

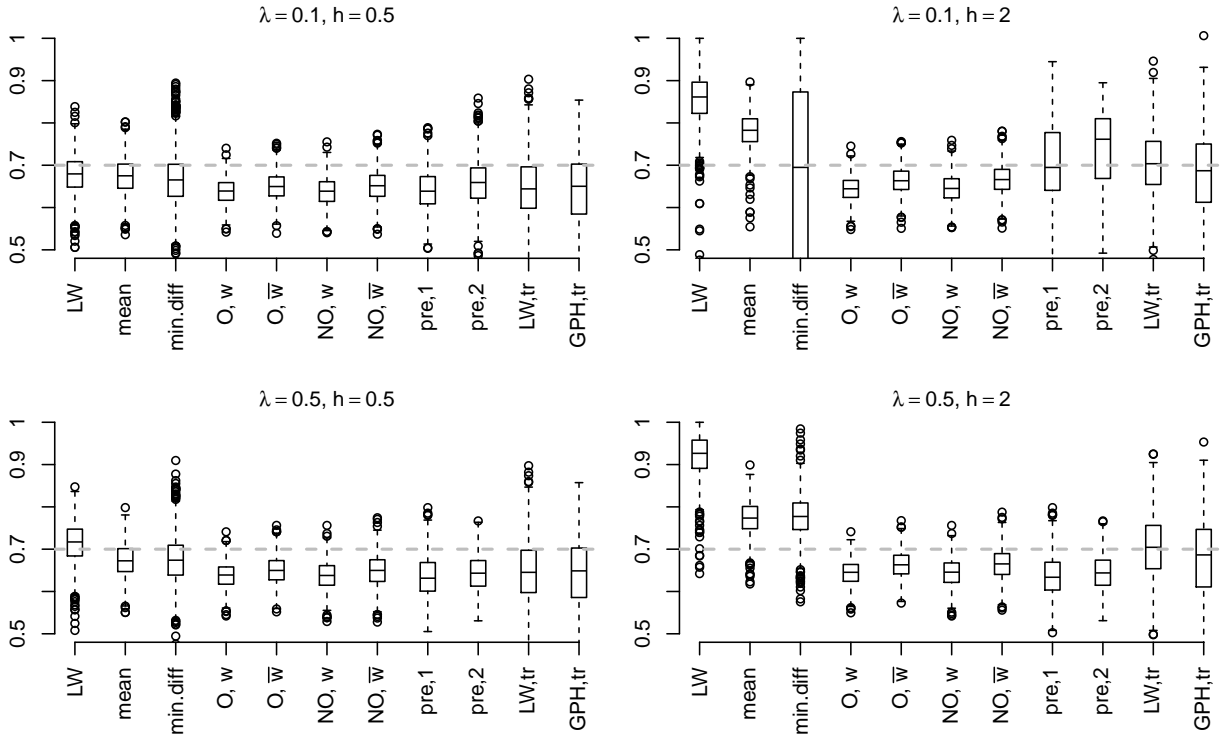


Fig. 3: Boxplots of the estimators based on the local Whittle estimator for the Hurst parameter $H = 0.7$ in time series with change-point (jump of height h after a proportion of τ of the data), each based on 1000 simulation runs with $n = 1000$ realizations of Pareto(3,1)-transformed fGn.

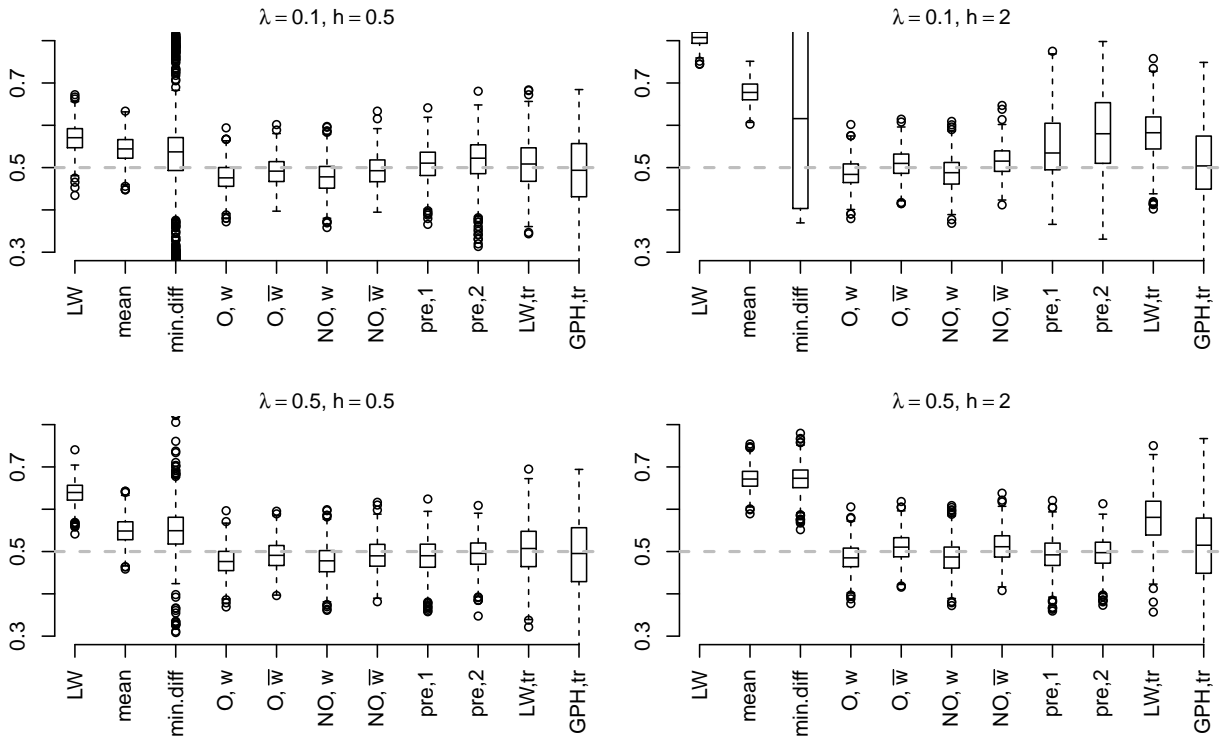


Fig. 4: Boxplots of the estimators based on the local Whittle estimator for the Hurst parameter $H = 0.5$ in time series with change-point (jump of height h after a proportion of τ of the data), based on each 1000 simulation runs with $n = 1000$ realizations of fGn.

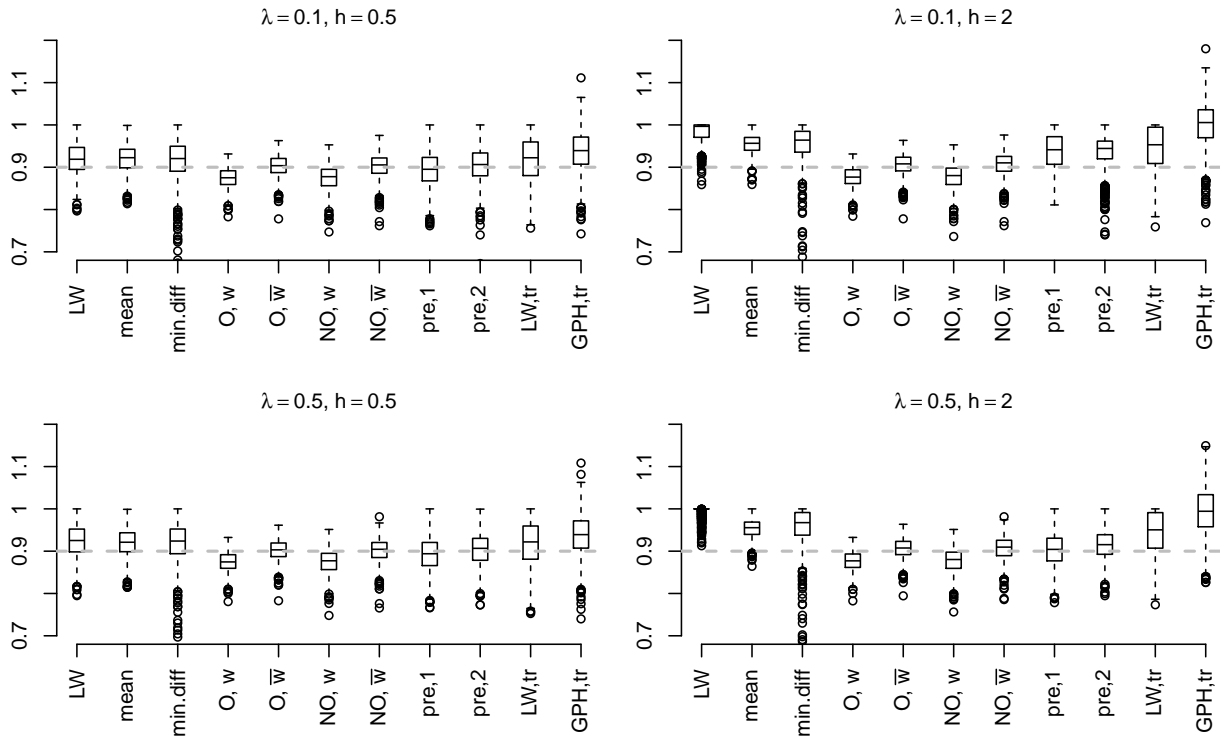


Fig. 5: Boxplots of the estimators based on the local Whittle estimator for the Hurst parameter $H = 0.9$ in time series with change-point (jump of height h after a proportion of τ of the data), based on each 1000 simulation runs with $n = 1000$ realizations of fGn.

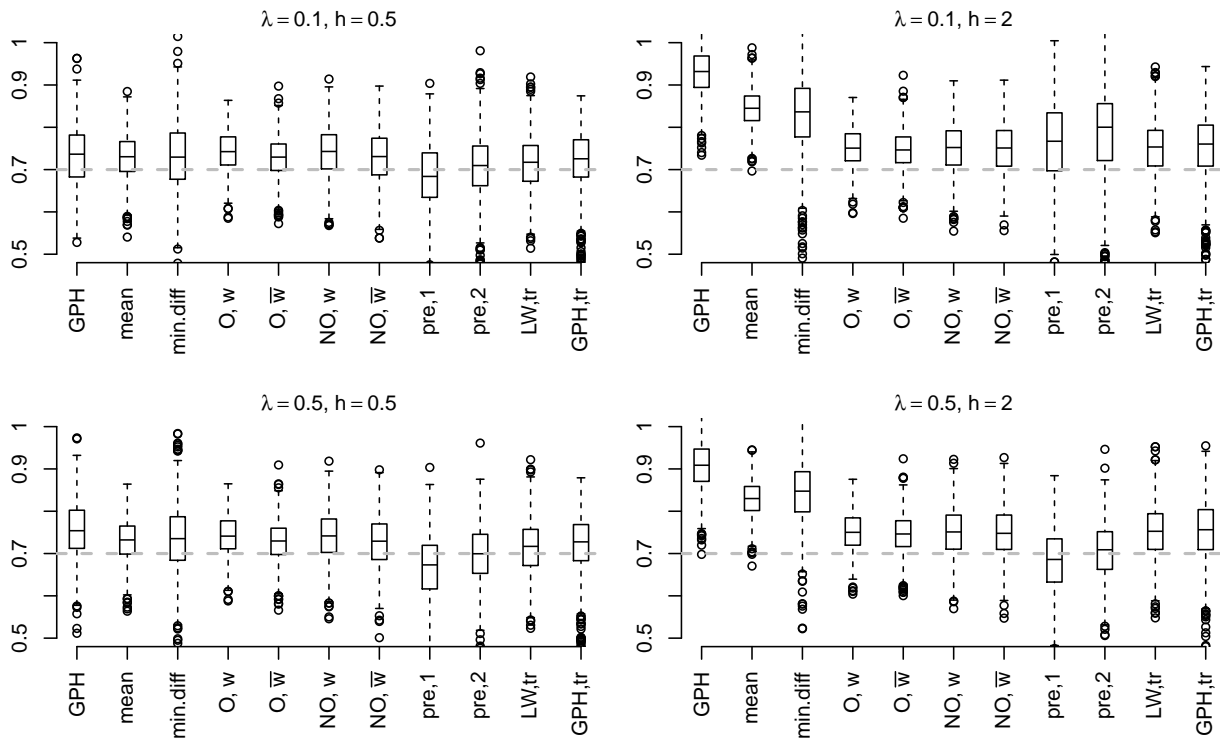


Fig. 6: Boxplots of the estimators based on the GPH estimator for the Hurst parameter $H = 0.7$ in time series with change-point (jump of height h after a proportion of τ of the data), based on each 1000 simulation runs with $n = 1000$ realizations of fGn.

Table 7: Estimated RMSE of non-overlapping blocks estimator with different block sizes for the differencing parameters $d = 0$ ($H = 0.5$) and $d = 0.2$ ($H = 0.7$) in time series without ($h = 0$) and with change-point (jump of height h after a proportion of τ of the data, each based on 1000 simulation runs with $n = 1000$ and $n = 1500$ values from the ARFIMA(0, d , -0.6) model and fGn), based on the local Whittle estimator.

ARFIMA(0, d , -0.6)										
h	$d = 0$					$d = 0.2$				
	\sqrt{n}	$2\sqrt{n}+1$	$3\sqrt{n}+1$	$4\sqrt{n}+1$	$7\sqrt{n}+1$	\sqrt{n}	$2\sqrt{n}+1$	$3\sqrt{n}+1$	$4\sqrt{n}+1$	$7\sqrt{n}+1$
$n = 1000$										
0	0.2661	0.1733	0.1399	0.1200	0.0935	0.2171	0.1782	0.1548	0.1390	0.1059
$\tau = 0.1$										
0.5	0.2644	0.1745	0.1404	0.1267	0.1040	0.2165	0.1779	0.1544	0.1387	0.1116
1	0.2662	0.1802	0.1415	0.1346	0.1250	0.2150	0.1803	0.1572	0.1421	0.1203
2	0.2691	0.1903	0.1463	0.1535	0.1618	0.2179	0.1834	0.1578	0.1515	0.1397
$\tau = 0.5$										
0.5	0.2646	0.1738	0.1447	0.1221	0.1020	0.2161	0.1764	0.1558	0.1378	0.1100
1	0.2667	0.1785	0.1508	0.1251	0.1198	0.2150	0.1787	0.1587	0.1387	0.1180
2	0.2684	0.1883	0.1636	0.1380	0.1539	0.2176	0.1821	0.1652	0.1444	0.1376
$n = 1500$										
0	0.2322	0.1506	0.1230	0.1069	0.0807	0.2051	0.1626	0.1408	0.1257	0.0962
$\tau = 0.1$										
0.5	0.2337	0.1515	0.1277	0.1095	0.0926	0.2055	0.1622	0.1425	0.1264	0.0986
1	0.2330	0.1524	0.1324	0.1087	0.1095	0.2053	0.1631	0.1435	0.1264	0.1054
2	0.2356	0.1588	0.1460	0.1136	0.1341	0.2054	0.1652	0.1514	0.1298	0.1215
$\tau = 0.5$										
0.5	0.2343	0.1545	0.1280	0.1121	0.0913	0.2057	0.1630	0.1430	0.1269	0.0969
1	0.2367	0.1589	0.1351	0.1173	0.1033	0.2054	0.1646	0.1425	0.1299	0.1043
2	0.2380	0.1664	0.1454	0.1316	0.1293	0.2057	0.1684	0.1522	0.1365	0.1198
fGn										
h	$d = 0$					$d = 0.2$				
	\sqrt{n}	$2\sqrt{n}+1$	$3\sqrt{n}+1$	$4\sqrt{n}+1$	$7\sqrt{n}+1$	\sqrt{n}	$2\sqrt{n}+1$	$3\sqrt{n}+1$	$4\sqrt{n}+1$	$7\sqrt{n}+1$
$n = 1000$										
$\tau = 0.5$										
2	0.0403	0.0386	0.0442	0.0459	0.0968	0.0360	0.0438	0.0496	0.0495	0.0763
$n = 1500$										
$\tau = 0.5$										
2	0.0324	0.0332	0.0402	0.0471	0.0755	0.0308	0.0423	0.0458	0.0508	0.0632

Table 8: Estimated RMSE of non-overlapping and overlapping blocks estimator based on the local Whittle estimator with different block sizes for the differencing parameters $d = 0$ ($H = 0.5$) and $d = 0.2$ ($H = 0.7$) in time series without ($h = 0$) and with change-point (jump of height h after a proportion of τ of the data), each based on 1000 simulation runs with $n = 1000$ and $n = 1500$ values from the ARFIMA(0, d , 1) model and fGn. The ARMA-structure is removed in several steps.

<u>non-overlapping</u>										
ARFIMA(0, d , 1)										
h	$d = 0$					$d = 0.2$				
	\sqrt{n}	$2\sqrt{n}+1$	$3\sqrt{n}+1$	$4\sqrt{n}+1$	$7\sqrt{n}+1$	\sqrt{n}	$2\sqrt{n}+1$	$3\sqrt{n}+1$	$4\sqrt{n}+1$	$7\sqrt{n}+1$
$n = 1000$										
0	0.1206	0.0595	0.0552	0.0633	0.0557	0.0554	0.0552	0.0569	0.0664	0.0633
$\tau = 0.1$										
0.5	0.1183	0.0639	0.0569	0.0664	0.0640	0.0528	0.0534	0.0580	0.0662	0.0671
1	0.1196	0.0708	0.0629	0.0776	0.0871	0.0527	0.0566	0.0599	0.0692	0.0708
2	0.1190	0.0806	0.0691	0.1029	0.1271	0.0488	0.0597	0.0586	0.0767	0.0869
$\tau = 0.5$										
0.5	0.1167	0.0607	0.0590	0.0646	0.0613	0.0513	0.0531	0.0603	0.0677	0.0673
1	0.1191	0.0668	0.0622	0.0701	0.0783	0.0512	0.0553	0.0587	0.0677	0.0689
2	0.1190	0.0717	0.0759	0.0758	0.1097	0.0526	0.0563	0.0656	0.0733	0.0854
$n = 1500$										
0	0.0745	0.0516	0.0512	0.0509	0.0476	0.0412	0.0485	0.0522	0.0538	0.0518
$\tau = 0.1$										
0.5	0.0751	0.0536	0.0558	0.0550	0.0519	0.0413	0.0514	0.0545	0.0579	0.0529
1	0.0783	0.0585	0.0643	0.0571	0.0706	0.0405	0.0505	0.0562	0.0588	0.0566
2	0.0781	0.0636	0.0769	0.0624	0.1033	0.0406	0.0533	0.0616	0.0581	0.0726
$\tau = 0.5$										
0.5	0.0771	0.0570	0.0563	0.0550	0.0511	0.0405	0.0502	0.0546	0.0574	0.0533
1	0.0789	0.0614	0.0636	0.0623	0.0621	0.0424	0.0506	0.0587	0.0592	0.0546
2	0.0812	0.0681	0.0750	0.0776	0.0903	0.0412	0.0559	0.0628	0.0659	0.0679
fGn										
h	$H = 0.5$					$H = 0.7$				
	\sqrt{n}	$2\sqrt{n}+1$	$3\sqrt{n}+1$	$4\sqrt{n}+1$	$7\sqrt{n}+1$	\sqrt{n}	$2\sqrt{n}+1$	$3\sqrt{n}+1$	$4\sqrt{n}+1$	$7\sqrt{n}+1$
$n = 1000$										
$\tau = 0.5$										
2	0.0417	0.0454	0.0517	0.0561	0.1028	0.0398	0.0456	0.0507	0.0563	0.0818
$n = 1500$										
$\tau = 0.5$										
2	0.0339	0.0396	0.0471	0.0554	0.0880	0.0335	0.0415	0.0455	0.0493	0.0608
<u>overlapping</u>										
ARFIMA(0, d , 1)										
h	$d = 0$					$d = 0.2$				
	\sqrt{n}	$2\sqrt{n}+1$	$3\sqrt{n}+1$	$4\sqrt{n}+1$	$7\sqrt{n}+1$	\sqrt{n}	$2\sqrt{n}+1$	$3\sqrt{n}+1$	$4\sqrt{n}+1$	$7\sqrt{n}+1$
$n = 1000$										
$\tau = 0.5$										
2	0.1181	0.0651	0.0706	0.0854	0.1046	0.0369	0.0479	0.0543	0.0661	0.0722

