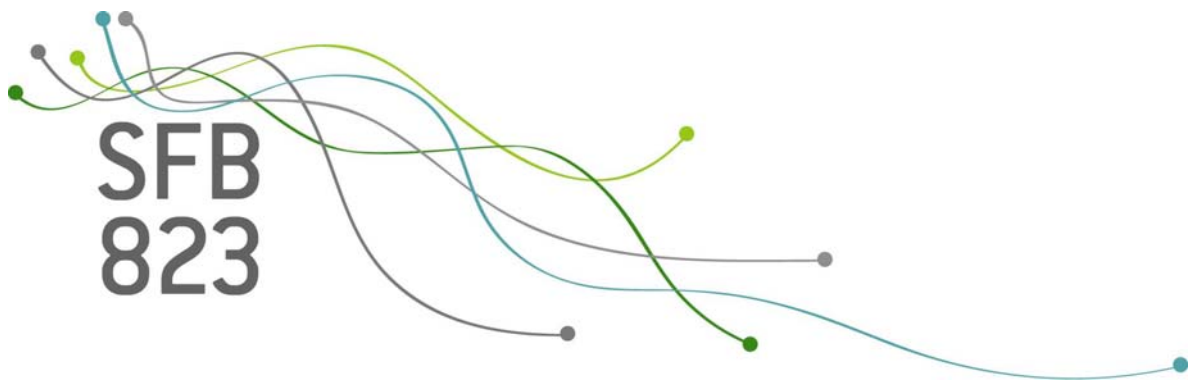


SFB  
823

# Subsampling for general statistics under long range dependence

Annika Betken, Martin Wendler

Nr. 36/2015



Discussion Paper



# Subsampling for General Statistics under Long Range Dependence

Annika Betken\*, Martin Wendler

*Ruhr-Universität Bochum*  
*e-mail: annika.betken@rub.de*

*Ernst-Moritz-Arndt-Universität Greifswald*  
*e-mail: martin.wendler@uni-greifswald.de*

**Abstract:** In the statistical inference for long range dependent time series, the shape of the limit distribution typically depends on unknown parameters. Therefore, we propose to use subsampling. We show the validity of subsampling for general statistics and long range dependent subordinated Gaussian processes, which satisfy mild regularity conditions. We apply our method to a self-normalized change-point test statistic and investigate the finite sample properties in a simulation study.

**MSC 2010 subject classifications:** Primary 60G15, 62G09; secondary 60G22.

**Keywords and phrases:** Subsampling, Gaussian Processes, Long Range Dependence, Change Point Test.

## 1. Introduction

### 1.1. Long Range Dependence

While most statistical research is done for independent data or short memory time series, in many applications there are also time series with long memory in the sense of slowly decaying correlations: in hydrology (starting with the work of Hurst [23]), in finance (e.g. Lo [30]), in the analysis of network traffic (e.g. Leland, Taqqu, Willinger and Wilson [29]) and many more fields of research.

As model of dependent time series, we will consider subordinated Gaussian processes: Let  $(\xi_n)_{n \in \mathbb{N}}$  be a stationary sequence of centered Gaussian variables with  $\text{Var}(\xi_n) = 1$  and covariance function  $\gamma$  satisfying

$$\gamma(k) := \text{Cov}(\xi_1, \xi_{k+1}) = k^{-D} L_\gamma(k), \quad (1)$$

for a  $D > 0$  and a slowly varying function  $L_\gamma$ . If  $D < 1$ , then the spectral density  $f$  of  $(\xi_n)_{n \in \mathbb{N}}$  is not continuous, but has a pole at 0. The spectral density has the form

$$f(x) = |x|^{D-1} L_f(x)$$

for a function  $L_f$ , which is slowly varying at the origin (see Proposition 1.1.14 in Pipiras and Taqqu [35]).

---

\*supported by Collaborative Research Center SFB 823 and by the German Academic Foundation

Furthermore, let  $G : \mathbb{R} \rightarrow \mathbb{R}$  be a measurable function such that  $E[G^2(\epsilon_1)] < \infty$ . The stochastic process  $(X_n)_{n \in \mathbb{N}}$  given by

$$X_n := G(\xi_n)$$

is called long range dependent if  $\sum_{n=0}^{\infty} |\text{Cov}(X_1, X_{n+1})| = \infty$ , and short range dependent if  $\sum_{n=0}^{\infty} |\text{Cov}(X_1, X_{n+1})| < \infty$ .

In limit theorems for the partial sum  $S_n = \sum_{i=1}^n X_i$ , the normalization required and the shape of the limit distribution depend not only on the decay of the covariances  $\gamma(k)$  as  $k \rightarrow \infty$ , but also on the function  $G$ . More precisely, Taqqu [43] and Dobruhsin, Major [18] independently proved that

$$\frac{1}{L_\gamma(n)^{r/2} n^H} \sum_{i=1}^n (X_i - E[X_i]) \xrightarrow{D} C(r, H) g_r Z_{r, H}(1),$$

if the Hurst parameter  $H := \max\{1 - \frac{rD}{2}, \frac{1}{2}\}$  is greater than  $\frac{1}{2}$ .  $r$  is the Hermite rank of the function  $G$ ,  $g_r$  is the first non-zero coefficient in the expansion of  $G$  as a sum of Hermite polynomials and  $Z_{r, H}$  is a Hermite process. For more details on Hermite polynomials and limit theorems for subordinated Gaussian processes, we recommend the book of Pipiras and Taqqu [35]. In this case ( $rD < 1$ ), the process  $(X_n)_{n \in \mathbb{N}}$  is long range dependent, as the covariances are not summable. Note that the limiting random variable  $C(r, H)Z_{r, H}(1)$  is Gaussian only if the Hermite rank  $r = 1$ .

If  $rD = 1$ , the process  $(X_n)_{n \in \mathbb{N}}$  might be short or long range dependent according to the slowly varying function  $L_\gamma$ . If  $rD > 1$ , the process is short range dependent. In this case, the partial sum  $\sum_{i=1}^n (X_i - E[X_i])$  has (with the proper normalization) always a Gaussian limit.

There are other models for long-memory processes: Fractionally integrated autoregressive moving average processes can show long range dependence, see Granger and Joyeux [20]. General linear processes with slowly decaying coefficients were studied by Surgailis [42].

## 1.2. Subsampling

In practical applications, the parameters  $D$ ,  $r$  and the slowly varying function  $L_\gamma$  are unknown and thus the scaling needed in the limit theorems and the shape of the asymptotic distribution are not known, either. That makes it difficult to use the asymptotic distribution for statistical inference. The situation gets even more complicated if one is not interested in partial sums, but in nonlinear statistical functionals. For example,  $U$ -statistics can have a linear combination of random variables related to different Hermite ranks as a limit, see Beutner and Zähle [10]. Self-normalized statistics typically converge to quotients of two random variables (e.g. McElroy and Politis [33]). The change-point test proposed by Berkes, Horváth, Kokoszka and Shao [8] converges under the alternative hypothesis to a supremum of a fractional Brownian bridge.

To overcome the problem of the unknown shape of the limit distribution and to avoid the estimation of nuisance parameters, one would like to use nonparametric methods. However, Lahiri [28] has shown that the popular moving block bootstrap might fail under long range dependence. Another nonparametric approach is subsampling (also called sampling window method), first studied by Politis and Romano [36], Hall and Jing [21], and Sherman and Carlstein [40]. The idea is the following: Let  $T_n = T_n(X_1, \dots, X_n)$  be a series of statistics converging in distribution to a random variable  $T$ , but as we typically only have one sample, we observe only one realization of  $T_n$  and can not estimate the distribution of  $T_n$ . But if  $l = l_n$  is a sequence with  $l_n \rightarrow \infty$  and  $l_n = o(n)$ , then  $T_l$  also converges in distribution to  $T$  and we have multiple (though dependent) realizations  $T_l(X_1, \dots, X_l)$ ,  $T_l(X_2, \dots, X_{l+1})$ , ...,  $T_l(X_{n-l+1}, \dots, X_n)$ , which we can use to calculate the empirical distribution function.

The validity of subsampling for the sample mean  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  under long range dependence has been investigated in the literature starting with Hall, Jing and Lahiri [22] for subordinated Gaussian processes. Nordman and Lahiri [34] and Zhang, Ho, Wendler, Wu [47] studied linear processes with slowly decaying coefficients. An alternative proof in the case of Gaussian processes can be found in the book of Beran, Feng, Ghosh, Kulik [7].

It was noted by Fan [19] that the proof of Hall et al. [22] can be easily generalized to other statistics than the sample mean. However, the assumptions on the Gaussian process are restrictive (see also McElroy and Politis [33]). Their conditions imply that the sequence  $(\xi_n)_{n \in \mathbb{N}}$  is completely regular, which might hold for some special cases (see Ibragimov and Rozanov [24]), but excludes many examples:

**Example 1** (Fractional Gaussian Noise). Let  $(B_H(t))_{t \in [0, \infty)}$  be a fractional Brownian motion, i.e. a centered, self-similar Gaussian process with covariance function

$$\mathbb{E}[B_H(t)B_H(s)] = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H})$$

for some  $H \in (\frac{1}{2}, 1)$ . Then  $(\xi_n)_{n \in \mathbb{N}}$  given by  $\xi_n = B_H(n) - B_H(n-1)$  is called fractional Gaussian noise. By self-similarity, we have

$$\begin{aligned} \text{corr}\left(\sum_{i=1}^n \xi_i, \sum_{j=2n+1}^{3n} \xi_j\right) &= \text{corr}(B_H(n), B_H(3n) - B_H(2n)) \\ &= \text{corr}(B_H(1), B_H(3) - B_H(2)). \end{aligned}$$

So the correlations of linear combinations of observations in the past and future do not vanish if the gap between past and future grows and thus fractional Gaussian noise is not completely regular.

Jach, McElroy, Politis [25] provided a more general result on the validity of subsampling. They assume that the function  $G$  has Hermite rank 1, that  $G$  is invertible and Lipschitz-continuous and that the process  $(\xi_n)_{n \in \mathbb{N}}$  has a causal representation as a functional of an independent sequence of random variables.

These assumptions are difficult to check in practice. Moreover, although not explicitly stated in their Theorem 4, the statistic  $T_n$  has to be Lipschitz-continuous (uniformly in  $n$ ), which is not satisfied by many robust estimators (see Section 3 for an example).

The main aim of this paper is to establish the validity of the subsampling method for general statistics  $T_n$  without any assumptions on the continuity of the statistic, on the function  $G$  and only mild assumptions on the Gaussian process  $(\xi_n)_{n \in \mathbb{N}}$ . We will give our main theorem in the next section. In Section 3, we will apply our theorem to a self-normalized, robust change-point statistic. The finite sample properties of this test will be investigated in Section 4. Finally, the proof of the main result and the lemmas needed can be found in Section 5.

## 2. Main Results

### 2.1. Statement of the Theorem

For a statistic  $T_n = T_n(X_1, \dots, X_n)$ , the subsampling estimator  $\hat{F}_{l,n}$  of the distribution function  $F_{T_n}$  with  $F_{T_n}(t) = P(T_n \leq t)$  is defined in the following way: Let for  $t \in \mathbb{R}$

$$\hat{F}_{l,n}(t) = \frac{1}{n-l+1} \sum_{j=1}^{n-l+1} 1_{\{T_l(X_j, \dots, X_{j+l-1}) \leq t\}}.$$

Our first assumption guarantees the convergence of the distribution function of  $T_n$ :

**Assumption 1.**  $(X_n)_{n \in \mathbb{N}}$  is a stochastic process and  $(T_n)_{n \in \mathbb{N}}$  is a sequence of statistics, such that  $T_n \rightarrow T$  in distribution for a random variable  $T$  with distribution function  $F_T$ .

This is a standard assumption for subsampling, see for example Politis and Romano [36]. If the distribution does not converge, we can not expect the distribution of  $T_l$  to be close to the distribution of  $T_n$ .

Next, we will formulate our conditions on the sequence  $(X_n)_{n \in \mathbb{N}}$  of random variables:

**Assumption 2.**  $X_n = G(\xi_n)$  for a measurable function  $G$  and a stationary, Gaussian process  $(\xi_n)_{n \in \mathbb{N}}$  with covariance function

$$\gamma(k) := \text{Cov}(\xi_1, \xi_{1+k}) = k^{-D} L_\gamma(k)$$

such that the following conditions hold:

1.  $D \in (0, 1]$  and  $L_\gamma$  is a slowly varying function with

$$\max_{\tilde{k} \in \{k+1, \dots, k+2l'-1\}} \left| L_\gamma(k) - L_\gamma(\tilde{k}) \right| \leq K \frac{l'}{k} \min \{L_\gamma(k), 1\}$$

for a constant  $K < \infty$  and all  $l' \in \{l_k, \dots, k\}$ .

2.  $(\xi_n)_{n \in \mathbb{N}}$  has a spectral density  $f$  with  $f(x) = |x|^{D-1} L_f(x)$  for a slowly varying function  $L_f$  which is bounded away from 0 on  $[0, \pi]$ .

While we have some regularity conditions on the underlying Gaussian process  $(\xi_n)_{n \in \mathbb{N}}$ , note that we do not have any conditions on the function  $G$ , no finite moments or continuity are required, which will make our results applicable for heavy-tailed random variables and robust test statistics. We will show that Assumption 2 holds for some standard examples of long range dependent Gaussian processes in the next subsection.

Furthermore, we need a restriction on the growth rate of the block length  $l$ :

**Assumption 3.** Let  $(l_n)_{n \in \mathbb{N}}$  be a nondecreasing sequence of integers such that  $l = l_n \rightarrow \infty$  as  $n \rightarrow \infty$  and  $l_n = \mathcal{O}(n^{(1+D)/2-\epsilon})$  for an  $\epsilon > 0$ .

If the dependence of the underlying process  $(\xi_n)_{n \in \mathbb{N}}$  gets stronger, the range of possible values for  $l$  gets smaller. A popular choice for the block length is  $l \approx C\sqrt{n}$  (see for example Hall et al. [22]), which is allowed for all  $D \in (0, 1]$ . Now we can state our main result:

**Theorem 1.** Under the Assumptions 1, 2 and 3 for all points of continuity  $t$  of  $F_T$ , we have

$$F_{T_n}(t) - \hat{F}_{l,n}(t) \xrightarrow{\mathcal{P}} 0.$$

If  $F_T$  is continuous, then

$$\sup_{t \in \mathbb{R}} \left| F_{T_n}(t) - \hat{F}_{l,n}(t) \right| \xrightarrow{\mathcal{P}} 0.$$

So we have a consistent estimator for the distribution function of  $T_n$  and it is possible to build tests and confidence intervals based on this estimator.

If  $D > 1$ , then the process  $(\xi_n)_{n \in \mathbb{N}}$  is strongly mixing by Theorem 9.8 in the book of Bradley [13] and the statements of our theorem hold by Corollary 3.2 in Politis and Romano [36] for any blocklength  $l$  satisfying  $l \rightarrow \infty$  and  $l = o(n)$ .

## 2.2. Examples for our Assumptions

We will now give two examples of Gaussian processes satisfying Assumption 2:

**Example 2** (Fractional Gaussian Noise). Fractional Gaussian Noise  $(\xi_n)_{n \in \mathbb{N}}$  with Hurst parameter  $H$  as introduced in Example 1 has the covariance function

$$\gamma(k) = \frac{1}{2} (|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}) = H(2H-1) (k^{-D} + h(k)k^{-D-1})$$

for  $D = 2 - 2H$  and a function  $h$  bounded by a constant  $M < \infty$ , which can be easily seen by means of a Taylor expansion. So  $L_\gamma(k) = H(2H-1)(1 + h(k)/k)$  and for all  $\tilde{k} \geq k$

$$\left| L_\gamma(k) - L_\gamma(\tilde{k}) \right| \leq H(2H-1) \left| \frac{h(k)}{k} - \frac{h(\tilde{k})}{\tilde{k}} \right| \leq H(2H-1) \frac{M}{k} =: K \frac{1}{k}$$

implying Part 1 of Assumption 2. For the second part, note that spectral density  $f$  of fractional Gaussian noise is given by

$$\begin{aligned} f(\lambda) &= C(H)(1 - \cos(\lambda)) \sum_{k=-\infty}^{\infty} |x + 2k\pi|^{D-3} \\ &= \lambda^{D-1} C(H) \frac{1 - \cos(\lambda)}{\lambda^2} \frac{\sum_{k=-\infty}^{\infty} |\lambda + 2k\pi|^{D-3}}{\lambda^{D-3}} \end{aligned}$$

see Sinai [41]. The slowly varying function

$$L_f(\lambda) = C(H) \frac{1 - \cos(\lambda)}{\lambda^2} \frac{\sum_{k=-\infty}^{\infty} |\lambda + 2k\pi|^{D-3}}{\lambda^{D-3}}$$

is bounded away from 0 because this obviously holds for the first factor  $(1 - \cos(\lambda))/\lambda^2$  and

$$\frac{\sum_{k=-\infty}^{\infty} |\lambda + 2k\pi|^{D-3}}{\lambda^{D-3}} \geq \frac{|\lambda + 20\pi|^{D-3}}{\lambda^{D-3}} = 1.$$

**Example 3** (Gaussian FARIMA processes). Let  $(\varepsilon_n)_{n \in \mathbb{Z}}$  be Gaussian white noise with variance  $\sigma^2 = \text{Var}(\varepsilon_0)$ . Then for a  $d \in (0, 1/2)$ , a FARIMA(0,  $d$ , 0) process  $(\xi_n)_{n \in \mathbb{N}}$  is given by

$$\xi_n = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \varepsilon_{n-j}.$$

According to Pipiras and Taquq [35], Section 1.3, it has the spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi} |1 - e^{-i\lambda}|^{-2d} = |\lambda|^{D-1} \frac{\sigma^2}{2\pi} \left( \frac{|\lambda|}{|1 - e^{-i\lambda}|} \right)^{1-D}$$

with  $D = 1 - 2d \in (0, 1)$ . As  $|1 - e^{-i\lambda}| \leq \lambda$ , part 2 of Assumption 2 holds. For part 1, we have by Corollary 1.3.4 of [35] that

$$\gamma(k) = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} \frac{\Gamma(k+d)}{\Gamma(k-d+1)}.$$

Recall that by the Stirling formula  $\Gamma(x) = (\frac{2\pi}{x})^{1/2} (\frac{x}{e})^x (1 + \mathcal{O}(x^{-1}))$  and consequently

$$\gamma(k) = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)} e^{-2d+1} k^{2d-1} \left( \frac{k+d}{k} \right)^{k+d} \left( \frac{k}{k-d+1} \right)^{k-d+1} \left( 1 + \mathcal{O}\left(\frac{1}{k}\right) \right).$$

Using a Taylor expansion of  $(k+d)(\log(k+d) - \log(k)) + (k-d+1)(\log(k) - \log(k-d+1))$ , it easily follows that

$$\gamma(k) = k^{-D} L_\gamma(k)$$

with  $L_\gamma(k) = C + \mathcal{O}(1/k)$  for a constant  $C$ . Part 1 of Assumption 2 follows in the same way as in Example 2.



It would be interesting to know if the subsampling for general statistics is consistent for long range dependent linear processes without the assumption of gaussianity, but this seems to be a very hard problem and goes beyond the scope of this paper.

### 3. Application

#### 3.1. Robust, Self-Normalized Change-Point Test

In this paper, the main motivation for considering subsampling procedures in order to approximate the distribution of test statistics consists in avoiding the choice of unknown parameters. As an example we will consider a self-normalized test statistic that can be applied to detect changes in the mean of long range dependent and heavy-tailed time series.

Given observations  $X_1, \dots, X_n$  with  $X_i = G(\xi_i) + \mu_i$  we are concerned with a decision on the change-point problem

$$\mathbf{H} : \mu_1 = \dots = \mu_n$$

against

$$\mathbf{A} : \mu_1 = \dots = \mu_k \neq \mu_{k+1} = \dots = \mu_n \text{ for some } k \in \{1, \dots, n-1\}.$$

Under the hypothesis  $\mathbf{H}$  we assume that the data generating process  $(X_n)_{n \in \mathbb{N}}$  is stationary, while under the alternative  $\mathbf{A}$  there is a change in location at an unknown point in time. This problem has been widely studied: Csörgő and Horváth [15] give an overview of parametric and non-parametric methods that can be applied in order to detect change-points in independent data.

Many commonly used testing procedures are based on Cusum (cumulated sum) test statistics, but when applied to data sets generated by long range dependent processes, these change-point tests often falsely reject the hypothesis of no change in the mean (see also Baek and Pipiras [4]). Furthermore, the performance of Cusum-like change-point tests is sensitive to outliers in the data.

In contrast, testing procedures that are based on rank statistics have the advantage of not being sensitive to outliers in the data. These were introduced by Antoch et al. [3] for detecting changes in the distribution function of independent random variables. Wilcoxon type rank tests have been studied by Wang [45] in the presence of linear long memory time series and by Dehling, Rooch and Taqqu [16] for subordinated Gaussian sequences.

Note that the normalization of the Wilcoxon change-point test statistic as proposed in [16] depends on the slowly varying function  $L_\gamma$ , the LRD parameter  $D$  and the Hermite rank  $r$  of the class of functions  $1_{\{X_i \leq x\}} - F(x)$ ,  $x \in \mathbb{R}$ . Although  $r = 1$  is assumed in many cases and while there are well-trying methods to estimate  $D$ , estimating  $L_\gamma$  does not seem to be an easy task. For this reason, the Wilcoxon change-point test does not seem to be suitable for application to real data.

So as to avoid these issues Betken [9] proposes an alternative normalization for the Wilcoxon change-point test. This normalization approach has originally been established by Lobato [31] for decision on the hypothesis that a short range dependent stochastic process is uncorrelated up to a lag of a certain order. In change-point analysis the normalization has recently been applied to several test statistics: Shao and Zhang [39] define a self-normalized Kolmogorov-Smirnov test statistic that serves to identify changes in the mean of short range dependent time series. Shao [38] adopted the normalization so as to define an alternative normalization for a Cusum test, which detects changes in the mean of short range dependent as well as long range dependent time series.

For the definition of the self-normalized Wilcoxon test statistic, we introduce the ranks  $R_i = \text{rank}(X_i) = \sum_{j=1}^n 1_{\{X_j \leq X_i\}}$  for  $i = 1, \dots, n$ . It seems natural to transfer the normalization that has been used by Shao [38] on the Cusum test statistic to the ranks in order to establish a self-normalized version of the Wilcoxon test statistic, which is robust to outliers in the data. Therefore, the corresponding two-sample test statistic is defined by

$$G_n(k) = \frac{\sum_{i=1}^k R_i - \frac{k}{n} \sum_{i=1}^n R_i}{\left\{ \frac{1}{n} \sum_{t=1}^k S_t^2(1, k) + \frac{1}{n} \sum_{t=k+1}^n S_t^2(k+1, n) \right\}^{1/2}},$$

where

$$S_t(j, k) = \sum_{h=j}^t (R_h - \bar{R}_{j,k}) \quad \text{with} \quad \bar{R}_{j,k} = \frac{1}{k-j+1} \sum_{t=j}^k R_t.$$

The self-normalized Wilcoxon change-point test rejects the hypothesis for large values of  $\max_{k \in \{\lfloor n\tau_1 \rfloor, \dots, \lfloor n\tau_2 \rfloor\}} |G_n(k)|$ , where  $0 < \tau_1 < \tau_2 < 1$ . The proportion of the data that is included in the calculation of the supremum is restricted by  $\tau_1$  and  $\tau_2$ . A common choice for these parameters is  $\tau_1 = 1 - \tau_2 = 0.15$ ; see Andrews [2].

For long range dependent subordinated Gaussian processes  $(X_n)_{n \in \mathbb{N}}$ , the asymptotic distribution of the test statistic under the hypothesis  $\mathbf{H}$  can be derived by the continuous mapping theorem (see Theorem 1 in Betken [9]):

$$T_n(\tau_1, \tau_2) := \max_{k \in \{\lfloor n\tau_1 \rfloor, \dots, \lfloor n\tau_2 \rfloor\}} |G_n(k)| \\ \Rightarrow \sup_{\tau_1 \leq \lambda \leq \tau_2} \frac{|Z_r(\lambda) - \lambda Z_r(1)|}{\left\{ \int_0^\lambda (Z_r(t) - \frac{t}{\lambda} Z_r(\lambda))^2 dt + \int_0^{1-\lambda} (Z_r^*(t) - \frac{t}{1-\lambda} Z_r^*(1-\lambda))^2 dt \right\}^{1/2}}.$$

Here,  $Z_r$  is a  $r$ -th order Hermite process with Hurst parameter  $H := \max\{1 - \frac{rD}{2}, \frac{1}{2}\}$  and  $Z_t^*(r) = Z_r(1) - Z_r(1-t)$ . A comparison of  $T_n(\tau_1, \tau_2)$  with the critical values of its limit distribution still presupposes determination of these parameters, but Assumption 1 holds and we can bypass the estimation of  $D$  and  $r$  by applying the subsampling procedure.

Note that even under the alternative  $\mathbf{A}$  (change in location), we have to find the quantiles of the distribution under the hypothesis (stationarity). But as the

block length  $l$  is much shorter than the sample size  $n$ , most blocks will not be contaminated by the change-point and so the distribution of the test statistic will not change. Even for blocks which cover the change, the power of the test is low (because there are only  $l$  observations) and thus the distribution will not change too much. The accuracy and the power of the test will be investigated by a simulation study in Section 4.

### 3.2. Data Examples

We will revisit some data sets which have been analyzed before in the literature. We will use the self-normalized Wilcoxon change-point test combined with subsampling and compare our findings to the conclusions of other authors.

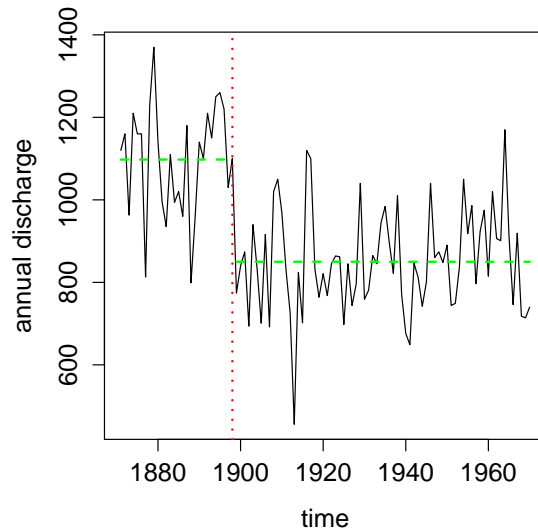


FIGURE 1. The annual discharge from the Nile at Aswan in  $10^8 m^3$  for 1871-1970. The dotted line indicates the location of the change-point; the dashed lines designate the sample means for the pre-break and post-break samples.

The plot in Figure 1 depicts the annual volume of discharge from the Nile river at Aswan in  $10^8 m^3$  for the years 1871 to 1970. The data set has been analyzed for the detection of a change-point by numerous authors under differing assumptions concerning the data generating random process and by usage of diverse methods. Amongst others, Cobb [14], MacNeill, Tang and Jandhyala [32], Wu and Zhao [46] and Shao [38] provided statistical significant evidence for a decrease of the Nile's annual discharge towards the end of the 19th century. The construction of the Aswan Low Dam between 1898 and 1902 serves as a popular explanation for an abrupt change in the data.

The value of the self-normalized Wilcoxon test statistic computed with respect to the data is given by  $T_n(\tau_1, \tau_2) = 13.48729$ . For a level of significance of 5%, the self-normalized Wilcoxon change-point test rejects the hypothesis for every possible value of  $H \in (\frac{1}{2}, 1)$ . Furthermore, we approximate the distribution of the self-normalized Wilcoxon test statistic by the sampling window method with block size  $l = \lfloor \sqrt{n} \rfloor$ . The subsampling-based test decision also indicates the existence of a change-point in the mean of the data.

In particular, previous analysis of the Nile data done by Wu and Zhao [46] and Balke [5] suggests that the change in the discharge volume occurred in 1899. We applied the self-normalized Wilcoxon test and the sampling window method to the corresponding prebreak and postbreak samples. Neither of these two approaches leads to rejection of the hypothesis, so that it seems reasonable to consider both samples as stationary. At this point, it is interesting to note that, based on the whole sample, local Whittle estimation with bandwidth parameter  $m = \lfloor n^{2/3} \rfloor$  suggests the existence of long range dependence characterized by an Hurst parameter  $\hat{H} = 0.962$ , whereas the estimates for the prebreak and postbreak samples given by  $\hat{H}_1 = 0.517$  and  $\hat{H}_2 = 0.5$ , respectively, should be considered as indication of short range dependent data. In this regard our findings support the conjecture of spurious long memory caused by a change-point and therefore coincide with the results of Shao [38].

The second data set consists of the seasonally adjusted monthly deviations of the temperature (degrees C) for the northern hemisphere during the years 1854 to 1989 from the monthly averages over the period 1950 to 1979. The data results from spatial averaging of temperatures measured over land and sea. At first sight the plot in Figure 2 may suggest an increasing trend as well as an abrupt change of the temperature deviations. Statistical evidence for a positive deterministic trend implies affirmation of the conjecture that there has been global warming during the last decades.

In scientific discourse the question of whether the Northern hemisphere temperature data acts as an indicator for global warming of the atmosphere is a controversial issue. Deo and Hurvich [17] provided some indication for global warming by fitting a linear trend to the data. Beran and Feng [6] considered a more general stochastic model by the assumption of so called semiparametric fractional autoregressive (SEMIFAR) processes. Their method did not deliver sufficient statistical evidence for a deterministic trend. Wang [44] applied another method for the detection of gradual change to the global temperature data and did not detect an increasing trend either. But he offers an alternative explanation for the occurrence of trend-like behavior by pointing out that it may have been generated by stationary long range dependent processes. In contrast, it is shown in Shao [38] that the existence of a change-point in the mean yields yet another explanation for the performance of the data.

The value of the self-normalized Wilcoxon test statistic computed with respect to the data is given by  $T_n(\tau_1, \tau_2) = 18.98636$ . Consequently, the self-normalized Wilcoxon change-point test would reject the hypothesis for every possible value of  $H \in (\frac{1}{2}, 1)$  at a level of significance of 1%. In addition, an application of the sampling window method with respect to the self-normalized

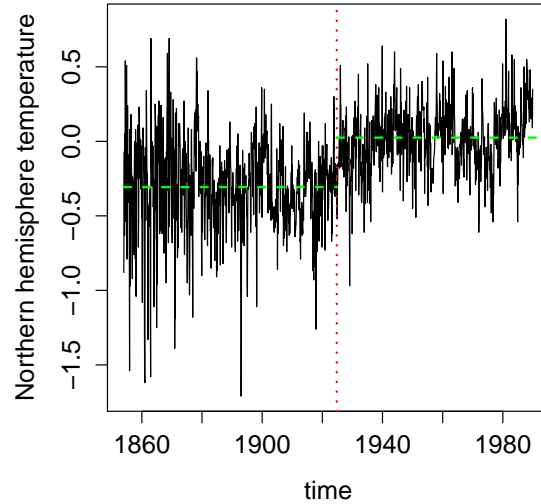


FIGURE 2. Monthly temperature for the northern hemisphere for the years 1854-1989, from the data base held at the Climate Research Unit of the University of East Anglia, Norwich, England. The numbers consist of the temperature (degrees C) difference from the monthly average over the period 1950-1979. The dotted line indicates the location of the potential change-point; the dashed lines designate the sample means for the pre-break and post-break samples.

Wilcoxon test statistic yields a test decision in favor of the alternative hypothesis for any possible choice of the block length. All in all, both testing procedures provide strong evidence for the existence of a change in the mean.

According to Shao [38] the change-point is located around October 1924. Based on the whole sample local Whittle estimation with bandwidth  $m = \lfloor n^{2/3} \rfloor$  provides an estimator  $\hat{H} = 0.811$ . The estimated Hurst parameters for the prebreak and postbreak sample are  $\hat{H}_1 = 0.597$  and  $\hat{H}_2 = 0.88$ , respectively. Neither of both testing procedures, i.e. subsampling with respect to the self-normalized Wilcoxon test statistic and comparison of the value of  $T_n(\tau_1, \tau_2)$  with the corresponding critical values of its limit distribution, provides evidence for another change-point in the prebreak or postbreak sample.

Therefore, it seems safe to conclude that the appearance of long memory in the postbreak sample is not caused by another change-point in the mean. The pronounced difference between the local Whittle estimators  $\hat{H}_1$  and  $\hat{H}_2$  suggests a change in the dependence structure of the times series. Another explanation might be a gradual change of the temperature in the postbreak period.

In both data examples, we find that the results obtained by subsampling and obtained by parameter estimation are in good accordance with each other. The

methods take into account possible long range dependence or heavy tails, but they still can detect a change in location.

#### 4. Simulations

We will now investigate the finite sample performance of the subsampling procedure with respect to the self-normalized Wilcoxon test and with respect to the classical Wilcoxon change-point test. Moreover, we will compare these results to the performance of the tests when the test decision is based on critical values obtained from the asymptotic distribution of the test statistic.

For this purpose, we consider subordinated Gaussian time series  $(X_n)_{n \in \mathbb{N}}$ ,  $X_n = G(\xi_n)$ , where  $(\xi_n)_{n \in \mathbb{N}}$  is fractional Gaussian noise (introduced in Example 2) with Hurst parameter  $H \in \{0.6, 0.7, 0.8, 0.9\}$  and covariance function

$$\gamma(k) \sim k^{-D} \left(1 - \frac{D}{2}\right) (1 - D)$$

where  $D = 2 - 2H$ , i.e.  $L_\gamma(k) \sim \left(1 - \frac{D}{2}\right) (1 - D)$ .

Initially, we take  $G(t) = t$ , so that  $(X_n)_{n \in \mathbb{N}}$  has normal marginal distributions. We also consider the transformation

$$G(t) = \left(\frac{\beta k^2}{(\beta - 1)^2(\beta - 2)}\right)^{-\frac{1}{2}} \left(k(\Phi(t))^{-\frac{1}{\beta}} - \frac{\beta k}{\beta - 1}\right)$$

with  $\Phi$  denoting the standard normal distribution function so as to generate Pareto-distributed data with parameters  $k, \beta > 0$  (referred to as Pareto( $\beta, k$ )). In both cases the Hermite rank  $r$  of  $1_{\{G(\xi_i) \leq x\}} - F(x), x \in \mathbb{R}$ , equals  $r = 1$  and

$$\left| \int_{\mathbb{R}} J_1(x) dF(x) \right| = \frac{1}{2\sqrt{\pi}};$$

see Dehling, Rooch and Taquq [16].

Under the above conditions the critical values of the asymptotic distribution of the self-normalized Wilcoxon test statistic are reported in Table 2 in Betken [9]. The limit of the Wilcoxon change-point test statistic can be found in [16], the corresponding critical values can be taken from Table 1 in [9].

The frequencies of rejections of both testing procedures are reported in Table 1 for the self-normalized Wilcoxon change-point test and in Table 2 for the classical Wilcoxon test (without self-normalization). The calculations are based on 5,000 realizations of time series with sample size  $n = 300$  and  $n = 500$ .

For the usual testing procedures the estimation of the Hermite rank  $r$ , the slowly varying function  $L_\gamma$  and the integral  $\int J_1(x) dF(x)$  is neglected. Yet, for every simulated time series we estimate the Hurst parameter  $H$  by the local Whittle estimator  $\hat{H}$  proposed in Künsch [27]. This estimator is based on an approximation of the spectral density by the periodogram at the Fourier frequencies. It depends on the spectral bandwidth parameter  $m = m(n)$  which

denotes the number of Fourier frequencies used for the estimation. If the bandwidth  $m$  satisfies  $\frac{1}{m} + \frac{m}{n} \rightarrow 0$  as  $n \rightarrow \infty$ , the local Whittle estimator is a consistent estimator for  $H$ ; see Robinson [37]. For convenience we always choose  $m = \lfloor n^{2/3} \rfloor$  in this article. The critical values corresponding to the estimated values of  $H$  are determined by linear interpolation.

Under the alternative **A** we analyze the power of the testing procedures (the frequency of rejection) by considering different choices for the height of the level shift (denoted by  $h$ ) and the location  $\lfloor n\tau \rfloor$  of the change-point. In the tables the columns that are superscribed by “ $h = 0$ ” correspond to the frequency of a type 1 error, i.e. the rejection rate under the hypothesis **H**.

For the self-normalized Wilcoxon change-point test (based on the asymptotic distribution), the empirical size almost equals the level of significance of 5% for normally distributed data (see Table 1). The sampling window method yields a rejection rate that slightly exceeds this level. For Pareto(3, 1) time series both testing procedures lead to similar results and tend to reject the hypothesis too often when there is no change. With regard to the empirical power it is notable that for fractional Gaussian noise time series the sampling window method yields considerably better power than the test based on asymptotic critical values. If Pareto(3, 1) distributed time series are considered, the empirical power of the subsampling procedure is still better than the empirical power that results from using asymptotic critical values. However, in this case the deviation of the rejection rates is rather small. While the empirical size is not much affected by the Hurst parameter  $H$ , the empirical power is lower for  $H = 0.8, 0.9$ .

Considering the classical Wilcoxon test (without self-normalization), it is notable that for both procedures, the empirical size is in most cases not close to the nominal level of significance (5%), ranging from 1.1% to 17.5% using subsampling and from 2.6% to 36.0% using asymptotic critical values (see Table 2). In general, the sampling window method becomes more conservative for higher values of the Hurst parameter  $H$ , while the test based on the asymptotic distribution becomes more liberal. Under the alternative the usual application of the Wilcoxon test yields better power than the sampling window method, especially for high values of  $H$ . But we emphasize that this comparison is problematic because the rejection frequencies under the hypothesis differ.

We conclude that the self-normalized Wilcoxon change-point test is more reliable than the classical change-point test. The reason might be that in the scaling of the classical test, the estimator  $\hat{H}$  of the Hurst parameter enters as a power of the sample size  $n$  and thus a small error in this estimation might lead to a large error in the value of the test statistic. By using the sampling window method for the self-normalized version, we avoid the estimation of unknown parameters so that the performance is similar to the performance of the classical testing procedure which compares the values of the test statistic with the corresponding critical values.

TABLE 1  
 Rejection rates of the self-normalized Wilcoxon change-point test obtained by subsampling (left) and by comparison with asymptotic critical values (right) for fractional Gaussian noise and Pareto(3, 1)-transformed fractional Gaussian noise of length  $n$  with Hurst parameter  $H$ .

		sampling window method					asymptotic distribution				
		$\tau = 0.25$			$\tau = 0.5$		$\tau = 0.25$			$\tau = 0.5$	
fGn	$n$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$
$H = 0.6$	300	0.064	0.313	0.742	0.570	0.964	0.044	0.209	0.521	0.424	0.861
	500	0.060	0.421	0.861	0.720	0.995	0.049	0.303	0.687	0.577	0.958
$H = 0.7$	300	0.070	0.171	0.423	0.313	0.763	0.053	0.108	0.268	0.228	0.611
	500	0.059	0.193	0.508	0.382	0.854	0.048	0.133	0.359	0.302	0.730
$H = 0.8$	300	0.067	0.117	0.234	0.208	0.494	0.048	0.081	0.144	0.141	0.362
	500	0.068	0.114	0.278	0.210	0.567	0.053	0.085	0.198	0.163	0.462
$H = 0.9$	300	0.074	0.097	0.161	0.169	0.397	0.057	0.065	0.106	0.125	0.308
	500	0.067	0.091	0.166	0.162	0.416	0.051	0.068	0.120	0.128	0.350
Pareto(3, 1)	$n$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$
$H = 0.6$	300	0.067	0.871	0.946	0.990	1.000	0.056	0.820	0.912	0.984	0.999
	500	0.066	0.946	0.994	0.999	1.000	0.061	0.920	0.970	0.996	1.000
$H = 0.7$	300	0.064	0.527	0.738	0.876	0.990	0.070	0.529	0.702	0.856	0.982
	500	0.068	0.684	0.893	0.942	0.998	0.076	0.663	0.820	0.940	0.995
$H = 0.8$	300	0.068	0.284	0.454	0.666	0.905	0.072	0.297	0.428	0.640	0.875
	500	0.063	0.379	0.581	0.714	0.933	0.069	0.369	0.510	0.715	0.920
$H = 0.9$	300	0.071	0.168	0.254	0.532	0.772	0.073	0.165	0.236	0.499	0.738
	500	0.064	0.219	0.340	0.547	0.802	0.068	0.199	0.296	0.529	0.782



TABLE 2  
 Rejection rates of the classical Wilcoxon change-point test obtained by subsampling (left) and by comparison with asymptotic critical values (right) for fractional Gaussian noise and Pareto(3, 1)-transformed fractional Gaussian noise of length  $n$  with Hurst parameter  $H$ .

		sampling window method					asymptotic distribution				
		$\tau = 0.25$			$\tau = 0.5$		$\tau = 0.25$			$\tau = 0.5$	
fGn	$n$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$
$H = 0.6$	300	0.054	0.223	0.411	0.439	0.784	0.026	0.096	0.160	0.223	0.727
	500	0.058	0.345	0.663	0.627	0.952	0.036	0.148	0.256	0.378	0.897
$H = 0.7$	300	0.049	0.095	0.158	0.206	0.466	0.035	0.067	0.228	0.167	0.665
	500	0.039	0.131	0.267	0.287	0.689	0.030	0.079	0.259	0.225	0.714
$H = 0.8$	300	0.029	0.038	0.048	0.075	0.179	0.077	0.153	0.421	0.245	0.673
	500	0.028	0.044	0.070	0.097	0.273	0.050	0.112	0.439	0.226	0.714
$H = 0.9$	300	0.009	0.014	0.009	0.021	0.060	0.36	0.484	0.739	0.524	0.830
	500	0.011	0.009	0.011	0.029	0.086	0.319	0.439	0.743	0.511	0.845
Pareto(3, 1)	$n$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$	$h = 0$	$h = 0.5$	$h = 1$	$h = 0.5$	$h = 1$
$H = 0.6$	300	0.130	0.963	0.861	0.996	0.991	0.108	0.938	0.985	0.998	1.000
	500	0.132	0.997	0.976	1.000	0.999	0.128	0.988	0.999	1.000	1.000
$H = 0.7$	300	0.175	0.802	0.680	0.955	0.949	0.179	0.833	0.969	0.974	0.999
	500	0.167	0.931	0.862	0.992	0.996	0.191	0.940	0.994	0.996	1.000
$H = 0.8$	300	0.160	0.496	0.347	0.776	0.808	0.204	0.729	0.925	0.918	0.993
	500	0.160	0.649	0.513	0.886	0.929	0.212	0.805	0.963	0.948	0.999
$H = 0.9$	300	0.097	0.128	0.071	0.403	0.550	0.309	0.712	0.901	0.848	0.966
	500	0.100	0.161	0.101	0.518	0.680	0.27	0.726	0.911	0.851	0.975

## 5. Proofs

### 5.1. Auxiliary Results

**Lemma 1.** *Under Assumption 2, there is a constant  $K_d < \infty$ , such that for all  $x_1, \dots, x_l \in \mathbb{R}$  with  $\text{Var}(\sum_{i=1}^l x_i \xi_i) = 1$*

$$\sum_{i=1}^l x_i^2 \leq K_d.$$

*Proof.* Recall that we can rewrite the covariances as

$$\gamma(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda) d\lambda$$

and that the spectral density  $f$  can be written as  $f(\lambda) = L_f(|\lambda|)|\lambda|^{D-1}$ . As by our assumptions  $L_f(x) \geq C_{\min}$  for a constant  $C_{\min} > 0$ , we can conclude that

$$\begin{aligned} 1 &= \text{Var}\left(\sum_{i=1}^l x_i \xi_i\right) = \sum_{j,k=1}^l x_j x_k \gamma(j-k) \\ &= \sum_{j,k=1}^l x_j x_k \int_{-\pi}^{\pi} e^{i(j-k)\lambda} f(\lambda) d\lambda = \sum_{j,k=1}^l x_j x_k \int_{-\pi}^{\pi} e^{i(j-k)\lambda} L_f(|\lambda|)|\lambda|^{D-1} d\lambda \\ &= 2 \int_0^{\pi} \sum_{j,k=1}^l x_j x_k e^{i(j-k)\lambda} L_f(\lambda) \lambda^{D-1} d\lambda = 2 \int_0^{\pi} \left| \sum_{j=1}^l x_j e^{-ij\lambda} \right|^2 L_f(\lambda) \lambda^{D-1} d\lambda \\ &\geq 2C_{\min} \pi^{D-1} \int_0^{\pi} \left| \sum_{j=1}^l x_j e^{-ij\lambda} \right|^2 d\lambda. \end{aligned}$$

Now we rewrite the integrand as

$$\begin{aligned} \left| \sum_{j=1}^l x_j e^{-ij\lambda} \right|^2 &= \sum_{j,k=1}^l x_j x_k e^{-ij\lambda} e^{ik\lambda} = \sum_{j=1}^l x_j^2 + \sum_{j \neq k} x_j x_k e^{-i(j-k)\lambda} \\ &= \sum_{j=1}^l x_j^2 + \sum_{j < k} x_j x_k \left( e^{-i(j-k)\lambda} + e^{-i(k-j)\lambda} \right) \\ &= \sum_{j=1}^l x_j^2 + 2 \sum_{j < k} x_j x_k \cos((k-j)\lambda) = \sum_{j,k=1}^l x_j x_k \cos((k-j)\lambda). \end{aligned}$$

We use this to calculate the integral:

$$\int_0^{\pi} \left| \sum_{j=1}^l x_j e^{-ij\lambda} \right|^2 d\lambda = \int_0^{\pi} \sum_{j,k=1}^l x_j x_k \cos((k-j)\lambda) d\lambda$$

$$\begin{aligned}
&= \sum_{j,k=1}^l x_j x_k \int_0^\pi \cos((k-j)\lambda) d\lambda \\
&= \sum_{j=1}^l x_j^2 \int_0^\pi \cos(0) d\lambda + \sum_{j \neq k} x_j x_k \underbrace{\int_0^\pi \cos((k-j)\lambda) d\lambda}_{=0} \\
&= \pi \sum_{j=1}^l x_j^2.
\end{aligned}$$

All in all this yields

$$1 = \text{Var}\left(\sum_{i=1}^l x_i \xi_i\right) \geq 2C_{\min} \pi^{D-1} \int_0^\pi \left| \sum_{j=1}^l x_j e^{-ij\lambda} \right|^2 d\lambda = 2C_{\min} \pi^D \sum_{j=1}^l x_j^2,$$

so the statement of the lemma holds with  $K_d = 1/(2C_{\min} \pi^D)$ .  $\square$

**Lemma 2.** *Under Assumption 2, there are constants  $K'_d < \infty, l_0 \in \mathbb{N}$ , such that for all  $l \geq l_0$  and  $x_1, \dots, x_l \in \mathbb{R}$  with  $\text{Var}(\sum_{i=1}^l x_i \xi_i) = 1$*

$$\left| \sum_{i=1}^l x_i \right| \leq K'_d l^{D/2}.$$

*Proof.* The statement of the proof is equivalent to existence of a constant  $C > 0$ , such that for all  $x_1, \dots, x_l \in \mathbb{R}$  with  $\sum_{i=1}^l x_i = 1$ , we have that

$$\text{Var}\left(\sum_{i=1}^l x_i \xi_i\right) \geq C l^{-D}.$$

Let  $x_1^*, \dots, x_l^* \in \mathbb{R}$  with  $\sum_{i=1}^l x_i^* = 1$  be the values that minimize  $\text{Var}(\sum_{i=1}^l x_i^* \xi_i)$ . Then  $\hat{\mu}_\xi(\xi_1, \dots, \xi_n) := \sum_{i=1}^l x_i^* \xi_i$  is the best linear unbiased estimator for  $\mu := E[\xi_1]$ . For a process  $(\zeta_n)_{n \in \mathbb{N}}$  with spectral density

$$f_\zeta(x) = \frac{1}{2\pi} |1 - e^{ix}|^{(D-1)/2}$$

we have by Corollary 5.2 of Adenstedt [1] that

$$\text{Var}(\hat{\mu}_\zeta(\zeta_1, \dots, \zeta_n)) \geq C_1 l^{-D}$$

for a constant  $C_1 > 0$ . Now we rewrite the spectral density  $f_\zeta$  of  $(\zeta_n)_{n \in \mathbb{N}}$  with the help of the spectral density  $f$  of  $(\xi_n)_{n \in \mathbb{N}}$  as

$$f_\zeta(x) = f(x) \frac{|1 - e^{ix}|^{(D-1)/2}}{2\pi |x|^{D-1} L_f(x)}$$

and observe that the function  $g$  with  $g(x) = \frac{|1 - e^{ix}|^{(D-1)/2}}{2\pi|x|^{D-1}L_f(x)}$  is bounded, as we assumed that  $L_f$  is bounded away from 0. So by Lemma 4.4 of [1], we have for all  $l \geq l_0$  that

$$\text{Var}(\hat{\mu}_\xi(\xi_1, \dots, \xi_n)) \geq \frac{1}{2g(0)} \text{Var}(\hat{\mu}_\zeta(\zeta_1, \dots, \zeta_n)) \geq Cl^{-D}.$$

□

Our next Lemma deals with the  $\rho$ -mixing coefficient, which is defined the following way: Let  $\mathcal{A}, \mathcal{B}$  be two  $\sigma$ -field, then

$$\rho(\mathcal{A}, \mathcal{B}) := \sup \text{corr}(X, Y),$$

where the supremum is taken over all  $\mathcal{A}$ -measurable random variables  $X$  and all  $\mathcal{B}$ -measurable random variables  $Y$ . For details, we recommend the book of Bradley [13].

**Lemma 3.** *Under Assumption 2, there are constants  $C_1, C_2 < \infty$ , such that*

$$\begin{aligned} \rho(k, l) &:= \rho(\sigma(\xi_i, 1 \leq i \leq l), \sigma(\xi_j, k+l+1 \leq j \leq k+2l)) \\ &\leq C_1 (k/l)^{-D} L_\gamma(k) + C_2 l^2 k^{-D-1} \max\{L_\gamma(k), 1\} \end{aligned}$$

for all  $k \in \mathbb{N}$  and all  $l \in \{l_k, \dots, k\}$ .

*Proof.* Kolmogorov and Rozanov [26] have proved that there exist real numbers  $a_1, a_2, \dots, a_l, b_1, b_2, \dots, b_l$ , such that

$$\rho(\sigma(\xi_i, 1 \leq i \leq l), \sigma(\xi_j, k+l+1 \leq j \leq k+2l)) = \text{Cov}\left(\sum_{i=1}^l a_i \xi_i, \sum_{j=1}^l b_j \xi_{k+l+j}\right).$$

and  $\text{Var}(\sum_{i=1}^l a_i \xi_i) = \text{Var}(\sum_{j=1}^l b_j \xi_{k+l+j}) = 1$ . The triangular inequality yields

$$\begin{aligned} &\left| \text{Cov}\left(\sum_{i=1}^l a_i \xi_i, \sum_{j=1}^l b_j \xi_{k+l+j}\right) \right| \\ &\leq \left| \sum_{i=1}^l a_i \sum_{j=1}^l b_j \right| |\gamma(k)| + \sum_{i=1}^l \sum_{j=1}^l |a_i| |b_j| |\gamma(k) - \gamma(k+l+j-i)|. \end{aligned}$$

We will treat these two summand separately. For the first part, we use Lemma 2 to obtain

$$\left| \sum_{i=1}^l a_i \sum_{j=1}^l b_j \right| |\gamma(k)| = \left| \sum_{i=1}^l a_i \right| \left| \sum_{j=1}^l b_j \right| |\gamma(k)| \leq K_d'^2 l^D L_\gamma(k) k^{-D}.$$

Before we deal with the second summand, we first observe that by Hölder's inequality and Lemma 1

$$\sum_{i=1}^l |a_i| \leq \sqrt{l \sum_{i=1}^l a_i^2} \leq \sqrt{K_d} \sqrt{l} \quad \text{and} \quad \sum_{j=1}^l |b_j| \leq \sqrt{l \sum_{j=1}^l b_j^2} \leq \sqrt{K_d} \sqrt{l}.$$

Furthermore, we have by our assumption

$$\sup_{|k-\tilde{k}| \leq 2l-1} \left| L_\gamma(k) - L_\gamma(\tilde{k}) \right| \leq K \frac{l}{k}$$

for a constant  $K$  and consequently for all  $\tilde{k} \in \{k+1, \dots, k+2l-1\}$

$$\begin{aligned} \left| \gamma(k) - \gamma(\tilde{k}) \right| &\leq L_\gamma(k) \left| k^{-D} - \tilde{k}^{-D} \right| + |L_\gamma(k) - L_\gamma(\tilde{k})| \tilde{k}^{-D} \\ &\leq L_\gamma(k) (k^{-D} - (k+2l-1)^{-D}) + |L_\gamma(k) - L_\gamma(\tilde{k})| k^{-D} \\ &\leq C_d k^{-D-1} l L_\gamma(k) + K \frac{l}{k} k^{-D} \max\{L_\gamma(k), 1\} \\ &\leq C_3 k^{-D-1} l \max\{L_\gamma(k), 1\} \end{aligned}$$

for some constants  $C_d, C_3$ . Combining this with the bounds for  $\sum_{i=1}^l |a_i|$ ,  $\sum_{j=1}^l |b_j|$ , we finally arrive at

$$\begin{aligned} \sum_{i=1}^l |a_i| \sum_{j=1}^l |b_j| |\gamma(k) - \gamma(k+l+j-i)| &\leq K_d l \max_{\tilde{k} \in \{k+1, \dots, k+2l-1\}} \left| \gamma(k) - \gamma(\tilde{k}) \right| \\ &= K_d C_3 k^{-D-1} l^2 \max\{L_\gamma(k), 1\}. \end{aligned}$$

□

## 5.2. Proof of the Main Result

Let  $t$  be a point of continuity of  $F_T$ . In order to simplify notation, we write  $N = n - l + 1$  and  $T_{l,i} = T_l(X_i, \dots, X_{i+l-1})$ . The triangular inequality yields

$$|\hat{F}_{l,n}(t) - F_{T_n}(t)| \leq |\hat{F}_{l,n}(t) - F_T(t)| + |F_T(t) - F_{T_n}(t)|.$$

The second term on the right-hand side of the above inequality converges to zero because of Assumption 1. As  $L_2$ -convergence implies stochastic convergence, it suffices to show that

$$\mathbb{E} \left( |\hat{F}_{l,n}(t) - F_T(t)|^2 \right) \longrightarrow 0$$

in order to prove that the first term converges to zero, as well. We have

$$\begin{aligned} & \mathbb{E} \left( |\hat{F}_{l,n}(t) - F_T(t)|^2 \right) \\ &= \mathbb{E} \left( \hat{F}_{l,n}^2(t) \right) - \left( \mathbb{E} \hat{F}_{l,n}(t) \right)^2 + (F_T(t))^2 - 2F_T(t) \mathbb{E} \hat{F}_{l,n}(t) + \left( \mathbb{E} \hat{F}_{l,n}(t) \right)^2 \\ &= \text{Var}(\hat{F}_{l,n}(t)) + \left| \mathbb{E} \hat{F}_{l,n}(t) - F_T(t) \right|^2. \end{aligned}$$

Furthermore, the stationarity of the process  $(X_n)_{n \in \mathbb{N}}$  and Assumption 1 imply

$$\mathbb{E} \hat{F}_{l,n}(t) = \frac{1}{N} \sum_{i=1}^N \mathbb{E} (1_{\{T_{l,i} \leq t\}}) = P(T_{l,1} \leq t) = F_{T_l}(t) \xrightarrow{l \rightarrow \infty} F_T(t).$$

It remains to show that  $\text{Var}(\hat{F}_{l,n}(t)) \rightarrow 0$ . Again, it follows by stationarity of  $(X_n)_{n \in \mathbb{N}}$  that

$$\begin{aligned} \text{Var}(\hat{F}_{l,n}(t)) &= \frac{1}{N} \text{Var}(1_{\{T_{l,1} \leq t\}}) + \frac{2}{N^2} \sum_{i=2}^N (N-i+1) \text{Cov}(1_{\{T_{l,1} \leq t\}}, 1_{\{T_{l,i} \leq t\}}) \\ &\leq \frac{2}{N} \sum_{i=1}^N |\text{Cov}(1_{\{T_{l,1} \leq t\}}, 1_{\{T_{l,i} \leq t\}})|. \end{aligned}$$

Recall that by Assumption 3, we have  $l \leq C_l n^{(1+D)/2-\epsilon}$  for some constants  $C_l$ ,  $\epsilon > 0$ . For  $n$  large enough such that  $l < \frac{1}{2} \lfloor n^{1-\epsilon/2} \rfloor$ , we split the sum of covariances into two parts:

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N |\text{Cov}(1_{\{T_{l,1} \leq t\}}, 1_{\{T_{l,i} \leq t\}})| \\ &= \frac{1}{N} \sum_{i=1}^{\lfloor n^{1-\epsilon/2} \rfloor} |\text{Cov}(1_{\{T_{l,1} \leq t\}}, 1_{\{T_{l,i} \leq t\}})| + \frac{1}{N} \sum_{i=\lfloor n^{1-\epsilon/2} \rfloor + 1}^N |\text{Cov}(1_{\{T_{l,1} \leq t\}}, 1_{\{T_{l,i} \leq t\}})| \\ &\leq \frac{\lfloor n^{1-\epsilon/2} \rfloor}{N} + \frac{1}{N} \sum_{k=\lfloor n^{1-\epsilon/2} \rfloor + 1}^N \rho(\sigma(X_i, 1 \leq i \leq l), \sigma(X_j, k \leq j \leq k+l-1)) \\ &\leq \frac{\lfloor n^{1-\epsilon/2} \rfloor}{N} + \frac{1}{N} \sum_{k=\lfloor n^{1-\epsilon/2} \rfloor - l}^{N-l-1} \rho(k, l), \end{aligned}$$

where

$$\rho(k, l) := \rho(\sigma(X_i, 1 \leq i \leq l), \sigma(X_j, k+l+1 \leq j \leq k+2l)).$$

Obviously, the first summand  $\lfloor n^{1-\epsilon/2} \rfloor / N$  converges to zero by Assumption 3. For the second summand, first note that as a consequence of Potter's Theorem

(Theorem 1.5.6 in Bingham et al. [12]), there is a constant  $C_L$  such that  $L_\gamma(k) \leq C_L k^{D\epsilon/2}$  for all  $k \in \mathbb{N}$ . This together with Lemma 3 yields

$$\begin{aligned}
& \frac{1}{N} \sum_{k=\lfloor n^{1-\epsilon/2} \rfloor - l}^{N-l-1} \rho(k, l) \\
& \leq C_L C_1 \frac{l^D}{N} \sum_{k=\lfloor n^{1-\epsilon/2} \rfloor / 2}^{N-l-1} k^{-D} k^{D\epsilon/2} + C_L C_2 \frac{l^2}{N} \sum_{k=\lfloor n^{1-\epsilon/2} \rfloor / 2}^{N-l-1} k^{-D-1} k^{D\epsilon/2} \\
& \leq C_L C_1 C_l^D 2^{D(1-\epsilon/2)} n^{D((1+D)/2-\epsilon)-(1-\epsilon/2)+\epsilon/2(1-\epsilon/2)} \\
& \quad + C_L C_2 C_l^2 2^{1+D(1-\epsilon/2)} n^{((1+D-2\epsilon)-(D+1)(1-\epsilon/2)+(1-\epsilon/2)D\epsilon/2)} \\
& \leq C \left( n^{-D((1-D)/2+\epsilon^2/4)} + n^{-\epsilon(\frac{3}{2}-D+D\epsilon/4)} \right) \xrightarrow{n \rightarrow \infty} 0
\end{aligned}$$

for some constant  $C < \infty$ . Thus, we have proved that  $\text{Var}(\hat{F}_{l,n}(t)) \rightarrow 0$  as  $n \rightarrow \infty$  and that the first conjecture of Theorem 1 holds.

The second assertion of Theorem 1 follows from

$$F_{T_n}(t) - \hat{F}_{l,n}(t) \xrightarrow{\mathcal{P}} 0$$

by the usual Glivenko-Cantelli argument for the uniform convergence of empirical distribution functions, see for example section 20 in the book of Billingsley [11].  $\square$

## References

- [1] ADENSTEDT, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *The Annals of Statistics* 1095–1107.
- [2] ANDREWS, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856.
- [3] ANTOCH, J., HUŠKOVÁ, M., JANIC, A. and LEDWINA, T. (2008). Data driven rank test for the change point problem. *Metrika* **68** 1–15.
- [4] BAEK, C., PIPIRAS, V. et al. (2014). On distinguishing multiple changes in mean and long-range dependence using local Whittle estimation. *Electronic Journal of Statistics* **8** 931–964.
- [5] BALKE, N. S. (1993). Detecting level shifts in time series. *Journal of Business & Economic Statistics* **11** 81–92.
- [6] BERAN, J. and FENG, Y. (2002). SEMIFAR models - a semiparametric framework for modelling trends, long-range dependence and nonstationarity. *Computational Statistics & Data Analysis* **40** 393–419.
- [7] BERAN, J., FENG, Y., GHOSH, S. and KULIK, R. (2013). *Long-memory processes*. Springer-Verlag Berlin Heidelberg.
- [8] BERKES, I., HORVÁTH, L., KOKOSZKA, P. and SHAO, Q.-M. (2006). On discriminating between long-range dependence and changes in mean. *The Annals of Statistics* **34** 1140–1165.

- [9] BETKEN, A. (2014). Testing for change-points in long-range dependent time series by means of a self-normalized Wilcoxon test. *arXiv preprint arXiv:1403.0265*.
- [10] BEUTNER, E. and ZÄHLE, H. (2014). Continuous mapping approach to the asymptotics of  $U$ - and  $V$ -statistics. *Bernoulli* **20** 846–877.
- [11] BILLINGSLEY, P. (1995). *Probability and measure*. John Wiley & Sons, Inc.
- [12] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular variation*. Cambridge University Press.
- [13] BRADLEY, R. C. (2007). *Introduction to strong mixing conditions*. Kendrick press.
- [14] COBB, G. W. (1978). The problem of the Nile: conditional solution to a changepoint problem. *Biometrika* **65** 243 - 251.
- [15] CSÖRGÖ, M. and HORVÁTH, L. (1997). *Limit theorems in change-point analysis*. Wiley Chichester; New York.
- [16] DEHLING, H., ROOCH, A. and TAQQU, M. S. (2013). Non-parametric change-point tests for long-range dependent data. *Scandinavian Journal of Statistics* **40** 153 - 173.
- [17] DEO, R. S. and HURVICH, C. M. (1998). Linear trend with fractionally integrated errors. *Journal of Time Series Analysis* **19** 379–397.
- [18] DOBRUSHIN, R. L. and MAJOR, P. (1979). Non-central limit theorems for non-linear functionals of Gaussian fields. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50** 27–52.
- [19] FAN, Z. (2012). *Statistical issues and developments in time series analysis and educational measurement*. BiblioBazaar.
- [20] GRANGER, C. W. and JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis* **1** 15–29.
- [21] HALL, P. and JING, B. (1996). On sample reuse methods for dependent data. *Journal of the Royal Statistical Society. Series B (Methodological)* 727–737.
- [22] HALL, P., JING, B.-Y. and LAHIRI, S. N. (1998). On the sampling window method for long-range dependent data. *Statistica Sinica* **8** 1189–1204.
- [23] HURST, H. E. (1956). Methods of using long-term storage in reservoirs. In *ICE Proceedings* **5** 519–543. Thomas Telford.
- [24] IBRAGIMOV, I. A. and ROZANOV, Y. A. (1978). *Gaussian random processes*. Springer, New York.
- [25] JACH, A., MCELROY, T. and POLITIS, D. N. (2012). Subsampling inference for the mean of heavy-tailed long-memory time series. *Journal of Time Series Analysis* **33** 96–111.
- [26] KOLMOGOROV, A. and ROZANOV, Y. A. (1960). On strong mixing conditions for stationary Gaussian processes. *Theory of Probability & Its Applications* **5** 204–208.
- [27] KÜNSCH, H. R. (1987). Statistical aspects of self-similar processes. In *Proceedings of the first World Congress of the Bernoulli Society* **1** 67–74. VNU Science Press Utrecht, The Netherlands.
- [28] LAHIRI, S. N. (1993). On the moving block bootstrap under long range



- dependence. *Statistics & Probability Letters* **18** 405–413.
- [29] LELAND, W. E., TAQQU, M. S., WILLINGER, W. and WILSON, D. V. (1994). On the self-similar nature of Ethernet traffic (extended version). *Networking, IEEE/ACM Transactions on* **2** 1–15.
- [30] LO, A. W. (1989). Long-term memory in stock market prices Technical Report, National Bureau of Economic Research.
- [31] LOBATO, I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association* **96** 1066–1076.
- [32] MACNEILL, I. B., TANG, S. M. and JANDHYALA, V. K. (1991). A search for the source of the Nile's change-points. *Environmetrics* **2** 341 – 375.
- [33] MCELROY, T. and POLITIS, D. (2007). Self-normalization for heavy-tailed time series with long memory. *Statistica Sinica* **17** 199.
- [34] NORDMAN, D. J. and LAHIRI, S. N. (2005). Validity of the sampling window method for long-range dependent linear processes. *Econometric Theory* **21** 1087–1111.
- [35] PIPIRAS, V. and TAQQU, M. S. (2011). *Long-range dependence and self-similarity*. Cambridge University Press.
- [36] POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics* 2031–2050.
- [37] ROBINSON, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics* 1630–1661.
- [38] SHAO, X. (2011). A simple test of changes in mean in the possible presence of long-range dependence. *Journal of Time Series Analysis* **32** 598–606.
- [39] SHAO, X. and ZHANG, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association* **105** 1228–1240.
- [40] SHERMAN, M. and CARLSTEIN, E. (1996). Replicate histograms. *Journal of the American Statistical Association* **91** 566–576.
- [41] SINAI, Y. G. (1976). Self-similar probability distributions. *Theory of Probability & Its Applications* **21** 64–80.
- [42] SURGAILIS, D. (1982). Zones of attraction of self-similar multiple integrals. *Lithuanian Mathematical Journal* **22** 327–340.
- [43] TAQQU, M. S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50** 53–83.
- [44] WANG, L. (2007). Gradual changes in long memory processes with applications. *Statistics* **41** 221–240.
- [45] WANG, L. (2008). Change-point detection with rank statistics in long-memory time-series models. *Australian & New Zealand Journal of Statistics* **50** 241–256.
- [46] WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69** 391 – 410.
- [47] ZHANG, T., HO, H.-C., WENDLER, M. and WU, W. B. (2013). Block sampling under strong dependence. *Stochastic Processes and their Applications* **123** 2323–2339.





