

# Nonparametric Change-Point Tests for Long-Range Dependent Data

HEROLD DEHLING

Fakultät für Mathematik, Ruhr-Universität Bochum

AENEAS ROOCH

Fakultät für Mathematik, Ruhr-Universität Bochum

MURAD S. TAQQU

Department of Mathematics and Statistics, Boston University

March 2, 2012

## Abstract

We propose a nonparametric change-point test for long-range dependent data, which is based on the Wilcoxon two-sample test. We derive the asymptotic distribution of the test statistic under the null-hypothesis that no change occurred. In a simulation study, we compare the power of our test with the power of a test which is based on differences of means. The results of the simulation study show that in the case of Gaussian data, our test has only slightly smaller power than the difference-of-means test. For heavy-tailed data, our test outperforms the difference-of-means test.

**Running Head:** Change-Point Tests for LRD Data

**Keywords:** Change-point problem, functional limit theorem, long memory, long-range dependent data, nonparametric test, Wilcoxon two-sample rank test.

# 1 Introduction and Statement of Main Results

In this paper we study change-point problems in long-range dependent time series. Specifically we consider the model where the observations are generated by a stochastic process  $(X_i)_{i \geq 1}$  of the type

$$X_i = \mu_i + \epsilon_i,$$

where  $(\epsilon_i)_{i \geq 1}$  is a long-range dependent stationary process with mean zero, and where  $(\mu_i)_{i \geq 1}$  are the unknown means. We focus on the case when  $(\epsilon_i)_{i \geq 1}$  is an instantaneous functional of a stationary Gaussian process with non-summable covariances, i.e.

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

We assume that  $(\xi_i)_{i \geq 1}$  is a mean-zero Gaussian process with  $E(\xi_i^2) = 1$  and long-range dependence, that is, with autocovariance function

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

where  $0 < D < 1$ , and where  $L(k)$  is a slowly varying function. Moreover,  $G : \mathbb{R} \rightarrow \mathbb{R}$  is a measurable function satisfying  $E(G(\xi_i)) = 0$ .

Based on observations  $X_1, \dots, X_n$ , we wish to test the hypothesis

$$H : \mu_1 = \dots = \mu_n$$

that there is no change in the means of the data against the alternative

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

We shall refer to this test problem as  $(H, A)$ .

In the case of independent observations, change-point problems have been much investigated, and both parametric as well as nonparametric tests have been studied. For an excellent survey, see the monograph by Csörgő and Horváth (1997). There are also many results for weakly dependent observations. E.g., Ling (2007) studies the asymptotic distribution of the Wald test for parameter changes in near epoch dependent processes. Aue, Hörmann, Horváth and Reimherr (2009) study tests for detecting breaks in the covariance structure of multivariate time series. Wied, Krämer and Dehling (2011) study change-points in the correlation structure of near epoch dependent processes.

In the case of long-range dependent data, much less is known. Csörgő and Horváth (1997) study so-called *cusum tests* (cumulative sum tests) for changes in the mean, which are based

on statistics of the type

$$\max_{1 \leq k \leq n} \left| \frac{1}{k} \sum_{i=1}^k X_i - \frac{1}{n} \sum_{i=1}^n X_i \right|.$$

Krämer, Sibbertsen and Kleiber (2002) observe that the presence of long-range dependence might lead to false rejection of the hypothesis of stationarity in change-point tests. Berkes, Horváth, Kokoszka and Shao (2006) propose cusum type tests for discriminating between long-range dependence and changes in the mean of a weakly dependent process. Giraitis, Leipus and Surgailis (1996) consider change-point problems for linear processes where they test the hypothesis of stationarity against the alternative of some break. Their tests are based on supremum functionals of the empirical process. Horvath and Kokoszka (1997) study cusum tests for change points for Gaussian processes. Wang (2003) studies cusum tests for change-points in multivariate processes that are subordinated to a Gaussian process. Wang (2008b) studies cusum tests for linear processes.

In the present paper we want to investigate rank tests for changes in the mean. Antoch, Hušková, Janic and Ledwina (2008) studied rank tests for i.i.d. observations. Wang (2008a) investigated rank tests for change-points in long-range dependent linear processes. In this paper, we study rank tests for changes in the mean of long-range dependent processes that have a representation as instantaneous nonlinear functionals of stationary Gaussian processes.

Tests for change-point problems usually start from the two-sample problem that one obtains when the change-point is regarded as being known, that is, where for a given  $k \in \{1, \dots, n-1\}$ , the alternative is

$$A_k : \mu_1 = \dots = \mu_k \neq \mu_{k+1} = \dots = \mu_n.$$

For the test problem  $(H, A_k)$ , the Wilcoxon two-sample rank test is a commonly used non-parametric test. The Wilcoxon test rejects for large and small values of the test statistic

$$W_{k,n} = \sum_{i=1}^k \sum_{j=k+1}^n 1_{\{X_i \leq X_j\}}.$$

$W_{k,n}$  counts the number of times the second part of the sample exceeds the first part of the sample. Note that this is the Mann-Whitney  $U$ -statistic representation of Wilcoxon's rank test statistic; see e.g. Lehmann (1975).

Tests for the change-point problem  $(H, A)$ , where one does not suppose that  $k$  is known, can be based on the vector of test statistics  $W_{1,n}, \dots, W_{n-1,n}$ . In a parametric model, the generalized likelihood ratio test rejects for large values of the maximum of the test statistics

obtained for the two-sample problems  $(H, A_k)$ ; see e.g. Csörgő and Horvath (1997). Also in the nonparametric case, a common procedure is to reject the null hypothesis for large values of  $W_n = \max_{k=1, \dots, n-1} |W_{k,n}|$ , though one could in principle take other functions of  $W_{1,n}, \dots, W_{n-1,n}$ .

In order to set the critical values of the test, we need to know the distribution of  $W_n$  under the null hypothesis of no change. Since the exact distribution is hard to obtain, we consider the asymptotic distribution for a large sample size. To this end we consider the process

$$W_{n, [n\lambda]} = \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n 1_{\{X_i \leq X_j\}}, \quad 0 \leq \lambda \leq 1,$$

parametrized by  $\lambda$ . After centering and normalizing, we obtain the process

$$W_n(\lambda) = \frac{1}{n d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n \left( 1_{\{X_i \leq X_j\}} - \int_{\mathbb{R}} F(x) dF(x) \right), \quad 0 \leq \lambda \leq 1. \quad (2)$$

Observe that the centering constant  $\int_{\mathbb{R}} F(x) dF(x)$  equals  $E(1_{\{X_1 \leq X'_1\}})$ , where  $X'_1$  is an independent copy of  $X_1$ . The latter is the proper normalization as the dependence between  $X_i$  and  $X_j$  vanishes asymptotically when  $|j - i| \rightarrow \infty$ . If the distribution function  $F$  is continuous, we get  $\int_{\mathbb{R}} F(x) dF(x) = \frac{1}{2}$ . Note that under the null hypothesis, the distribution of  $W_n(\lambda)$  does not depend on the common mean  $\mu := \mu_1 = \dots = \mu_n$ . Thus, we may assume without loss of generality that  $\mu = 0$  and hence that  $X_i = G(\xi_i)$ .

In order to analyze the asymptotic distribution of the process  $(W_n(\lambda))_{0 \leq \lambda \leq 1}$ , we apply the empirical process invariance principle of Dehling and Taqqu (1989) to the sequence  $(G(\xi_i))_{i \geq 1}$ . We consider the Hermite expansion

$$1_{\{G(\xi_i) \leq x\}} - F(x) = \sum_{q=1}^{\infty} \frac{J_q(x)}{q!} H_q(\xi_i),$$

where  $H_q$  is the  $q$ -th order Hermite polynomial and where

$$J_q(x) = E H_q(\xi_i) 1_{\{G(\xi_i) \leq x\}}.$$

We define the Hermite rank of the class of functions  $\{1_{\{G(\xi_i) \leq x\}} - F(x), x \in \mathbb{R}\}$ , by

$$m := \min\{q \geq 1 : J_q(x) \neq 0 \text{ for some } x \in \mathbb{R}\}, \quad (3)$$

and the normalizing constants

$$d_n^2 = \text{Var} \left( \sum_{j=1}^n H_m(\xi_j) \right). \quad (4)$$

We make the further assumption that  $0 < D < \frac{1}{m}$ , in which case

$$d_n^2 \sim c_m n^{2-mD} L^m(n), \quad (5)$$

where  $c_m = \frac{2m!}{(1-Dm)(2-Dm)}$ . Here we use the symbol  $a_n \sim b_n$  to denote that  $a_n/b_n \rightarrow 1$  as  $n \rightarrow \infty$ . Then by the two-parameter empirical process invariance principle of Dehling and Taqqu (1989),

$$(d_n^{-1}[n\lambda](F_{[n\lambda]}(x) - F(x)))_{x \in [-\infty, \infty], \lambda \in [0, 1]} \rightarrow \left( \frac{J_m(x)}{m!} Z_m(\lambda) \right)_{x \in [-\infty, \infty], \lambda \in [0, 1]},$$

weakly in  $D([-\infty, \infty] \times [0, 1])$ , where  $(Z_m(\lambda))_{\lambda \in [0, 1]}$  is an  $m$ -th order Hermite process, defined in Taqqu (1979), formula (1.3). A spectral representation of  $Z_m$  is given in formula (1.7) of Dehling and Taqqu (1989). Observe the separation of variables in the limit: the limiting process  $Z_m(x, \lambda)$  is expressible as  $\frac{J_m(x)}{m!} Z_m(\lambda)$ .

The process  $(Z_m(\lambda))_{\lambda \geq 0}$  is self-similar with parameter

$$H = 1 - \frac{mD}{2} \in \left(\frac{1}{2}, 1\right),$$

that is,  $Z_m(c\lambda)$  and  $c^H Z_m(\lambda)$  have the same finite-dimensional distributions for all constants  $c > 0$ .<sup>1</sup> The process  $Z_m(\lambda)$  is not Gaussian when  $m \geq 2$ . When  $m = 1$ ,  $Z_1(\lambda)$  is the standard Gaussian fractional Brownian motion, often denoted  $B_H(\lambda)$ . It is Gaussian with mean zero and covariance

$$E(Z_1(\lambda_1) Z_1(\lambda_2)) = \frac{1}{2} \{ \lambda_1^{2H} + \lambda_2^{2H} - |\lambda_1 - \lambda_2|^{2H} \} \quad (6)$$

**Theorem 1** *Suppose that  $(\xi_i)_{i \geq 1}$  is a stationary Gaussian process with mean zero, variance 1, and autocovariance function as in (1) with  $0 < D < \frac{1}{m}$ . Moreover let  $G : \mathbb{R} \rightarrow \mathbb{R}$  be a measurable function and define*

$$X_k = G(\xi_k).$$

*Assume that  $X_k$  has a continuous distribution function  $F$ . Let  $m$  denote the Hermite rank of the class of functions  $1_{\{G(\xi_i) \leq x\}} - F(x)$ ,  $x \in \mathbb{R}$ , as defined in (3) and let  $d_n > 0$  satisfy (5). Then*

$$\frac{1}{n d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n \left( 1_{\{X_i \leq X_j\}} - \int_{\mathbb{R}} F(x) dF(x) \right), \quad 0 \leq \lambda \leq 1, \quad (7)$$

*converges in distribution towards the process*

$$\frac{1}{m!} (Z_m(\lambda) - \lambda Z_m(1)) \int_{\mathbb{R}} J_m(x) dF(x), \quad 0 \leq \lambda \leq 1.$$

---

<sup>1</sup>Do not confuse the index  $H$ , which is called *Hurst parameter* with the other  $H$ 's used in this paper to denote *hypothesis* and *Hermite polynomial*.

## 2 Proof of Theorem 1

We introduce the empirical distribution functions of the first  $k$  and the last  $(n - k)$  observations, respectively:

$$\begin{aligned} F_k(x) &= \frac{1}{k} \sum_{i=1}^k 1_{\{X_i \leq x\}} \\ F_{k+1,n}(x) &= \frac{1}{n - k} \sum_{i=k+1}^n 1_{\{X_i \leq x\}} \end{aligned}$$

so that  $[n\lambda]F_{[n\lambda]}(x) = \sum_{i=1}^{[n\lambda]} 1_{\{X_i \leq x\}}$  and  $(n - [n\lambda])F_{[n\lambda]+1,n} = \sum_{i=[n\lambda]+1}^n 1_{\{X_i \leq x\}}$ . Then we obtain the following representation

$$\begin{aligned} & \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n \left( 1_{\{X_i \leq X_j\}} - \int_{\mathbb{R}} F(x) dF(x) \right) \\ &= [n\lambda] \sum_{j=[n\lambda]+1}^n \left( F_{[n\lambda]}(X_j) - \int_{\mathbb{R}} F(x) dF(x) \right) \\ &= [n\lambda](n - [n\lambda]) \left( \int_{\mathbb{R}} F_{[n\lambda]}(x) dF_{[n\lambda]+1,n}(x) - \int_{\mathbb{R}} F(x) dF(x) \right) \\ &= [n\lambda](n - [n\lambda]) \int_{\mathbb{R}} (F_{[n\lambda]}(x) - F(x)) dF_{[n\lambda]+1,n}(x) \\ &\quad + [n\lambda](n - [n\lambda]) \int_{\mathbb{R}} F(x) d(F_{[n\lambda]+1,n} - F)(x) \\ &= [n\lambda](n - [n\lambda]) \int_{\mathbb{R}} (F_{[n\lambda]}(x) - F(x)) dF_{[n\lambda]+1,n}(x) \\ &\quad - [n\lambda](n - [n\lambda]) \int_{\mathbb{R}} (F_{[n\lambda]+1,n}(x) - F(x)) dF(x), \end{aligned} \tag{8}$$

where in the third line we used the relation  $E(K(X)) = \int_{\mathbb{R}} K(x) dF(x)$ , where  $F$  and  $E$  are respectively, the empirical distribution and the empirical mean of  $X$ , and where, in the final step, we have used integration by parts, namely  $\int_{\mathbb{R}} G dF = 1 - \int_{\mathbb{R}} F dG$ , if  $F$  and  $G$  are two distribution functions. We now apply the empirical process non-central limit theorem of Dehling and Taqqu (1989), which states that in the space  $D([-\infty, \infty] \times [0, 1])$ , equipped with the supremum norm,

$$\left( d_n^{-1} [n\lambda] (F_{n\lambda}(x) - F(x)) \right)_{x \in [-\infty, \infty], \lambda \in [0, 1]} \xrightarrow{\mathcal{D}} (J(x)Z(\lambda))_{x \in [-\infty, \infty], \lambda \in [0, 1]}, \tag{9}$$

where

$$J(x) = J_m(x) \quad \text{and} \quad Z(\lambda) = Z_m(\lambda)/m!.$$

Here,  $\xrightarrow{\mathcal{D}}$  denotes convergence in distribution of random variables with values in the non-separable metric space  $D = D([-\infty, \infty] \times [0, 1])$ , as defined e.g. in Shorack and Wellner (1986), Section 2.3. Given  $D$ -valued random variables  $(W_n)_{n \geq 1}$  and  $W$ , we say that  $W_n$  converges in distribution to  $W$  if

$$Ef(W_n) \rightarrow Ef(W)$$

holds for all functions  $f : D \rightarrow \mathbb{R}$  that are bounded, uniformly continuous with respect to the supremum norm on  $D$  and measurable with respect to the  $\sigma$ -field  $\mathcal{D}^B$  on  $D$  that is generated by the open balls. We can then apply the Skorohod-Dudley-Wichura representation theorem; see Theorem 2.3.4 in Shorack and Wellner (1986). Note that  $C([-\infty, \infty] \times [0, 1])$  is a separable subset of  $D$ , and that  $P((J(x)Z(\lambda))_{x \in [-\infty, \infty], \lambda \in [0, 1]} \in C([-\infty, \infty] \times [0, 1])) = 1$ , since  $F$  is assumed to be continuous. Thus the conditions of the Skorohod-Dudley-Wichura theorem are met with  $M_s = C([-\infty, \infty] \times [0, 1])$ . Hence we may assume without loss of generality that convergence holds almost surely with respect to the supremum norm on the function space  $D([0, 1] \times [-\infty, \infty])$ , i.e.

$$\sup_{\lambda, x} |d_n^{-1}[n\lambda](F_{n\lambda}(x) - F(x)) - J(x)Z(\lambda)| \longrightarrow 0 \quad a.s. \quad (10)$$

Note that, by definition,

$$(n - [n\lambda])(F_{[n\lambda]+1, n}(x) - F(x)) = n(F_n(x) - F(x)) - [n\lambda](F_{[n\lambda]}(x) - F(x)). \quad (11)$$

Applying (10) to each of the terms on the right-hand side of (11), we obtain the following limit theorem for the two parameter empirical process of the observations  $X_{[n\lambda]+1}, \dots, X_n$ .

$$\sup_{\lambda, x} |d_n^{-1}(n - [n\lambda])(F_{[n\lambda]+1, n}(x) - F(x)) - J(x)(Z(1) - Z(\lambda))| \longrightarrow 0. \quad (12)$$

Thus we get, for the first term in the right-hand side of (8),

$$\begin{aligned} & \frac{1}{n d_n} [n\lambda](n - [n\lambda]) \int_{\mathbb{R}} (F_{[n\lambda]}(x) - F(x)) dF_{[n\lambda]+1, n}(x) - (1 - \lambda) \int_{\mathbb{R}} J(x)Z(\lambda) dF(x) \quad (13) \\ &= \frac{n - [n\lambda]}{n} \int_{\mathbb{R}} d_n^{-1}[n\lambda](F_{[n\lambda]}(x) - F(x)) dF_{[n\lambda]+1, n}(x) - \frac{n - [n\lambda]}{n} \int_{\mathbb{R}} J(x)Z(\lambda) dF(x) \\ & \quad + \left\{ \frac{n - [n\lambda]}{n} - (1 - \lambda) \right\} \int_{\mathbb{R}} J(x)Z(\lambda) dF(x) \\ &= \frac{n - [n\lambda]}{n} \int_{\mathbb{R}} \{d_n^{-1}[n\lambda](F_{[n\lambda]}(x) - F(x)) - J(x)Z(\lambda)\} dF_{[n\lambda]+1, n}(x) \\ & \quad + \frac{n - [n\lambda]}{n} \int_{\mathbb{R}} J(x)Z(\lambda) d(F_{[n\lambda]+1, n} - F)(x) \\ & \quad + \left\{ \frac{n - [n\lambda]}{n} - (1 - \lambda) \right\} \int_{\mathbb{R}} J(x)Z(\lambda) dF(x). \end{aligned}$$

The first term on the right-hand side converges to 0 almost surely, by (10). The third term converges to zero as

$$\sup_{0 \leq \lambda \leq 1} \left| \frac{n - [n\lambda]}{n} - (1 - \lambda) \right| \rightarrow 0.$$

Regarding the second term, note that  $\int_{\mathbb{R}} J(x) dF(x) = E(J(X_i))$  and hence

$$\begin{aligned} & \frac{n - [n\lambda]}{n} \int_{\mathbb{R}} J(x) Z(\lambda) d(F_{[n\lambda]+1, n} - F)(x) \\ &= Z(\lambda) \frac{1}{n} \sum_{i=[n\lambda]+1}^n (J(X_i) - E(J(X_i))) \\ &= Z(\lambda) \left\{ \frac{1}{n} \sum_{i=1}^n (J(X_i) - E(J(X_i))) - \frac{1}{n} \sum_{i=1}^{[n\lambda]} (J(X_i) - E(J(X_i))) \right\}. \end{aligned}$$

By the ergodic theorem,  $\frac{1}{n} \sum_{i=1}^n (J(X_i) - E(J(X_i))) \rightarrow 0$ , almost surely. Thus, given  $\epsilon > 0$ , we can find  $n_0 = n_0(\epsilon, \omega)$  such that, for all  $n \geq n_0$ ,

$$\frac{1}{n} \sum_{i=1}^n (J(X_i) - E(J(X_i))) \leq \epsilon.$$

Hence, for  $n \geq \max\{n_0, \frac{1}{\epsilon} \max_{1 \leq k \leq n_0} |\sum_{i=1}^k (J(X_i) - E(J(X_i)))|\}$ , we have

$$\begin{aligned} & \max_{1 \leq k \leq n} \left| \sum_{i=1}^k (J(X_i) - E(J(X_i))) \right| \\ & \leq \max \left( \max_{1 \leq k \leq n_0} \left| \sum_{i=1}^k (J(X_i) - E(J(X_i))) \right|, \max_{n_0 \leq k \leq n} \left| \sum_{i=1}^k (J(X_i) - E(J(X_i))) \right| \right) \\ & \leq \epsilon n. \end{aligned}$$

and thus

$$\max_{1 \leq k \leq n} \left| \sum_{i=1}^k (J(X_i) - E(J(X_i))) \right| = o(n), \text{ as } n \rightarrow \infty, \quad a.s. \quad (14)$$

Hence (13) converges to zero almost surely, uniformly in  $\lambda \in [0, 1]$ .

Regarding the second term on the right-hand side of (8) we obtain

$$\begin{aligned} & \frac{1}{n d_n} [n\lambda] (n - [n\lambda]) \int_{\mathbb{R}} (F_{[n\lambda]+1, n} - F(x)) dF(x) - \lambda \int_{\mathbb{R}} J(x) (Z(1) - Z(\lambda)) dF(x) \quad (15) \\ &= \frac{[n\lambda]}{n} \int_{\mathbb{R}} \{ d_n^{-1} (n - [n\lambda]) (F_{[\lambda n]+1, n}(x) - F(x)) - J(x) (Z(1) - Z(\lambda)) \} dF(x) \\ & \quad - \left( \lambda - \frac{[n\lambda]}{n} \right) \int_{\mathbb{R}} J(x) (Z(1) - Z(\lambda)) dF(x). \end{aligned}$$



Both terms on the right-hand side converge to zero a.s., uniformly in  $\lambda \in [0, 1]$ . For the first term, this follows from (12). For the second term, this holds since  $\sup_{0 \leq \lambda \leq 1} \left| \frac{[n\lambda]}{n} - \lambda \right| \rightarrow 0$ , as  $n \rightarrow \infty$ . Using (8) and the fact that the right-hand sides of (13) and (15) converge to zero uniformly in  $0 \leq \lambda \leq 1$ , we have proved that the normalized Wilcoxon two-sample test statistic (7) converges in distribution towards

$$\int_{\mathbb{R}} (1 - \lambda)Z(\lambda)J(x)dF(x) - \int_{\mathbb{R}} \lambda(Z(1) - Z(\lambda))J(x)dF(x), \quad 0 \leq \lambda \leq 1.$$

Finally, we observe that

$$\int_{\mathbb{R}} (1 - \lambda)Z(\lambda)J(x)dF(x) - \int_{\mathbb{R}} \lambda(Z(1) - Z(\lambda))J(x)dF(x) = (Z(\lambda) - \lambda Z(1)) \int_{\mathbb{R}} J(x)dF(x),$$

and thus we have established Theorem 1.  $\square$

**Remark 1** (1) In our proof we have used an integration by parts in order to express our test statistic as a functional of the empirical process. A similar integration by parts technique was used by Dehling and Taqqu (1989, 1991). Note that in the present paper, we use a one-dimensional integration by parts formula, whereas Dehling and Taqqu use a two-dimensional integration by parts. The latter would not work here, as the kernel  $1_{\{x \leq y\}}$  does not have locally bounded variation.

(2) As a corollary to Theorem 1, we obtain for *fixed*  $\lambda \in [0, 1]$  the asymptotic distribution of the Wilcoxon two-sample test, where the two samples are

$$\begin{aligned} &X_1, \dots, X_{[n\lambda]} \\ &X_{[n\lambda]+1}, \dots, X_n. \end{aligned}$$

After centering and normalization, the corresponding test statistic is

$$U_n = \frac{1}{n d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n \left( 1_{\{X_i \leq X_j\}} - \int_{\mathbb{R}} F(x)dF(x) \right).$$

From Theorem 1 we obtain that  $U_n$  converges in distribution to

$$\frac{1}{m!} (Z_m(\lambda) - \lambda Z_m(1)) \int_{\mathbb{R}} J_m(x)dF(x).$$

This can be rewritten as

$$\frac{1}{m!} ((1 - \lambda)Z_m(\lambda) - \lambda(Z_m(1) - Z_m(\lambda))) \int_{\mathbb{R}} J_m(x)dF(x),$$

an expression which is useful for the next remark.

(3) A different two-sample problem arises when we observe samples from two independent long-range dependent processes  $(X_i)_{i \geq 1}$  and  $(X'_i)_{i \geq 1}$  with identical joint distributions. In this case, the samples are

$$\begin{aligned} X_1, \dots, X_{[n\lambda]} \\ X'_1, \dots, X'_{n-[n\lambda]}. \end{aligned}$$

The normalized two-sample Wilcoxon test statistic for this case is

$$U'_n = \frac{1}{n d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=1}^{n-[n\lambda]} \left( 1_{\{X_i \leq X'_j\}} - \int_{\mathbb{R}} F(x) dF(x) \right).$$

Following the proof of Theorem 1, and making the appropriate changes where needed, one can show that  $U'_n$  converges in distribution towards

$$\frac{1}{m!} ((1 - \lambda)Z_m(\lambda) - \lambda Z'_m(1 - \lambda)) \int_{\mathbb{R}} J_m(x) dF(x),$$

where  $(Z'_m(\lambda))$  is an independent copy of  $(Z_m(\lambda))_{0 \leq \lambda \leq 1}$ . Note that the limit distributions in the two models are different, as the joint distribution of  $(Z_m(\lambda), Z_m(1) - Z_m(\lambda))$  is different from that of  $(Z_m(\lambda), Z'_m(1 - \lambda))$ . This is a result of the fact that the Hermite process does not have independent increments.

Observe that this is in contrast to the short-range dependent case. In this situation, the Wilcoxon two-sample test statistic has the same distribution in both models; see Dehling and Fried (2012). Roughly speaking, the dependence washes away in the limit for short-range dependence, but not for long-range dependence.

### 3 Application

In what follows, we will consider the model

$$X_i = \mu_i + G(\xi_i), \quad i = 1, \dots, n, \tag{16}$$

where  $(\xi_i)_{i \geq 1}$  is a mean-zero Gaussian process with  $\text{Var}(\xi_i) = 1$  and autocovariance function  $\rho(k)$  satisfying (1). We wish to test the hypothesis

$$H : \mu_1 = \dots = \mu_n \tag{17}$$

against the alternative

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

### 3.1 Wilcoxon-type rank test

The change-point test based on Wilcoxon's rank test will reject the null hypothesis for large values of

$$W_n = \frac{1}{n d_n} \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^k \sum_{j=k+1}^n \left( 1_{\{X_i \leq X_j\}} - \frac{1}{2} \right) \right|. \quad (18)$$

We first note an invariance property of the test statistic  $W_n$ .

**Lemma 1** *The test statistic  $W_n$  is invariant under strictly monotone transformations of the data, i.e.*

$$\frac{1}{n d_n} \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^k \sum_{j=k+1}^n \left( 1_{\{G(X_i) \leq G(X_j)\}} - \frac{1}{2} \right) \right| = \frac{1}{n d_n} \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^k \sum_{j=k+1}^n \left( 1_{\{X_i \leq X_j\}} - \frac{1}{2} \right) \right|$$

for all strictly monotone functions  $G : \mathbb{R} \rightarrow \mathbb{R}$ .

*Proof.* If  $G$  is strictly increasing, this is obvious, as  $G(X_i) \leq G(X_j)$  if and only if  $X_i \leq X_j$ . If  $G$  is strictly decreasing,  $G(X_i) \leq G(X_j)$  if and only if  $X_j \leq X_i$ , and thus

$$1_{\{G(X_i) \leq G(X_j)\}} = 1 - 1_{\{X_i \leq X_j\}}.$$

Hence we get

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

and thus the lemma is proved.  $\square$

Since  $X_i = \mu_i + G(\xi_i)$ , under the null hypothesis,

$$W_n = \frac{1}{n d_n} \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^k \sum_{j=k+1}^n \left( 1_{\{G(\xi_i) \leq G(\xi_j)\}} - \frac{1}{2} \right) \right|. \quad (19)$$

Thus, applying Theorem 1 and the continuous mapping theorem, under the null hypothesis,  $W_n$  converges in distribution, as  $n \rightarrow \infty$ , to

$$\sup_{0 \leq \lambda \leq 1} \left| \frac{Z_m(\lambda)}{m!} - \lambda \frac{Z_m(1)}{m!} \right| \left| \int_{\mathbb{R}} J_m(x) dF(x) \right|.$$

In order to set critical values for the asymptotic test based on  $W_n$ , we need to calculate the distribution of the last expression.

In what follows, we will assume that  $G$  is a strictly monotone function. In this case, combining (19) and Lemma 1 we get that

$$W_n = \frac{1}{n d_n} \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^k \sum_{j=k+1}^n \left( 1_{\{\xi_i \leq \xi_j\}} - \frac{1}{2} \right) \right|.$$

Note that in this case,

$$J_1(x) = E(\xi 1_{\{\xi \leq x\}}) = \int_{-\infty}^x t \varphi(t) dt = -\varphi(x),$$

where  $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$  denotes the standard normal density function. In the last step, we have used the fact that  $\varphi'(t) = -t \varphi(t)$ . Thus  $J_1(x) \neq 0$  for all  $x$  and hence the class of functions  $\{1_{\{\xi \leq x\}}, x \in \mathbb{R}\}$  has Hermite rank  $m = 1$ . Moreover, as  $F$  is the normal distribution function we obtain

$$\int_{\mathbb{R}} J_1(x) dF(x) = - \int_{\mathbb{R}} \varphi(x) \varphi(x) dx = \int_{\mathbb{R}} -(\varphi(s))^2 ds = -\frac{1}{2\sqrt{\pi}}. \quad (20)$$

We have thus proved the following theorem.

**Theorem 2** *Let  $(\xi_i)_{i \geq 1}$  be a stationary Gaussian process with mean zero, variance 1 and autocovariance function as in (1) with  $0 < D < 1$ . Moreover, let  $G : \mathbb{R} \rightarrow \mathbb{R}$  be a strictly monotone function and define  $X_i = \mu_i + G(\xi_i)$ . Then, under the null hypothesis  $H: \mu_1 = \dots = \mu_n$ , the test statistic  $W_n$ , as defined in (18), converges in distribution towards*

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

where  $(Z_1(\lambda))_{\lambda \geq 0}$  denotes the standard fractional Brownian motion process with Hurst parameter  $H = 1 - D/2 \in (1/2, 1)$ . The normalizing constant in (18) satisfies  $n d_n \sim \left( \frac{L(n)}{(1-D)(1-D/2)} \right)^{1/2} n^{2-D/2}$ .

– Insert Figure 1 here –

We have evaluated the distribution of  $\sup_{0 \leq \lambda \leq 1} |Z_1(\lambda) - \lambda Z_1(1)|$  by simulating 10,000 realizations of a standard fractional Brownian motion  $(Z_1^{(j)}(t))_{0 \leq t \leq 1}$ ,  $1 \leq j \leq 10,000$ ,  $t = \frac{i}{1000}$ ,  $0 \leq i \leq 1000$ , using the `fArma` package in R. For each realization, we have

calculated  $M_j := \max_{1 \leq i \leq 1000} |Z_1^{(j)}(\frac{i}{1000}) - \frac{i}{1000} Z_1^{(j)}(1)|$  as a numerical approximation to  $\sup_{0 \leq \lambda \leq 1} |Z_1^{(j)}(\lambda) - \lambda Z_1^{(j)}(1)|$ . The empirical distribution

$$F_M(x) := \frac{1}{10,000} \#\{1 \leq j \leq 10,000 : M_j \leq x\}$$

of these 10,000 maxima was used as approximation to the distribution of  $\sup_{0 \leq \lambda \leq 1} |Z_1(\lambda) - \lambda Z_1(1)|$ ; see Figure 1 for the estimated probability density and the empirical distribution function, for the Hurst parameter  $H = 0.7$ . We have calculated the corresponding upper  $\alpha$ -quantiles

$$q_\alpha := \inf\{x : F_M(x) \geq 1 - \alpha\} \quad (21)$$

for  $H = 0.6, 0.7, 0.9$ ; that is,  $D = 2 - 2H = 0.8, 0.6, 0.2$ ; see Table 1.

– Insert Table 1 here –

### 3.2 A difference-of-means test

As an alternative, we also consider a test based on differences of means of the observations. We consider the test statistic

$$D_n := \max_{1 \leq k \leq n-1} |D_{k,n}|, \quad (22)$$

where

$$D_{k,n} := \frac{1}{n d_n} \sum_{i=1}^k \sum_{j=k+1}^n (X_i - X_j).$$

One would reject the null hypothesis (17) if  $D_n$  is large. To obtain the asymptotic distribution of the test statistic, we apply the functional non-central limit theorem of Taqqu (1979) and obtain that

$$\begin{aligned} \frac{1}{n d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n (X_i - X_j) &= \frac{1}{n d_n} \left( (n - [n\lambda]) \sum_{i=1}^{[n\lambda]} G(\xi_i) - [n\lambda] \sum_{j=[n\lambda]+1}^n G(\xi_j) \right) \\ &\xrightarrow{w} a_m \left( (1 - \lambda) \frac{Z_m(\lambda)}{m!} - \lambda \left( \frac{Z_m(1)}{m!} - \frac{Z_m(\lambda)}{m!} \right) \right) \\ &= a_m \frac{1}{m!} (Z_m(\lambda) - \lambda Z_m(1)), \end{aligned}$$

where  $m$  denotes the Hermite rank of  $G(\xi)$  and where  $a_m = E(G(\xi)H_m(\xi))$  is the Hermite coefficient. Applying the continuous mapping theorem, we obtain the following theorem concerning the asymptotic distribution of the  $D_n$ .

**Theorem 3** Let  $(\xi_i)_{i \geq 1}$  be a stationary Gaussian process with mean zero, variance 1 and autocovariance function as in (1) with  $0 < D < 1/m$ . Moreover, let  $G : \mathbb{R} \rightarrow \mathbb{R}$  be a measurable function satisfying  $E(G^2(\xi)) < \infty$  and define  $X_i = \mu_i + G(\xi_i)$ . Then, under the null hypothesis  $H : \mu_1 = \dots = \mu_n$ , the test statistic  $D_n$ , as defined in (22), converges in distribution towards

$$\frac{|a_m|}{m!} \sup_{0 \leq \lambda \leq 1} |Z_m(\lambda) - \lambda Z_m(1)|,$$

where  $(Z_m(\lambda))$  denotes the  $m$ -th order Hermite process with Hurst parameter  $H = 1 - Dm/2 \in (1/2, 1)$ .

**Remark 2** (i) Observe that  $D_{k,n}$  may be rewritten as

$$D_{k,n} = \frac{k(n-k)}{n d_n} \left( \frac{1}{k} \sum_{i=1}^k X_i - \frac{1}{n-k} \sum_{i=k+1}^n X_i \right).$$

Thus, the difference-of-means test is indeed equal to the cusum test, which was investigated by Horvath and Kokoszka (1997) and Wang (2003). Theorem 3 can be obtained as a corollary of the results of the above mentioned papers. We have presented the short proof in the present paper for the ease of reference.

(ii) For a strictly increasing function  $G$ , the Hermite rank is 1, since

$$E(G(\xi)H_1(\xi)) = \int_{\mathbb{R}} G(s)H_1(s)\varphi(s)ds = \int_0^\infty s\varphi(s)(G(s) - G(-s))ds > 0.$$

Similarly, for a strictly decreasing function  $G$  we obtain  $E(G(\xi)H_1(\xi)) < 0$ . Thus, in these cases the test statistic  $D_n$  converges under the null hypothesis towards

$$|a_1| \sup_{0 \leq \lambda \leq 1} |Z_1(\lambda) - \lambda Z_1(1)|,$$

where  $Z_1$  is fractional Brownian motion with index  $H = 1 - D/2$ . Note that in this case, up to a norming constant, the limit distribution of the difference-of-means test is the same as for the test based on Wilcoxon's rank statistics.

## 4 Difference-of-means test under fractional Gaussian noise

We want to obtain a lower bound for the power of the difference-of-means test for fractional Gaussian noise, i.e. for the model

$$X_i = \mu_i + \xi_i, \quad i = 1, \dots, n.$$

While the distribution of the difference-of-means test statistic  $D_n$ , as defined in (22), is not explicitly known, one can calculate the exact distribution of

$$D_{k,n} = \frac{1}{n d_n} \sum_{i=1}^k \sum_{j=k+1}^n (X_i - X_j),$$

for the case of fractional Gaussian noise that we consider here. Recall that fractional Gaussian noise can be obtained by differencing fractional Brownian motion, that is

$$\xi_k = B_H(k) - B_H(k-1), \quad (23)$$

where  $(B_H(t))_{t \geq 0}$  is the standard fractional Brownian motion with Hurst parameter  $H$ , which we denoted  $Z_1$ . Its covariance is given in (6). The random variables  $\xi_k$  have then mean zero, variance 1 and autocovariances

$$\rho(k) \sim H(2H-1)k^{2H-2} = \left(1 - \frac{D}{2}\right) (1-D)k^{-D}, \quad (24)$$

where  $D = 2 - 2H$ , and thus they have long-range dependence.

Consider the following alternative

$$H_{\lambda,h} : E(X_i) = 0 \text{ for } i = 1, \dots, [n\lambda] \text{ and } E(X_i) = h \text{ for } i = [n\lambda] + 1, \dots, n.$$

We shall compute the exact distribution of  $D_{k,n}$  under  $H_{\lambda,h}$  and thus obtain a lower bound for the power of the test based on  $D_n$ , defined in (22), since

$$P(D_n \geq q_\alpha) \geq P(D_{[n\lambda],n} \geq q_\alpha),$$

where  $\{D_n \geq q_\alpha\}$  is the rejection region and  $q_\alpha$  is given in (21).

First note that  $d_n = n^H$  and thus  $nd_n = n^{1+H}$ , where  $H$  is again the Hurst coefficient.  $D_{k,n}$  has a normal distribution with mean

$$E(D_{k,n}) = \frac{1}{n^{1+H}} \sum_{i=1}^k \sum_{j=k+1}^n (E(X_i) - E(X_j)).$$

Thus a small calculation shows that

$$E(D_{k,n}) = \begin{cases} -\frac{1}{n^{1+H}} k (n - [n\lambda]) h & \text{if } k \leq [n\lambda] \\ -\frac{1}{n^{1+H}} (n - k) [n\lambda] h & \text{if } k \geq [n\lambda]. \end{cases}$$

Note that  $\max_{1 \leq k \leq n-1} |E(D_{k,n})| = |E(D_{[n\lambda],n})| = \frac{1}{n^{1+H}}(n - [n\lambda])[n\lambda]h \sim n^{1-H}\lambda(1-\lambda)h$ . Since the variance of  $D_{k,n}$  is not changed by the level shift, we get

$$\begin{aligned}
\text{Var}(D_{k,n}) &= \text{Var}\left(\frac{1}{n^{1+H}} \sum_{i=1}^k \sum_{j=k+1}^n (\xi_i - \xi_j)\right) \\
&= \text{Var}\left(\frac{1}{n^{1+H}} \left((n-k) \sum_{i=1}^k \xi_i - k \sum_{j=k+1}^n \xi_j\right)\right) \\
&= \text{Var}\left(\frac{1}{n^{1+H}} ((n-k)B_H(k) - k(B_H(n) - B_H(k)))\right) \\
&= \text{Var}\left(\frac{1}{n^{1+H}} (nB_H(k) - kB_H(n))\right).
\end{aligned}$$

By the self-similarity of fractional Brownian motion, we finally get

$$\begin{aligned}
\text{Var}(D_{k,n}) &= \text{Var}\left(B_H\left(\frac{k}{n}\right) - \frac{k}{n}B_H(1)\right) \\
&= \text{Var}\left(B_H\left(\frac{k}{n}\right)\right) + \frac{k^2}{n^2}\text{Var}(B_H(1)) - 2\frac{k}{n}\text{Cov}\left(B_H\left(\frac{k}{n}\right), B_H(1)\right) \\
&= \left(\frac{k}{n}\right)^{2H} + \frac{k^2}{n^2} - \frac{k}{n} \left( \left(\frac{k}{n}\right)^{2H} + 1 - \left(1 - \frac{k}{n}\right)^{2H} \right) \\
&= \left(\frac{k}{n}\right)^{2H} + \left(\frac{k}{n}\right)^2 - \left(\frac{k}{n}\right)^{2H+1} - \frac{k}{n} + \frac{k}{n} \left(1 - \frac{k}{n}\right)^{2H} \\
&= \left(\frac{k}{n}\right)^{2H} \left(1 - \frac{k}{n}\right) - \frac{k}{n} \left(1 - \frac{k}{n}\right) + \frac{k}{n} \left(1 - \frac{k}{n}\right)^{2H}.
\end{aligned}$$

Defining

$$\sigma^2(\lambda) = \lambda^{2H}(1-\lambda) - \lambda(1-\lambda) + \lambda(1-\lambda)^{2H},$$

we thus obtain

$$\text{Var}(D_{k,n}) = \sigma^2(k/n).$$

The variance is maximal for  $k = n/2$ , in which case we obtain

$$\text{Var}(D_{n/2,n}) = \frac{1}{2^{2H}} - \frac{1}{4} \approx 0.13.$$

The distribution of  $D_{k,n}$  gives a lower bound for the power of the difference-of-means test at



the alternative  $H_{\lambda,h}$  considered above. We have

$$\begin{aligned}
P(D_n \geq q_\alpha) &\geq P(|D_{[n\lambda],n}| \geq q_\alpha) \\
&= P(D_{[n\lambda],n} \leq -q_\alpha) + P(D_{[n\lambda],n} \geq q_\alpha) \\
&= P\left(\frac{D_{[n\lambda],n} + n^{1-H}\lambda(1-\lambda)h}{\sqrt{\sigma^2(\lambda)}} \leq \frac{-q_\alpha + n^{1-H}\lambda(1-\lambda)h}{\sqrt{\sigma^2(\lambda)}}\right) \\
&\quad + P\left(\frac{D_{[n\lambda],n} + n^{1-H}\lambda(1-\lambda)h}{\sqrt{\sigma^2(\lambda)}} \geq \frac{q_\alpha + n^{1-H}\lambda(1-\lambda)h}{\sqrt{\sigma^2(\lambda)}}\right) \\
&\approx \Phi\left(\frac{-q_\alpha + n^{1-H}\lambda(1-\lambda)h}{\sqrt{\sigma^2(\lambda)}}\right) + 1 - \Phi\left(\frac{q_\alpha + n^{1-H}\lambda(1-\lambda)h}{\sqrt{\sigma^2(\lambda)}}\right),
\end{aligned}$$

where  $\Phi$  is the c.d.f. of a standard normal random variable. E.g., for  $H = 0.7$ ,  $\lambda = \frac{1}{2}$ , we get  $q_{0.05} = 0.87$  using Table 1 and thus

$$P(D_n \geq q_{0.05}) \geq \Phi\left(\frac{-0.87 + n^{0.3}h/4}{\sqrt{\sigma^2(1/2)}}\right) \approx \Phi(-2.42 + 0.70 h n^{0.3}).$$

In this way, for sample size  $n = 500$  and level shift  $h = 1$  we get  $\Phi(2.07) \approx .98$  as lower bound on the power of the difference-of-means test. For the same sample size, but  $h = 0.5$ , we get the lower bound  $\Phi(-0.18) \approx 0.43$ .

## 5 Simulations

In this section, we will present the results of a simulation study involving the Wilcoxon type rank test (18) and the difference-of-means test (22). We first investigate whether these tests reach their asymptotic level when applied in a finite sample setting, for sample sizes ranging from  $n = 10$  to  $n = 1,000$ . Secondly, we compare the power of the two tests for sample size  $n = 500$  at various different alternatives

$$A_k : \mu_1 = \dots = \mu_k \neq \mu_{k+1} = \dots = \mu_n.$$

We let both the break point  $k$  and the level shift  $h := \mu_{k+1} - \mu_k$  vary. Specifically, we choose  $k = 25, 50, 150, 250$  and  $h = 0.5, 1, 2$ .

As a basis for our simulations, we have taken realizations  $\xi_1, \dots, \xi_n$  of a fractional Gaussian noise (fGn) process with Hurst parameter  $H$ ; respectively  $D = 2 - 2H$ , see (23) and (24). We have repeated each simulation 10,000 times.

## 5.1 Normally distributed data

In our first simulations, we took  $G(t) = t$ , so that  $(X_i)_{i \geq 1}$  is fGn.  $F$  is then the c.d.f.  $\Phi$  of a standard normal random variable. In order to determine the critical values for the test statistics  $W_n$  and  $D_n$ , we have applied Theorem 2 and Theorem 3. Since  $G$  is strictly increasing, Theorem 2 yields that, under the null hypothesis,  $W_n$  has approximately the same distribution as

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

Since  $G$  is the first Hermite polynomial, its Hermite rank is  $m = 1$  and the associated Hermite coefficient is  $a_1 = 1$ . Hence, Theorem 3 yields that, under the null hypothesis, the test statistic  $D_n$  has approximately the same distribution as

$$\sup_{\lambda \in [0,1]} |Z_1(\lambda) - \lambda Z_1(1)|,$$

We have calculated asymptotic critical values for both tests by using the upper 5%-quantiles of  $\sup_{\lambda \in [0,1]} |Z_1(\lambda) - \lambda Z_1(1)|$ , as given in Table 1. Thus the Wilcoxon-type test rejected the null hypothesis when  $W_n \geq \frac{1}{2\sqrt{\pi}} q_\alpha$ , while the difference-of-means test rejected when  $D_n \geq q_\alpha$ , where  $q_\alpha$  is given in (21).

– **Insert Table 2 here** –

We have checked whether the tests reach their asymptotic level of 5% and counted the number of rejections of the null hypothesis in 10,000 simulations, where the null hypothesis was true. We see in Table 2 that both tests perform well already for moderate sample sizes of  $n = 50$ , with the notable exception of the Wilcoxon-type test when  $H = 0.9$ , i.e. when we have very strong dependence. In that case, convergence in Theorem 2 appears to be very slow so that the asymptotic critical values are misleading when applied in a finite sample setting.

– **Insert Figure 2 here** –

In order to analyze how well the tests detect break points, we have introduced a level

shift  $h$  at time  $[n\lambda]$ , i.e. we consider the time series

$$X_i = \begin{cases} \xi_i & \text{for } i = 1, \dots, [n\lambda] \\ \xi_i + h & \text{for } i = [n\lambda] + 1, \dots, n. \end{cases}$$

We have done this for several choices of  $\lambda$  and  $h$ , for sample size  $n = 500$ . As Table 3 shows, both tests detect breaks very well – and the better, the larger the level shift is and the more in the middle the shift takes place. When the break occurs in the middle, both tests perform equally well. Breaks at the beginning are better detected by the difference-of-means test.

– **Insert Table 3 here** –

## 5.2 Heavy-Tailed Data

In the second set of simulations, we took Pareto distributed data. Note that the Pareto( $\beta, k$ ) distribution has distribution function

$$F(x) = \begin{cases} 1 - \left(\frac{k}{x}\right)^\beta & \text{if } x \geq k \\ 0 & \text{else,} \end{cases}$$

where the scale parameter  $k$  is the smallest possible value for  $x$  and where  $\beta$  is a shape parameter. The Pareto distribution has finite expected value when  $\beta > 1$  and finite variance when  $\beta > 2$ . The expected value and the variance are given by

$$\begin{aligned} \mu &= E(X) = \frac{\beta k}{\beta - 1}, \quad \beta > 1 \\ \sigma^2 &= \text{Var}(X) = \frac{\beta k^2}{(\beta - 1)^2(\beta - 2)}, \quad \beta > 2. \end{aligned}$$

In order to obtain Pareto( $\beta, k$ )-distributed  $X = G(\xi)$ , we take  $G$  to be the quantile transform, i.e.  $G(t) = k(\Phi(t))^{-1/\beta}$  where  $\Phi$  denotes the standard normal distribution function, so that for  $x \geq k$

$$P(X_i \geq x) = P(G(\xi_i) \geq x) = P\left(\left(\Phi(\xi_i)\right)^{-1/\beta} \geq \frac{x}{k}\right) = P\left(\xi_i \leq \Phi^{-1}\left(\left(\frac{k}{x}\right)^\beta\right)\right) = \left(\frac{k}{x}\right)^\beta.$$

Since we want the  $X_i$  to be centered, we will in fact take

$$G(t) = k(\Phi(t))^{-1/\beta} - \frac{\beta k}{\beta - 1}. \tag{25}$$

If we want  $X_i$  to be standardized to have mean 0 and variance 1, we will consider  $Z = (X - \mu)/\sigma$  and take

$$G(t) = \left( \frac{\beta k^2}{(\beta - 1)^2(\beta - 2)} \right)^{-1/2} \left( k(\Phi(t))^{-1/\beta} - \frac{\beta k}{\beta - 1} \right) \quad (26)$$

The corresponding distribution function of  $Z$  is then

$$F_Z(z) = \begin{cases} 1 - \left( \frac{k}{\sigma z + \mu} \right)^\beta & \text{if } z \geq \frac{k - \mu}{\sigma} \\ 0 & \text{else.} \end{cases} \quad (27)$$

and its density function is

$$f_Z(z) = \begin{cases} k^\beta \beta \sigma (\sigma z + \mu)^{-\beta - 1} & \text{if } z \geq \frac{k - \mu}{\sigma} \\ 0 & \text{else} \end{cases} \quad (28)$$

**Pareto(3, 1) Data:** We first performed simulations with Pareto(3, 1) data, i.e. heavy-tailed data with finite variance. In this case,  $\beta = 3$ ,  $k = 1$  and we have  $E(X) = \frac{3}{2}$  and  $\text{Var}(X) = \frac{3}{4}$ . For a better comparison with the simulations involving fractional Gaussian noise, we also standardize the data, i.e. we consider (see (26)),

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

The probability density function of the standardized  $X$  is given by (see (28)),

$$f(x) = \begin{cases} 3\sqrt{\frac{3}{4}} \left( \sqrt{\frac{3}{4}} x + \frac{3}{2} \right)^{-4} & \text{if } x \geq -\sqrt{\frac{1}{3}} \\ 0 & \text{else.} \end{cases}$$

$G$  is again strictly decreasing, and by the above results, the Hermite rank of the class of functions  $\{I_{\{G(\xi_i) \leq x\}} - F(x), x \in \mathbb{R}\}$  is  $m = 1$ , the Hermite rank of  $G$  itself is  $m = 1$  and  $|\int_{\mathbb{R}} J_1(x) dF(x)| = (2\sqrt{\pi})^{-1}$ , see (20). Numerical integration yields

$$a_1 = E[\xi G(\xi)] = \sqrt{\frac{4}{3}} \int_{-\infty}^{\infty} s \Phi(s)^{-1/3} \varphi(s) ds \approx -0.6784.$$

Table 4 shows the observed level of the tests, for various sample sizes and various Hurst parameters. For sample sizes up to  $n = 1,000$ , the "difference-of-means" test has level larger than 10%. We conjecture that this is due to the slow convergence in Theorem 3. This conjecture is supported by the outcomes of simulations with sample sizes  $n = 2,000$  and  $n = 10,000$ ; see Table 4.

– **Insert Table 4 here** –

Table 5 gives the observed power of the "difference-of-means" test and the Wilcoxon-type test, for sample size  $n = 500$  and various values of the break points and height of level shift. The results show that the Wilcoxon-type test has larger power than the "difference-of-means" test for small level shift  $h$ , but that the "difference-of-means" test outperforms the Wilcoxon type test for larger level shifts.

– **Insert Table 5 here** –

However, the above comparison is not really meaningful, since the "difference-of-means" test has a realized level of approximately 10% while the Wilcoxon-type test has level close to 5%; see Table 4. Thus we have calculated the finite sample 5%-quantiles of the distribution of the "difference-of-means" test, using a Monte-Carlo simulation; see Table 6. For example, for  $n = 500$  and  $H = 0.7$ , the corresponding critical value is 0.70. Thus we reject the null hypothesis of no break point if the "difference-of-means" test statistic is greater than 0.70.

– **Insert Table 6 here** –

The value of 0.70 should be contrasted to the asymptotic ( $n = \infty$ ) value of 0.59. To obtain the asymptotic critical values, we proceeded as follows: according to Theorem 3, the asymptotic distribution, under the null hypothesis, of the test statistic  $D_n$  equals the distribution of

$$|a_1| \sup_{0 \leq \lambda \leq 1} |Z(\lambda) - \lambda Z(1)|.$$

Thus the asymptotic upper  $\alpha$ -quantiles of  $D_n$  can be calculated as  $|a_1|q_\alpha$ , where  $q_\alpha$  is the upper  $\alpha$ -quantile of the distribution of  $\sup_{0 \leq \lambda \leq 1} |Z(\lambda) - \lambda Z(1)|$ , as tabulated in Table 1.

– **Insert Table 7 here** –

We have then calculated the power of the "difference-of-means" test through simulation,

with  $n = 500$ ,  $H = 0.7$  and the finite sample quantile critical value of 0.70 rather than the asymptotic value of 0.59 (see Table 6). Table 7 shows the power of the test. We can now compare the results of the Wilcoxon-type test given in the right-hand side of Table 5 with the finite sample “difference-of-means” test results given in Table 7. We see that the Wilcoxon-type test has better power than the “difference-of-means” test, except for large level shifts at an early time. Such changes are detected more often by the “difference-of-means” test.

**Pareto(2, 1) Data:** We now choose  $k = 1$  and  $\beta = 2$ , so that the  $X$  have finite expectation, but infinite variance. In order to have centered data, we take  $G$  as in (25), i.e.

$$G(t) = \frac{1}{\sqrt{\Phi(t)}} - 2.$$

– **Insert Table 8 here** –

We will now compare both tests, i.e. the Wilcoxon-type test and the difference of means test, although the latter can strictly speaking not be applied because it requires data with finite variance, respectively  $G \in L_2(\mathbb{R}, \mathcal{N})$ . By Theorem 2, under the null hypothesis of no change, the Wilcoxon-type test statistic  $W_n$  converges in distribution towards

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

In fact, as a consequence of Lemma 1, even the finite sample distribution of  $W_n$  is the same as for normally distributed data. Table 8 gives the measured level of the Wilcoxon-type test (the asymptotic level is 5%) and Table 9 suggests it has good power, especially for small shifts in the middle of the observations.

– **Insert Table 9 here** –

Let us now consider the difference-of-means test. Note that, strictly speaking, Theorem 3 cannot be applied in the case of the Pareto data with  $\beta = 2$  because it requires the variance of the data to be finite. It is interesting nevertheless to use the asymptotic test suggested by Theorem 3. Since  $G$  is strictly monotone, the Hermite rank of  $G$  is  $m = 1$  as well, by the

remark following Theorem 3. Using numerical integration, we have calculated

$$a_1 = E[\xi G(\xi)] = \int_{-\infty}^{\infty} s(\Phi(s)^{-1/2} - 2)\varphi(s) ds = \int_{-\infty}^{\infty} s\Phi(s)^{-1/2}\varphi(s) ds \approx -1.40861.$$

We clearly see in Table 8 that the difference-of-means test very often falsely rejects the null hypothesis, that is detects breaks where there are none, while the Wilcoxon-type test is robust. Table 9 shows that both tests have good power, but again, the Wilcoxon tests is clearly better, especially for small shifts in the middle of the observations.

## 6 Conclusion

We have investigated the properties of a nonparametric change-point test in the presence of long-range dependent data. Our test is based on Wilcoxon's two sample rank test statistic. We analytically derive the asymptotic distribution of our test statistic, and we perform a simulation study to calculate its power. The simulation study shows that already for moderate sample sizes, our test attains the asymptotic level. For LRD Gaussian data, our test has almost the same power as a difference-of-means test. For heavy-tailed data, our new test outperforms the difference-of-means test. In a subsequent paper, we plan to derive the asymptotic distribution of our test statistic under local alternatives.

### Acknowledgement.

Herold Dehling and Aeneas Rooch were supported in part by the Collaborative Research Grant 823, Project C3 *Analysis of Structural Change in Dynamic Processes*, of the German Research Foundation. Murad S. Taqqu was supported in part by the NSF grants DMS-0608669 and DMS-1007616 at Boston University. The authors wish to thank the referee of this paper for his/her careful reading of the manuscript and for insightful and constructive comments.

## References

- [1] Antoch, J., Hušková, M., Janic, M.A. & Ledwina, T. (2008). Data driven rank tests for the change point problem. *Metrika* **68**, 1–15.
- [2] Aue, A., Hörmann, S., Horváth, L. & Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series. *The Annals of Statistics* **37**, 4046–4087.
- [3] Berkes, I., Horváth, L., Kokoszka, P. & Shao, Q.-M. (2006). On discriminating between long-range dependence and changes in the mean. *The Annals of Statistics* **34**, 1140–1165.

- [4] Csörgő, M. & Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. J. Wiley & Sons, Chichester.
- [5] Dehling, H. & Fried, R. (2012). Asymptotic Distribution of Two-Sample Empirical U-Quantiles with Applications to Robust Tests for Shifts in Location. *Journal of Multivariate Analysis* **105**, 124–140.
- [6] Dehling, H. & Taqqu, M.S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *The Annals of Statistics* **17**, 1767–1783.
- [7] Dehling, H. & Taqqu, M.S. (1991). Bivariate symmetric statistics of long-range dependent observations. *Journal of Statistical Planning and Inference* **28**, 153–165.
- [8] Giraitis, L., Leipus, R. & Surgailis, D. (1996). The change-point problem for dependent observations. *Journal of Statistical Planning and Inference* **53**, 297–310.
- [9] Horváth, L. & Kokoszka, P. (1997). The effect of long-range dependence on change-point estimators. *Journal of Statistical Planning and Inference* **64**, 57–81.
- [10] Krämer, W. & Sibbertsen, P. & Kleiber, C. (2002). Long memory versus structural change in financial time series. *Allgemeines Statistisches Archiv* **86**, 83–96.
- [11] Lehmann, E. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco.
- [12] Ling, S. (2007). Testing for change points in time series models and limiting theorems for NED sequences. *The Annals of Statistics* **35**, 1213–1227.
- [13] Shorack, G.R. & Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley & Sons, New York.
- [14] Taqqu, M.S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50**, 53–83.
- [15] Wang, L. (2003). Limit theorems in change-point problems with multivariate long-range dependent observations. *Statistics & Decisions* **21**, 283–300.
- [16] Wang, L. (2008a). Change-Point Detection with Rank Statistics in Long-Memory Time-Series Models. *Australian & New Zealand Journal of Statistics* **50**, 241–256.
- [17] Wang, L. (2008b). Change-in-mean problem for long memory time series models with applications. *Journal of Statistical Computation and Simulation* **78**, 653–668.
- [18] Wied, D., Krämer, W. & Dehling, H. (2011). Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory*, Available on CJO 2011 doi:10.1017/S0266466611000661.

Herold Dehling, Fakultät für Mathematik, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany.

E-mail: herold.dehling@rub.de



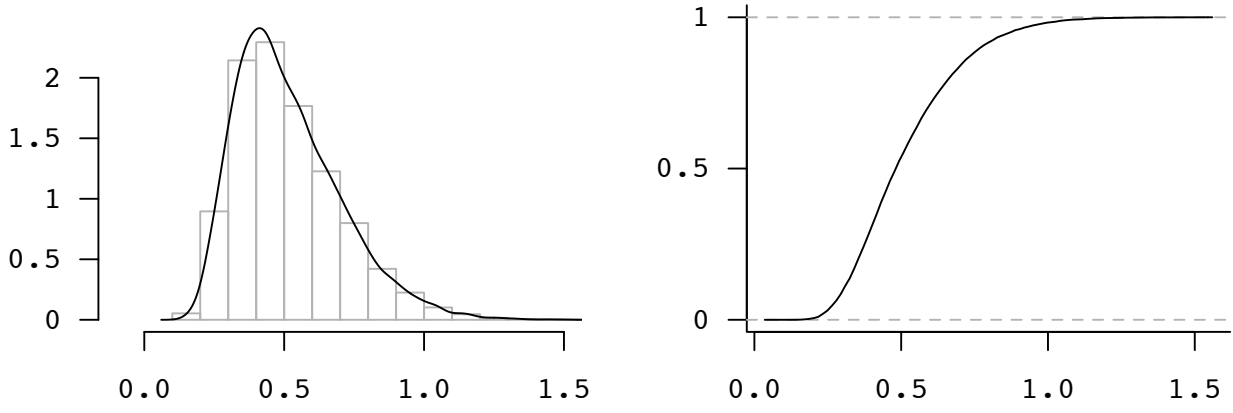


Figure 1: Probability density (left) and distribution function (right) of  $\sup_{0 \leq \lambda \leq 1} |Z(\lambda) - \lambda Z(1)|$ ; calculations based on 10,000 simulation runs.

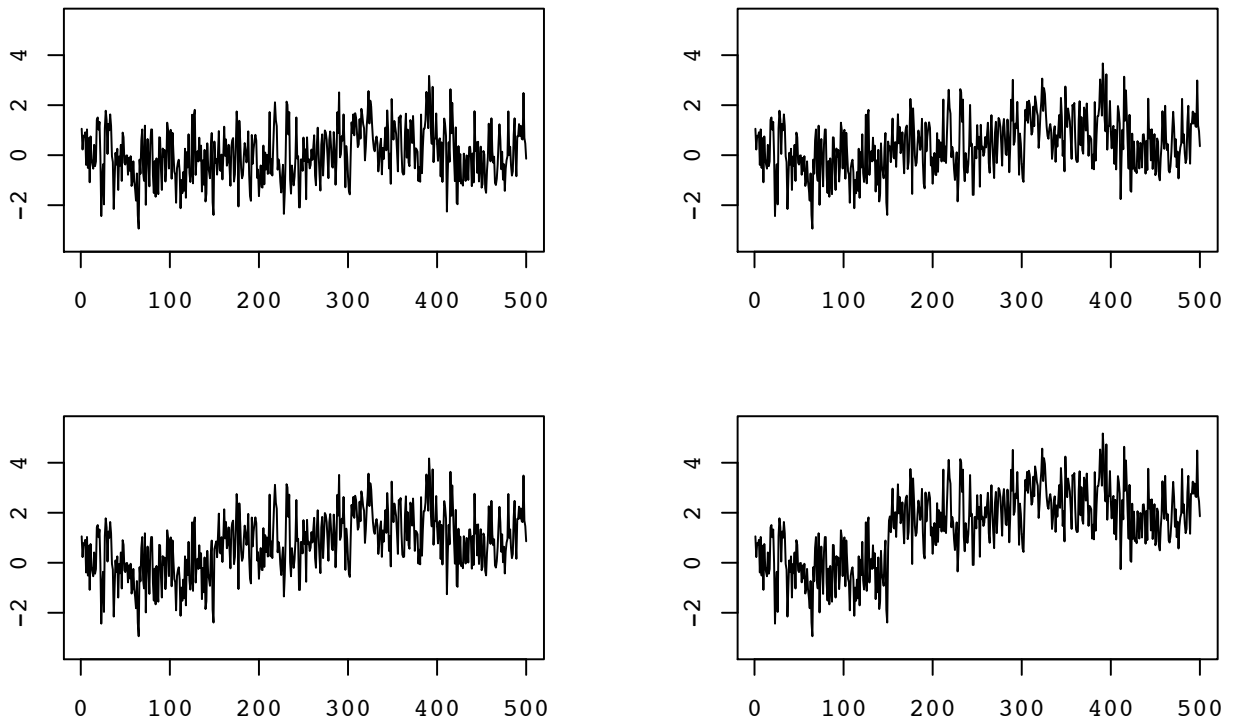


Figure 2: fGn without breaks (top left) and with a jump after observation 150 (this is  $[\lambda n]$  with  $\lambda = 0.3$ ) of height  $h = 0.5$  (top right),  $h = 1$  (bottom left) and  $h = 2$  (bottom right).

Table 1: Upper  $\alpha$ -quantiles of  $\sup_{0 \leq \lambda \leq 1} |Z(\lambda) - \lambda Z(1)|$ , where  $Z$  is a standard fBm, for different LRD parameters  $H$ , based on 10,000 repetitions.

$H / \alpha$	0.10	0.05	0.01
0.6	0.98	1.10	1.34
0.7	0.77	0.87	1.06
0.9	0.38	0.44	0.54

Table 2: Level of “difference-of-means” test (left) and level of “Wilcoxon-type” test (right) for fGn time series with LRD parameter  $H$ ; 10,000 simulation runs. Both tests have asymptotically level 5%.

$n / H$	0.6	0.7	0.9	$n / H$	0.6	0.7	0.9
10	0.024	0.031	0.038	10	0.013	0.026	0.326
50	0.039	0.042	0.047	50	0.042	0.050	0.167
100	0.042	0.046	0.044	100	0.044	0.051	0.140
500	0.047	0.052	0.049	500	0.051	0.052	0.100
1000	0.047	0.052	0.053	1000	0.051	0.054	0.095

Table 3: Power of “difference-of-means” test (left) and power of “Wilcoxon-type” test (right) for  $n = 500$  observations of fGn with LRD parameter  $H = 0.7$ , different break points  $[\lambda n]$  and different level shifts  $h$ . Both tests have asymptotically level 5%. The calculations are based on 10,000 simulation runs.

$h / \lambda$	0.05	0.1	0.3	0.5	$h / \lambda$	0.05	0.1	0.3	0.5
0.5	0.060	0.090	0.388	0.524	0.5	0.058	0.086	0.386	0.525
1	0.090	0.254	0.952	0.985	1	0.077	0.184	0.948	0.985
2	0.261	0.965	1.000	1.000	2	0.119	0.763	1.000	1.000

Table 4: Level of “difference-of-means” test (left) and level of Wilcoxon-type test (right) for standardised Pareto(3,1)-transformed fGn with LRD parameter  $H$ ; 10,000 simulation runs. Both tests have asymptotically level 5%.

$n / H$	0.6	0.7	0.9	$n / H$	0.6	0.7	0.9
10	0.104	0.109	0.117	10	0.013	0.026	0.326
50	0.138	0.127	0.126	50	0.042	0.050	0.167
100	0.145	0.125	0.126	100	0.044	0.051	0.140
500	0.140	0.103	0.119	500	0.051	0.052	0.100
1000	0.131	0.101	0.123	1000	0.051	0.054	0.095
2000	0.120	0.086	0.115				
10000	0.106	0.069	0.101				

Table 5: Power of “difference-of-means” test (left) and power of “Wilcoxon-type” test (right) for  $n = 500$  observations of standardised Pareto(3,1)-transformed fGn with LRD parameter  $H = 0.7$ , different break points  $[\lambda n]$  and different level shifts  $h$ . Both tests have asymptotically level 5%. The calculations are based on 10,000 simulation runs.

$h / \lambda$	0.05	0.1	0.3	0.5	$h / \lambda$	0.05	0.1	0.3	0.5
0.5	0.116	0.177	0.756	0.864	0.5	0.088	0.294	0.983	0.998
1	0.177	0.693	0.998	1.000	1	0.115	0.655	1.000	1.000
2	0.815	0.998	1.000	1.000	2	0.138	0.944	1.000	1.000

Table 6: 5%-quantiles of the finite sample distribution of the “difference-of-means” test under the null hypothesis for Pareto(3,1)-transformed fGn with different LRD parameter  $H$  and different sample sizes  $n$ . The calculations are based on 10,000 simulation runs.

$H / n$	10	50	100	500	1,000	2,000	10,000	$\infty$
0.6	1.02	1.04	1.02	0.93	0.90	0.88	0.85	0.75
0.7	0.84	0.82	0.79	0.70	0.68	0.65	0.62	0.59
0.9	0.47	0.47	0.44	0.43	0.42	0.41	0.38	0.30

Table 7: Power of the “difference-of-means” test, based on the finite sample quantiles, for  $n = 500$  observations of Pareto(3,1)-transformed fGn with LRD parameter  $H = 0.7$ , different break points  $[\lambda n]$  and different level shifts  $h$  (left). The calculations are based on 10,000 simulation runs.

$h / \lambda$	0.05	0.1	0.3	0.5
0.5	0.053	0.078	0.566	0.733
1	0.078	0.379	0.994	0.999
2	0.423	0.994	1.000	1.000

Table 8: Level of “difference-of-means” test (left) and level of “Wilcoxon-type” test (right) for Pareto(2,1)-transformed fGn with LRD parameter  $H$ ; 10,000 simulation runs.

$n / H$	0.6	0.7	0.9	$n / H$	0.6	0.7	0.9
10	0.104	0.104	0.107	10	0.013	0.026	0.326
50	0.159	0.138	0.120	50	0.042	0.050	0.167
100	0.181	0.151	0.122	100	0.044	0.051	0.140
500	0.223	0.148	0.124	500	0.051	0.052	0.100
1000	0.232	0.151	0.130	1000	0.051	0.054	0.095

Table 9: Power of “difference-of-means” test (left) and power of “Wilcoxon-type” test (right) for  $n = 500$  observations of Pareto(2,1)-transformed fGn with LRD parameter  $H = 0.7$ , different break points  $[\lambda n]$  and different level shifts  $h$ . The calculations are based on 10,000 simulation runs.

$h / \lambda$	0.05	0.1	0.3	0.5	$h / \lambda$	0.05	0.1	0.3	0.5
0.5	0.148	0.156	0.272	0.350	0.5	0.075	0.180	0.878	0.960
1	0.156	0.200	0.741	0.853	1	0.097	0.401	0.997	1.000
2	0.199	0.651	0.996	0.999	2	0.122	0.744	1.000	1.000