Power of change-point tests for long-range dependent data

Herold Dehling, Aeneas Rooch,
Murad S. Taqqu

February 28, 2013

POWER OF CHANGE-POINT TESTS FOR LONG-RANGE DEPENDENT DATA

HEROLD DEHLING, AENEAS ROOCH, AND MURAD S. TAQQU

ABSTRACT. We investigate the power of the CUSUM test and the Wilcoxon change-point tests for a shift in the mean of a process with long-range dependent noise. We derive analytic formulas for the power of these tests under local alternatives. These results enable us to calculate the asymptotic relative efficiency (ARE) of the CUSUM test and the Wilcoxon change point test. We obtain the surprising result that for Gaussian data, the ARE of these two tests equals 1, in contrast to the case of i.i.d. noise when the ARE is known to be $3/\pi$.

CONTENTS

1. INTRODUCTION

Statistical tests for the presence of changes in the structure of time series are of great importance in a wide range of scientific discussions, e.g. regarding economic, technological and climate data. Many procedures for detecting changes and for estimating change-points have been proposed in the literature; see e.g. Csörgő and Horvath (1997) for a detailed exposition. In the case of independent data, the theory is quite satisfactory. For various types of change-point models, statistical procedures have been proposed and their properties investigated. In constrast, the situation is different for dependent data, such as encountered in time series models. For dependent data, most research has focused on linear procedures, such as cumulative sum (CUSUM) tests, and there are many open problems when it comes to other types of test procedures, e.g. those used in robust statistics.

In the present paper, we study the change-point problem for long-range dependent data. Specifically, we will test the hypothesis that the process is stationary against the alternative that there is a change in the mean. The classical test statistic for this problem is the CUSUM

statistic,

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

When the test statistic is large, one infers that there is a change in the mean. The CUSUM test has good properties when the underlying process is Gaussian. However, this test is not robust against possible outliers in the data and against heavy-tailed distributions. Recently, Dehling, Rooch and Taqqu (2013) have proposed a robust alternative to the CUSUM test, which is based on the Wilcoxon two-sample rank statistic. The corresponding "Wilcoxon change-point test" uses the test statistic

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

One rejects the null hypothesis when this test statistic is large. In their paper, Dehling, Rooch and Taqqu (2013) investigated the asymptotic distribution of the Wilcoxon change-point test under the null hypothesis of no change. Moreover, they performed a simulation study to compare the finite sample performance and the power of the CUSUM test based on (1) and the Wilcoxon change-point test based on (2).[1]

In the present paper, we study the power of the CUSUM test and the Wilcoxon change-point test for a shift in the mean of a long-range dependent process. We will calculate the power under local alternatives, where the height of the shift decreases with the sample size n in such a way that the tests have non-trivial limit power as $n \to \infty$. These results enable us to compute the asymptotic relative efficiency (ARE) of the CUSUM and the Wilcoxon change-point tests, which is defined as the limit of the ratio of the sample sizes required to obtain a given power. We obtain the surprising result that the ARE of these two tests equals 1 in the case of long-range dependent Gaussian data. This is in contrast with the case of i.i.d. and short-range dependent data, where the ARE of the Wilcoxon change-point test with respect to the CUSUM test is $3/\pi$.

We consider a model where the observations are generated by a stochastic process $(X_i)_{i \geq 1}$ of the type

$$(3) \qquad X_i = \mu_i + \epsilon_i,$$

where $(\epsilon_i)_{i \geq 1}$ is a long-range dependent stationary process with mean zero, finite variance 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 $(\xi_i)_{i \geq 1}$ 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 long-range dependent (LRD), mean-zero Gaussian process with variance $E(\xi_i^2) = 1$ and autocovariance function

$$(4) \qquad \rho(k) = k^{-D} L(k), \quad k \geq 1,$$

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

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

$$H : \mu_1 = \ldots = \mu_n$$

---

[1]Dehling et al. (2013) called the CUSUM test, the "difference of means test", and called the Wilcoxon change-point test, the "Wilcoxon-type" test.

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

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

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

Dehling, Rooch and Taqqu (2013) have studied two tests for this change-point problem, namely the Wilcoxon change-point test which is based on the test statistic

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

and the CUSUM test which uses the test statistic

(7) $$D_n := \frac{1}{n \, d_n} \max_{1 \leq k \leq n-1} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} (X_j - X_i) \right|.$$

Observe that the normalization d_n, which will be specified below, is the same for both tests. These tests are similar in spirit. They compare the first part of the sample to the second part. The Wilcoxon change-point test (6) involves the rank of the data whereas the CUSUM test (7) involves their values. One rejects the null hypothesis of no change when these test statistics are large.

Dehling, Rooch and Taqqu (2013) investigated the asymptotic distribution of these test statistics under the null hypothesis H of no change in the means. In addition, they calculated the power of these tests numerically via a Monte-Carlo simulation. In this paper, we will compute the power of the above test statistics under a local alternative. More specifically, we shall consider the following sequence of alternatives

(8) $$A_{\tau, h_n}(n) : \mu_i = \begin{cases} \mu & \text{for } i = 1, \ldots, [n\tau] \\ \mu + h_n & \text{for } i = [n\tau] + 1, \ldots, n, \end{cases}$$

where $0 \leq \tau \leq 1$. Observe that the level shift h_n depends on the sample size n.

2. Power of the CUSUM Test under Local Alternatives

We will first investigate the asymptotic distribution of the process

(9) $$D_n(\lambda) := \frac{1}{n \, d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} (X_j - X_i), \ 0 \leq \lambda \leq 1.$$

To do so, we consider the Hermite expansion of $G(\xi_i)$, namely

$$G(\xi_i) = \sum_{q=1}^{\infty} \frac{a_q}{q!} H_q(\xi_i),$$

where H_q is the q-th order Hermite polynomial. We define the Hermite rank of the function G as

$$m = \min\{q \geq 1 : a_q \neq 0\}$$

and introduce the normalization constants

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

We suppose $0 < D < \frac{1}{m}$, in which case

(11) $$d_n^2 \sim \kappa_m \, n^{2-mD} \, L^m(n),$$

where $\kappa_m = 2(m!)/(1 - Dm)(2 - Dm)$. Here we use the symbol $a_n \sim b_n$ to denote $a_n/b_n \to 0$ as $n \to \infty$.

Under the null hypothesis H of no level shift, we get that the process $(D_n(\lambda))_{0 \leq \lambda \leq 1}$ in (9) converges in distribution towards the process

$$(12) \qquad \frac{a_m}{m!}(\lambda Z_m(1) - Z_m(\lambda))_{0 \leq \lambda \leq 1},$$

where m denotes the Hermite rank of G and where

$$(13) \qquad a_m = E(H_m(\xi)G(\xi));$$

see Dehling, Rooch and Taqqu (2013), proof of Theorem 3. The process $(Z_m(\lambda))_{\lambda \geq 0}$ denotes the m-th order Hermite process with Hurst parameter $H = 1 - \frac{Dm}{2} \in (\frac{1}{2}, 1)$. It is Gaussian (namely fractional Brownian motion) when $m = 1$, but it is non-Gaussian when $m \geq 2$. For various representations of the Hermite process $(Z_m(\lambda))_{\lambda \geq 0}$, see Pipiras and Taqqu (2010).

In view of (3), under the alternative A in (5), we need to consider

$$(14) \qquad D_n(\lambda) = \frac{1}{n\, d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} (G(\xi_j) - G(\xi_i)) + \frac{1}{n\, d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} (\mu_j - \mu_i).$$

Observe that the statistic $D_n(\lambda)$ presumes that the jump occurs at time $[n\lambda] + 1$, whereas the local alternative $A_{\tau,h_n}(n)$ involves a jump at $[n\tau]+1$. There will therefore be an interplay between λ and τ. In fact, under the local alternative $A_{\tau,h_n}(n)$ in (8), we get

$$(15) \qquad \frac{1}{n\, d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} (\mu_j - \mu_i) = \begin{cases} \frac{h_n}{n\, d_n}[\lambda n](n - [\tau n]) & \text{for } \lambda \leq \tau \\ \frac{h_n}{n\, d_n}(n - [\lambda n])[\tau n] & \text{for } \lambda \geq \tau. \end{cases}$$
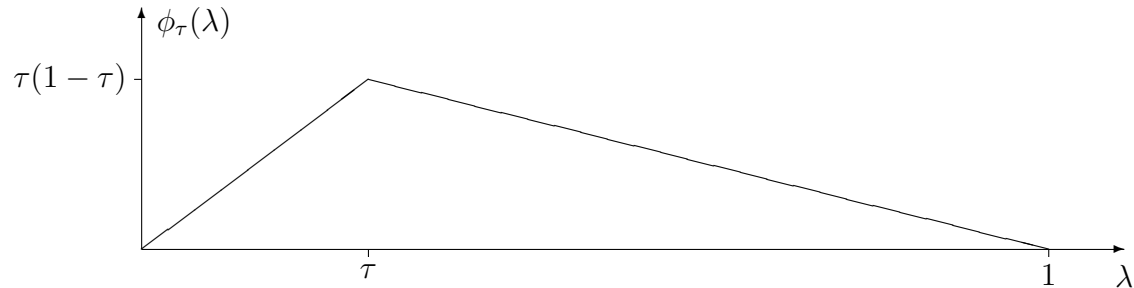

FIGURE 1. Graph of the function ϕ_τ; see (16).

We introduce the function $\phi_\tau : [0,1] \to \mathbb{R}$ by

$$(16) \qquad \phi_\tau(\lambda) = \begin{cases} \lambda(1 - \tau) & \text{for } \lambda \leq \tau \\ (1 - \lambda)\tau & \text{for } \lambda \geq \tau, \end{cases}$$

which takes its maximum value $\tau(1 - \tau)$ at $\lambda = \tau$; see Figure 1. Note that for large n, we get

$$(17) \qquad \frac{1}{n\, d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} (\mu_j - \mu_i) \sim \frac{n\, h_n}{d_n} \phi_\tau(\lambda).$$

Thus, in order for the second term in (14) to converge as $n \to \infty$, we have to choose the level shift $h_n \sim c\, d_n/n$. When n is large, this is exactly the order of the level shift that can be detected with a nontrivial power, that is with a power which is neither 0 nor 1.

Theorem 2.1. *Let $(\xi_i)_{i\geq 1}$ be a stationary Gaussian process with mean zero, variance 1 and autocovariance function as in (4) with $0 < D < \frac{1}{m}$. Moreover, let $G : \mathbb{R} \to \mathbb{R}$ be a measurable function satisfying $E(G^2(\xi)) < \infty$ and define $X_i = \mu_i + G(\xi_i)$. Then under the local alternative $A_{\tau,h_n}(n)$ with*

$$(18) \qquad h_n \sim \frac{d_n}{n} c,$$

for an arbitrary constant c, the process $(D_n(\lambda))_{0\leq\lambda\leq 1}$ in (14) converges in distribution to the process

$$(19) \qquad \left(\frac{a_m}{m!}\left(\lambda Z_m(1) - Z_m(\lambda)\right) + c\,\phi_\tau(\lambda)\right)_{0\leq\lambda\leq 1},$$

where $(Z_m(\lambda))_{\lambda\geq 0}$ denotes the m-th order Hermite process with Hurst parameter $H = 1 - \frac{Dm}{2} \in (\frac{1}{2}, 1)$, where a_m is given by (13) and $\phi_\tau(\lambda)$ by (16).

Proof. We use the decomposition (14). The first term on the right hand side has the same distribution as $D_n(\lambda)$ under the hypothesis, and thus converges in distribution to $\frac{a_m}{m!}(\lambda Z_m(1) - Z_m(\lambda))$. Regarding the second term, we observe that by (18) and (15) we get

$$\frac{1}{n\,d_n}\sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n}(\mu_j - \mu_i) \quad \sim \quad \begin{cases} \frac{c}{n^2}[\lambda n](n - [\tau n]) & \text{for } \lambda \leq \tau \\ \frac{c}{n^2}(n - [\lambda n])[\tau n] & \text{for } \lambda \geq \tau. \end{cases}$$

$$\to \quad c\,\phi_\tau(\lambda),$$

uniformly in $\lambda \in [0,1]$, as $n \to \infty$. $\qquad\square$

Remark 2.2. (i) Observe that for $c = 0$ we recover the limit distribution under the null hypothesis. Thus, Theorem 2.1 is a generalization of the results obtained previously under the null hypothesis. The limit process is a fractional bridge process. When $m = 1$, this process is a fractional Gaussian bridge. For $m > 1$, the process is non-Gaussian.

(ii) Under the local alternative, i.e. when $c \neq 0$, the limit process is the sum of a fractional bridge process and the deterministic function $c\,\phi_\tau$.

As an application of the continuous mapping theorem, we obtain the following corollary.

Corollary 2.3. *Under the alternative $A_{\tau,h_n}(n)$ with $h_n \sim \frac{d_n}{n}c$, D_n as defined in (7) converges in distribution to*

$$(20) \qquad \sup_{0\leq\lambda\leq 1}\left|\frac{a_m}{m!}\left(\lambda Z_m(1) - Z_m(\lambda)\right) + c\,\phi_\tau(\lambda)\right|.$$

Remark 2.4. (i) The limit distribution (20) depends on the constant c. For $c = 0$, we obtain the limit distribution under the null hypothesis. Quantiles of this limit distribution were calculated numerically via a Monte-Carlo simulation by Dehling, Rooch and Taqqu (2013), Table 1. Increasing the value of $|c|$ leads to a shift of the distribution to the right. If $c = \infty$, that is, if h_n tends slower to zero than $\frac{d_n}{n}c$ for any $c > 0$, then the correct normalization for $D_n(\lambda)$ should go to ∞ at a higher rate which would kill the random part $(\lambda Z_m(1) - Z_m(\lambda))$ in (19), and hence the level shift could be detected precisely. The power of the asymptotic test would be equal to 1 in this case.

(ii) For a given $\tau \in [0,1]$, the function $\phi_\tau(\lambda)$ takes its maximum value in $\lambda = \tau$, and this maximum value equals $\tau(1 - \tau)$. Thus, for values of τ close to 0 and close to 1, $\tau(1 - \tau)$ is close to 0, and thus the effect of adding the term $c\phi_\tau(\lambda)$ is rather small. As a result, the power of the test is small at level shifts that occur very early or very late in the process.

(iii) The higher the level shift and the closer λ is to τ, the easier it is to detect the level shift.

(iv) If the observations are short-range dependent, one can typically detect level shifts h_n of size $\frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}$, but here, because of long-range dependence, the level shifts that can be detected are of smaller order $\frac{d_n}{n} \sim \frac{cn^{1-Dm/2}L(n)}{n} = cn^{-Dm/2}L(n)$; note that $Dm < 1$.

We will now apply Corollary 2.3 in order to make power calculations for the change-point test that rejects for large values of D_n. Under the null hypothesis of no level shift,

$$D_n \xrightarrow{\mathcal{D}} \sup_{0 \leq \lambda \leq 1} \frac{|a_m|}{m!} |\lambda Z_m(1) - Z_m(\lambda)|.$$

If we denote by q_α the upper α-quantile of the distribution of $\sup_{0 \leq \lambda \leq 1} |\lambda Z_m(1) - Z_m(\lambda)|$, we obtain

$$\lim_{n \to \infty} P_H \left(D_n \geq \frac{|a_m|}{m!} q_\alpha \right) = P \left(\sup_{0 \leq \lambda \leq 1} \frac{|a_m|}{m!} |\lambda Z_m(1) - Z_m(\lambda)| \geq \frac{|a_m|}{m!} q_\alpha \right) = \alpha,$$

where P_H indicates the probability under the null hypothesis H. Thus, the test that rejects the null hypothesis H when $D_n \geq \frac{|a_m|}{m!} q_\alpha$ has asymptotic level α. If h_n is chosen as in (18), we obtain under the alternative $A_{\tau, h_n}(n)$

(21)
$$\lim_{n \to \infty} P_{A_{\tau, h_n}(n)} \left(D_n \geq \frac{|a_m|}{m!} q_\alpha \right) = P \left(\sup_{0 \leq \lambda \leq 1} \left| \frac{a_m}{m!} (\lambda Z_m(1) - Z_m(\lambda)) + c\,\phi_\tau(\lambda) \right| \geq \frac{|a_m|}{m!} q_\alpha \right).$$

Thus, for large n, the power of our test at the alternative $A_{\tau, h_n}(n)$ is approximately given by the right-hand side of (21).

We may also apply Corollary 2.3 in order to determine the size of a level shift at time $[\tau n]$ that can be detected with a given probability β. First, we calculate $c = c(\alpha, \beta)$ such that

(22)
$$P \left(\sup_{0 \leq \lambda \leq 1} \left| \frac{a_m}{m!} (\lambda Z_m(1) - Z_m(\lambda)) + c\,\phi_\tau(\lambda) \right| \geq \frac{|a_m|}{m!} q_\alpha \right) = \beta.$$

Thus, by (21), we get that the asymptotic power of the test at the alternative $A_{\tau, h_n}(n)$ is equal to β. Thus, given a sample size n, we can detect a level shift of size $h_n = \frac{d_n}{n} c(\alpha, \beta)$ at time $[\tau n]$ with probability β with a level α test based on the test statistic D_n.

3. Power of the Wilcoxon Change-Point Test under Local Alternatives

In the context of the Wilcoxon change-point test, the Hermite rank is not that of the function G, but of the class of functions

(23)
$$1_{\{G(\xi_i) \leq x\}} - F(x), \ x \in \mathbb{R},$$

where $F(x) = E(1_{\{G(\xi_i) \leq x\}}) = P(G(\xi_i) \leq x)$. We define the Hermite expansion of the class of functions (23) as

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

where H_q is again the q-th order Hermite polynomial and where the coefficients are

(24)
$$J_q(x) = E\left(H_q(\xi_i) 1_{\{G(\xi_i) \leq x\}} \right).$$

We define the Hermite rank of the class of functions (23) as

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

Theorem 3.1. *Suppose that $(\xi_i)_{i \geq 1}$ is a stationary Gaussian process with mean zero, variance 1 and autocovariance function as in (4) with $0 \leq D < \frac{1}{m}$. Moreover, let $G : \mathbb{R} \to \mathbb{R}$ be a measurable function, and assume that $G(\xi_k)$ has continuous distribution function $F(x)$. Let m denote the Hermite rank of the class of functions (23), let d_n be as in (11), and let the level shift h_n be as in (18). Then, under the sequence of alternatives A_{τ, h_n}, defined in (8), if $h_n \to 0$ as $n \to \infty$, the process*

$$(25) \quad \left(\frac{1}{n \, d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} \left(1_{\{X_i \leq X_j\}} - \frac{1}{2} \right) - \frac{n}{d_n} \phi_\tau(\lambda) \int_{\mathbb{R}} (F(x + h_n) - F(x)) dF(x) \right)_{0 \leq \lambda \leq 1}$$

converges in distribution towards the process

$$(26) \quad \left(\frac{\int_{\mathbb{R}} J_m(x) dF(x)}{m!} \big(Z_m(\lambda) - \lambda Z_m(1) \big) \right)_{0 \leq \lambda \leq 1},$$

where $(Z_m(\lambda))_{\lambda \geq 0}$ denotes the m-th order Hermite process with Hurst parameter $H = 1 - \frac{Dm}{2} \in (\frac{1}{2}, 1)$ and where $J_m(x)$ is defined as in (24).

Remark 3.2. (i) The normalization d_n and the processes $(Z_m(\lambda))_{\lambda \geq 0}$ in Theorem 2.1 and Theorem 3.1 are the same.

(ii) Since, by assumption, the distribution $F(x)$ of $G(\xi_k)$ is continuous, it follows from integration by parts that $\int_{\mathbb{R}} F(x) dF(x) = \frac{1}{2}$. This explains the $1/2$ in (25) because $\int_{\mathbb{R}} F(x) dF(x) = E(1_{\{X_1 \leq X_1'\}})$, where X_1' is an independent copy of X_1. The independence assumption is reasonable as the dependence between X_i and X_j vanishes asymptotically when $|i - j| \to \infty$.

(iii) As noted at the beginning of the proof, the first part of (25) converges to (26) under the null hypothesis. We show in the proof that the second part of (25) compensates for the presence of the alternatives A_{τ, h_n}.

(iv) We make no assumption about the exact order of the sequence $(h_n)_{n \geq 1}$. Theorem 3.1 holds under the very general assumption that $h_n \to 0$, as $n \to \infty$.

(v) If we choose $(h_n)_{n \geq 1}$ as in (18), the centering constants in (25) converge, provided some technical assumptions are satisfied. To see this, observe that

$$\frac{n}{d_n} \phi_\tau(\lambda) \int_{\mathbb{R}} (F(x + h_n) - F(x)) dF(x) \quad \sim \quad \frac{n \, h_n}{d_n} \phi_\tau(\lambda) \int_{\mathbb{R}} \frac{F(x + h_n) - F(x)}{h_n} dF(x)$$

$$\to \quad c \, \phi_\tau(\lambda) \int_{\mathbb{R}} f(x) dF(x)$$

$$= \quad c \, \phi_\tau(\lambda) \int_{\mathbb{R}} f^2(x) dx.$$

The convergence in the next to last step requires some justification – this holds, e.g. if F is differentiable with bounded derivative $f(x)$.

Corollary 3.3. *Suppose that $(\xi_i)_{i \geq 1}$ is a stationary Gaussian process with mean zero, variance 1 and autocovariance function as in (4) with $0 \leq D < \frac{1}{m}$. Moreover, let $G : \mathbb{R} \to \mathbb{R}$ be a measurable function, and assume that $G(\xi_k)$ has a distribution function $F(x)$ with bounded density $f(x)$. Let m denote the Hermite rank of the class of functions $1_{\{G(\xi_i) \leq x\}} - F(x)$, $x \in \mathbb{R}$. Then, under the sequence of alternatives A_{τ, h_n}, defined in (8), with $h_n \sim c\frac{d_n}{n}$ we obtain that*

$$(27) \quad \left(\frac{1}{n \, d_n} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} \left(1_{\{X_i \leq X_j\}} - \frac{1}{2} \right) \right)_{0 \leq \lambda \leq 1}$$

converges in distribution to the process

$$\left(\frac{\int_{\mathbb{R}} J_m(x)dF(x)}{m!}(Z_m(\lambda) - \lambda Z_m(1)) + c\phi_\tau(\lambda)\int_{\mathbb{R}} f^2(x)dx\right)_{0 \leq \lambda \leq 1}.$$

Proof of Theorem 3.1. In our proof, we will make use of the limit theorem that was derived in Dehling, Rooch and Taqqu (2013) under the null hypothesis. They showed (see Theorem 1) that

$$(28) \qquad \frac{1}{n\,d_n}\sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n}\left(1_{\{G(\xi_i)\leq G(\xi_j)\}} - \frac{1}{2}\right) \to \frac{\int_{\mathbb{R}} J_m(x)dF(x)}{m!}(Z_m(\lambda) - \lambda Z_m(1)).$$

In order to make use of this result, we will decompose the test statistic into a term whose distribution is the same both under the null hypothesis as well as under the alternative, and a second term which, after proper centering converges to zero. As in Dehling, Rooch and Taqqu (2013), we will express the test statistic as a functional of the empirical distribution function of the $G(\xi_i)$, namely

$$F_k(x) = \frac{1}{k}\sum_{i=1}^{k} 1_{\{G(\xi_i)\leq x\}}.$$

Given integers k, l with $k \leq l$ we denote by $F_{k,l}(x)$ the empirical distribution function based on $G(\xi_k), \ldots, G(\xi_l)$, i.e.

$$(29) \qquad F_{k,l}(x) = \frac{1}{l-k+1}\sum_{i=k}^{l} 1_{\{G(\xi_i)\leq x\}}.$$

Recall that under the local alternative, we have

$$X_i = \begin{cases} G(\xi_i) + \mu & \text{for } i = 1, \ldots, [n\tau] \\ G(\xi_i) + \mu + h_n & \text{for } i = [n\tau] + 1, \ldots, n. \end{cases}$$

Thus, we obtain for $\lambda \leq \tau$,

$$
\begin{aligned}
\sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n} 1_{\{X_i\leq X_j\}} &= \sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{[n\tau]} 1_{\{G(\xi_i)+\mu\leq G(\xi_j)+\mu\}} + \sum_{i=1}^{[n\lambda]}\sum_{j=[n\tau]+1}^{n} 1_{\{G(\xi_i)+\mu\leq G(\xi_j)+\mu+h_n\}} \\
&= \sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{[n\tau]} 1_{\{G(\xi_i)\leq G(\xi_j)\}} + \sum_{i=1}^{[n\lambda]}\sum_{j=[n\tau]+1}^{n} 1_{\{G(\xi_i)\leq G(\xi_j)+h_n\}} \\
&= \sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n} 1_{\{G(\xi_i)\leq G(\xi_j)\}} + \sum_{i=1}^{[n\lambda]}\sum_{j=[n\tau]+1}^{n} \left(1_{\{G(\xi_i)\leq G(\xi_j)+h_n\}} - 1_{\{G(\xi_i)\leq G(\xi_j)\}}\right) \\
(30) \qquad &= \sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n} 1_{\{G(\xi_i)\leq G(\xi_j)\}} + \sum_{i=1}^{[n\lambda]}\sum_{j=[n\tau]+1}^{n} 1_{\{G(\xi_j)<G(\xi_i)\leq G(\xi_j)+h_n\}}.
\end{aligned}
$$

In the same way, we obtain for $\lambda \geq \tau$,

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

$$= \sum_{i=1}^{[n\tau]} \sum_{j=[n\lambda]+1}^{n} 1_{\{G(\xi_i)+\mu \leq G(\xi_j)+\mu+h_n\}} + \sum_{i=[n\tau]+1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} 1_{\{G(\xi_i)+\mu+h_n \leq G(\xi_j)+\mu+h_n\}}$$

$$= \sum_{i=1}^{[n\tau]} \sum_{j=[n\lambda]+1}^{n} 1_{\{G(\xi_i) \leq G(\xi_j)+h_n\}} + \sum_{i=[n\tau]+1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} 1_{\{G(\xi_i) \leq G(\xi_j)\}}$$

$$= \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} 1_{\{G(\xi_i) \leq G(\xi_j)\}} + \sum_{i=1}^{[n\tau]} \sum_{j=[n\lambda]+1}^{n} \left(1_{\{G(\xi_i) \leq G(\xi_j)+h_n\}} - 1_{\{G(\xi_i) \leq G(\xi_j)\}}\right)$$

$$(31) \qquad = \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} 1_{\{G(\xi_i) \leq G(\xi_j)\}} + \sum_{i=1}^{[n\tau]} \sum_{j=[n\lambda]+1}^{n} 1_{\{G(\xi_j) < G(\xi_i) \leq G(\xi_j)+h_n\}}$$

Thus, in order to prove Theorem 3.1, it suffices to show that the following two terms,

$$(32) \quad \frac{1}{n\,d_n} \sup_{0 \leq \lambda \leq \tau} \left| \sum_{i=1}^{[n\lambda]} \sum_{j=[n\tau]+1}^{n} 1_{\{G(\xi_j) < G(\xi_i) \leq G(\xi_j)+h_n\}} - n^2 \lambda(1-\tau) \int_{\mathbb{R}} (F(x+h_n)-F(x))dF(x) \right|$$

and

$$(33) \quad \frac{1}{n\,d_n} \sup_{\tau \leq \lambda \leq 1} \left| \sum_{i=1}^{[n\tau]} \sum_{j=[n\lambda]+1}^{n} 1_{\{G(\xi_j) < G(\xi_i) \leq G(\xi_j)+h_n\}} - n^2 \tau(1-\lambda) \int_{\mathbb{R}} (F(x+h_n)-F(x))dF(x) \right|$$

both converge to zero in probability. We first show this for (32). Observe that

$$\sum_{i=1}^{[n\lambda]} \sum_{j=[n\tau]+1}^{n} 1_{\{G(\xi_j) < G(\xi_i) \leq G(\xi_j)+h_n\}} - n^2 \lambda(1-\tau) \int_{\mathbb{R}} (F(x+h_n) - F(x))\, dF(x)$$

$$= [n\lambda] \sum_{j=[n\tau]+1}^{n} \left(F_{[n\lambda]}(G(\xi_i)+h_n) - F_{[n\lambda]}(G(\xi_i))\right)$$

$$\quad - n^2 \lambda(1-\tau) \int_{\mathbb{R}} (F(x+h_n) - F(x))\, dF(x)$$

$$= [n\lambda](n-[n\tau]) \int_{\mathbb{R}} \left(F_{[n\lambda]}(x+h_n) - F_{[n\lambda]}(x)\right) dF_{[n\tau]+1,n}(x)$$

$$\quad - n^2 \lambda(1-\tau) \int_{\mathbb{R}} (F(x+h_n) - F(x))\, dF(x)$$

$$= [n\lambda](n-[n\tau]) \left(\int_{\mathbb{R}} (F_{[n\lambda]}(x+h_n) - F_{[n\lambda]}(x))dF_{[n\tau]+1,n}(x) - \int_{\mathbb{R}} (F(x+h_n) - F(x))dF(x) \right)$$

$$\quad + \left([n\lambda](n-[n\tau]) - n^2 \lambda(1-\tau)\right) \int_{\mathbb{R}} (F(x+h_n) - F(x))dF(x)$$

Note that $|[n\lambda](n-[n\tau]) - n^2\lambda(1-\tau)| \leq n$ and $|\int_{\mathbb{R}}(F(x+h_n) - F(x))dF(x)| \leq 1$. Thus

$$\frac{1}{n\,d_n} \left([n\lambda](n-[n\tau]) - n^2\lambda(1-\tau)\right) \int_{\mathbb{R}} (F(x+h_n) - F(x))dF(x) \leq \frac{1}{d_n} \to 0,$$

as $n \to \infty$. Hence, in order to show that (32) converges to zero in probability, it suffices to show that

(34)
$$\frac{1}{n\, d_n}[n\lambda](n-[n\tau]) \left(\int_{\mathbb{R}} (F_{[n\lambda]}(x+h_n) - F_{[n\lambda]}(x)) dF_{[n\tau]+1,n}(x) - \int_{\mathbb{R}} (F(x+h_n) - F(x)) dF(x) \right)$$

converges to zero, in probability. In order to prove this, we rewrite the difference of the integrals in (34) as

$$(35) \quad \int_{\mathbb{R}} \left(F_{[n\lambda]}(x+h_n) - F_{[n\lambda]}(x)\right) dF_{[n\tau]+1,n}(x) - \int_{\mathbb{R}} (F(x+h_n) - F(x)) \, dF(x)$$

$$= \int_{\mathbb{R}} \left(F_{[n\lambda]}(x+h_n) - F_{[n\lambda]}(x)\right) - (F(x+h_n) - F(x)) \, dF_{[n\tau]+1,n}(x)$$

$$+ \int_{\mathbb{R}} (F(x+h_n) - F(x)) \, d(F_{[n\tau]+1,n} - F)(x)$$

$$= \int_{\mathbb{R}} \left(F_{[n\lambda]}(x+h_n) - F_{[n\lambda]}(x)\right) - (F(x+h_n) - F(x)) \, dF_{[n\tau]+1,n}(x)$$

$$- \int_{\mathbb{R}} \left(F_{[n\tau]+1,n}(x) - F(x)\right) d(F(x+h_n) - F(x)),$$

where we have used integration by parts in the final step. Thus, in order to prove that (34) converges to zero, it suffices to show that the following two terms converge in probability, as $n \to 0$,

$$(36) \quad \frac{1}{d_n}[n\lambda] \int_{\mathbb{R}} \left((F_{[n\lambda]}(x+h_n) - F_{[n\lambda]}(x)) - (F(x+h_n) - F(x))\right) dF_{[n\tau]+1,n}(x) \to 0$$

$$(37) \quad \frac{1}{d_n}(n - [n\tau]) \int_{\mathbb{R}} \left(F_{[n\tau]+1,n}(x) - F(x)\right) d(F(x+h_n) - F(x)) \to 0.$$

In order to prove (36) and (37), we now apply the empirical process non-central limit theorem of Dehling and Taqqu (1989) which states that

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

where

$$J(x) = J_m(x) = E\left(1_{\{G(\xi_i)\leq x\}}H_m(\xi_i)\right) \quad \text{and} \quad Z(\lambda) = \frac{Z_m(\lambda)}{m!}.$$

By the Dudley-Wichura version of Skorohod's representation theorem (see Shorack and Wellner (1986), Theorem 2.3.4) 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.

$$(38) \quad \sup_{\lambda\in[0,1],x\in\mathbb{R}} \left|d_n^{-1}[n\lambda](F_{[n\lambda]}(x) - F(x)) - J(x)Z(\lambda)\right| \longrightarrow 0 \quad a.s.$$

Note that by definition, for any $\lambda \leq \tau$

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

Hence, we may deduce from (38) the following limit theorem for the empirical distribution of the observations $X_{[n\lambda]+1}, \ldots, X_{[n\tau]}$,

$$(39) \quad \sup_{0\leq\lambda\leq\tau,x\in\mathbb{R}} \left|d_n^{-1}([n\tau] - [n\lambda])(F_{[n\lambda]+1,[n\tau]}(x) - F(x)) - J(x)(Z(\tau) - Z(\lambda))\right| \to 0,$$

almost surely. In the same way, we obtain

$$(40) \qquad \sup_{0\leq\lambda\leq1, x\in\mathbb{R}} \left| d_n^{-1}(n-[n\lambda])(F_{[n\lambda]+1,n}-F(x)) - J(x)(Z(1)-Z(\lambda)) \right| \to 0,$$

almost surely. Now we return to (36) and write

$$(41) \qquad \left| \int_{\mathbb{R}} \frac{1}{d_n}[n\lambda]\left((F_{[n\lambda]}(x+h_n) - F_{[n\lambda]}(x)) - (F(x+h_n) - F(x))\right) dF_{[n\tau]+1,n}(x) \right|$$

$$\leq \left| \int_{\mathbb{R}} (J(x+h_n) - J(x))Z(\lambda) dF_{[n\tau]+1,n}(x) \right|$$

$$+ \sup_{x\in\mathbb{R}, 0\leq\lambda\leq1} \left| \frac{1}{d_n}[n\lambda]\left((F_{[n\lambda]}(x+h_n) - F_{[n\lambda]}(x)) - (F(x+h_n) - F(x))\right) \right.$$

$$\left. -(J(x+h_n) - J(x))Z(\lambda) \right|$$

$$\leq \left| \int_{\mathbb{R}} (J(x+h_n) - J(x)) dF_{[n\tau]+1,n}(x) \right| \sup_{0\leq\lambda\leq1} |Z(\lambda)|$$

$$+ \sup_{x\in\mathbb{R}, 0\leq\lambda\leq1} \left| \frac{1}{d_n}[n\lambda]\left((F_{[n\lambda]}(x+h_n) - F_{[n\lambda]}(x)) - (F(x+h_n) - F(x))\right) \right.$$

$$\left. -(J(x+h_n) - J(x))Z(\lambda) \right|.$$

The second term on the right-hand side converges to zero by (38). Concerning the first term, note that

$$(42) \qquad J(x) = \int_{\mathbb{R}} 1_{\{G(y)\leq x\}} H_m(y)\phi(y) dy = - \int_{\mathbb{R}} 1_{\{x\leq G(y)\}} H_m(y)\phi(y) dy,$$

where $\phi(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ denotes the standard normal density function. For the second identity, we have used the fact that $G(\xi)$, by assumption, has a continuous distribution, and that $\int_{\mathbb{R}} H_m(y)\phi(y) dy = 0$, for $m \geq 1$. Using (42) we thus obtain

$$(43) \qquad \int_{\mathbb{R}} J(x) dF_{[n\tau]+1,n}(x) = -\int_{\mathbb{R}}\int_{\mathbb{R}} 1_{\{x\leq G(y)\}} H_m(y)\phi(y) dy \, dF_{[n\tau]+1,n}(x)$$

$$= -\int_{\mathbb{R}}\int_{\mathbb{R}} 1_{\{x\leq G(y)\}} dF_{[n\tau]+1,n}(x) H_m(y)\phi(y) dy$$

$$= -\int_{\mathbb{R}} F_{[n\tau]+1,n}(G(y)) H_m(y)\phi(y) dy,$$

and, using analogous arguments,

$$(44) \qquad \int_{\mathbb{R}} J(x+h_n) dF_{[n\tau]+1,n}(x) = -\int_{\mathbb{R}} F_{[n\tau]+1,n}(G(y) - h_n) H_m(y)\phi(y) dy.$$

By the Glivenko-Cantelli theorem, applied to the stationary, ergodic process $(G(\xi_i))_{i\geq1}$, we get $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0$, almost surely. Since

$$F_{[n\tau]+1,n}(x) = \frac{n}{n-[n\tau]} F_n(x) - \frac{[n\tau]}{n-[n\tau]} F_{[n\tau]}(x),$$

we get that, almost surely,

$$(45) \qquad \sup_{x\in\mathbb{R}} \left| F_{[n\tau]+1,n}(x) - F(x) \right| \to 0.$$

Returning to the first term on the right-hand side of (41), we obtain, using (43) and (44),

$$\left| \int_{\mathbb{R}} (J(x + h_n) - J(x)) \, dF_{[n\tau]+1,n}(x) \right|$$

$$= \left| \int_{\mathbb{R}} \left(F_{[n\tau]+1,n}(G(y) - h_n) - F_{[n\tau]+1,n}(G(y)) \right) H_m(y)\phi(y)dy \right|$$

$$\leq \int_{\mathbb{R}} |F(G(y) - h_n) - F(G(y))| \, |H_m(y)| \, \phi(y)dy$$

$$+ 2 \sup_x \left| F_{[n\tau]+1,n}(x) - F(x) \right| \int_{\mathbb{R}} |H_m(y)| \, \phi(y)dy.$$

Both terms on the right-hand-side converge to zero; the second one by (45), the first one by continuity of F, the fact that $h_n \to 0$, and Lebesgue's dominated convergence theorem. In both cases, we have made use of the fact that $\int |H_m(y)|\phi(y)dy < \infty$. Thus we have finally established (36). In order to prove (37), we observe that

$$\frac{1}{d_n}(n - [n\tau]) \int_{\mathbb{R}} \left(F_{[n\tau]+1,n}(x) - F(x) \right) d(F(x + h_n) - F(x))$$

$$\leq \left| \int_{\mathbb{R}} J(x)(Z(\tau) - Z(1))d(F(x + h_n) - F(x)) \right|$$

$$+ \sup_{x \in \mathbb{R}} \left| \frac{1}{d_n}(n - [n\tau])(F_{[n\tau]+1,n}(x) - F(x)) - J(x)(Z(\tau) - Z(1)) \right|$$

$$\leq \left| \int_{\mathbb{R}} J(x)d(F(x + h_n) - F(x)) \right| |Z(\tau) - Z(1)|$$

$$+ \sup_{x \in \mathbb{R}} \left| \frac{1}{d_n}(n - [n\tau])(F_{[n\tau]+1,n}(x) - F(x)) - J(x)(Z(\tau) - Z(1)) \right|.$$

The second term on the right-hand side converges to zero, by (40). Concerning the first term, note that

$$\int_{\mathbb{R}} J(x)d(F(x + h_n) - F(x)) = E\left(J(G(\xi_i) - h_n) - J(G(\xi_i)) \right).$$

Applying Lebesgue's dominated convergence theorem and making use of the fact that, by assumption, J is continuous, we obtain that $\int_{\mathbb{R}} J(x)d(F(x + h_n) - F(x)) \to 0$. In this way, we have finally proved that (32) converges to zero, in probability. By similar arguments, we can prove this for (33), which finally ends the proof of Theorem 3.1. □

4. ARE OF THE WILCOXON CHANGE-POINT TEST AND THE CUSUM TEST FOR LRD DATA

In this section, we calculate the asymptotic relative efficiency (ARE) of the Wilcoxon change-point test with respect to the CUSUM test. To do so, we calculate the number of observations needed to detect a small level shift h at time $[\tau n]$ with a test of given level α and given power β, both for the Wilcoxon change-point test and the CUSUM test, and denote these numbers by n_W and n_C, respectively. We then define the asymptotic relative efficiency of the Wilcoxon change-point test with respect to the CUSUM test by

$$(46) \qquad\qquad ARE(W, C) = \lim_{h \to 0} \frac{n_C}{n_W}.$$

It will turn out that the limit (46) exists and that the asymptotic relative efficiency does not depend on the choice of τ, α, β. If this limit is larger than 1, then the CUSUM test requires

a larger sample size to detect the level shift, and hence the Wilcoxon change-point test is (asymptotically) more efficient.

In the remaining part of this section, we will focus on the case when $m = 1$ both for the CUSUM as well as the Wilcoxon change-point test, i.e. when the Hermite rank of $G(\xi_1)$ and of the class of functions $1_{\{G(\xi_1) \leq x\}} - F(x), x \in \mathbb{R}$, are both equal to 1. This is the case, for example, when G is a strictly monotone function. In this case

$$\int_{\mathbb{R}} J_1(x) dF(x) = -\frac{1}{2\sqrt{\pi}},$$

see Relation (20) in Dehling, Rooch and Taqqu (2013), showing that the Hermite rank of the class of functions $1_{\{G(\xi_1) \leq x\}} - F(x), x \in \mathbb{R}$ equals 1. Focusing now on $G(\xi_1)$ and using integration by parts, we get that the first order Hermite coefficient a_1 of G equals

$$a_1 = E(G(\xi_1)\xi_1) = \int_{\mathbb{R}} G(x)x\phi(x)dx = -\int_{\mathbb{R}} G(x)\phi'(x)dx = \int_{\mathbb{R}} \phi(x)dG(x) > 0,$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ denotes the standard normal density function. Thus, the Hermite rank of $G(\xi_i)$ equals 1, as well.

In this case, i.e. when $m = 1$, the Hermite process arising as limit in Theorem 2.1, Theorem 3.1 and Corollary 3.3 is fractional Brownian motion $(B_H(\lambda))_{0 \leq \lambda \leq 1}$. Note that fractional Brownian motion is symmetric, i.e. $(-B_H(\lambda))_{0 \leq \lambda \leq 1}$ has the same distribution as $(B_H(\lambda))_{0 \leq \lambda \leq 1}$. Thus the limit processes in Theorem 2.1 and Corollary 3.1 can also be written as

$$(47) \qquad \left(|a_1|(B_H(\lambda) - \lambda B_H(1)) + c\,\phi_\tau(\lambda) \right)_{0 \leq \lambda \leq 1},$$

$$(48) \qquad \left(\left| \int_{\mathbb{R}} J_1(x) dF(x) \right| (B_H(\lambda) - \lambda B_H(1)) + c\,\phi_\tau(\lambda) \int_{\mathbb{R}} f^2(x)dx \right)_{0 \leq \lambda \leq 1}.$$

As preparation, we first calculate a quantity that is related to the asymptotic relative efficiency, namely the ratio of the sizes of level shifts that can be detected by the two tests, based on the same number of observations n, again for given values of τ, α, β. We denote the corresponding level shifts by $h_W(n)$ and $h_C(n)$, respectively, assuming that these numbers depend on n in the following way:

$$(49) \qquad h_W(n) \;=\; c_W \frac{d_n}{n}$$

$$(50) \qquad h_C(n) \;=\; c_C \frac{d_n}{n}.$$

In order to simplify the following considerations, we take a one-sided change-point test, thus rejecting the hypothesis of no change-point for large values of

$$\max_{1 \leq k \leq n-1} \sum_{i=1}^{k} \sum_{j=k+1}^{n} (X_j - X_i)$$

or

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

respectively. These are the appropriate tests when testing against the alternative of a non-negative level shift. In order to obtain tests that have asymptotically level α, the CUSUM

test and the Wilcoxon change-point test reject the null-hypothesis when

$$(51) \qquad \frac{1}{n\, d_n\, |a_1|} \max_{1 \le k \le n-1} \sum_{i=1}^{k} \sum_{j=k+1}^{n} (X_j - X_i) \ge q_\alpha,$$

$$(52) \qquad \frac{1}{n\, d_n\, |\int_{\mathbb{R}} J_1(x) dF(x)|} \max_{k=1,\dots,n-1} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(1_{\{X_i \le X_j\}} - \frac{1}{2} \right) \ge q_\alpha,$$

where q_α denotes the upper α quantile of the distribution of $\sup_{0 \le \lambda \le 1}(B_H(\lambda) - \lambda B_H(1))$. This follows from Theorem 2.1 and Corollary 3.3 after applying the continuous mapping theorem. The constants a_1 and the functions $J_1(x)$ are defined in (13) and (24), respectively, and have just been computed. Under the sequence of alternatives $A_{\tau, h_C(n)}$, the asymptotic distribution of the test statistic in (51) is given by

$$\sup_{0 \le \lambda \le 1} \left(B_H(\lambda) - \lambda B_H(1) + \frac{c_C}{|a_1|} \phi_\tau(\lambda) \right);$$

see Theorem 2.1. Under the sequence of alternatives $A_{\tau, h_W(n)}$, the asymptotic distribution of the test statistic in (52) is given by

$$\sup_{0 \le \lambda \le 1} \left(B_H(\lambda) - \lambda B_H(1) + \frac{c_W \int_{\mathbb{R}} f^2(x) dx}{|\int_{\mathbb{R}} J_1(x) dF(x)|} \phi_\tau(\lambda) \right);$$

see Corollary 3.3. Thus, the asymptotic power of the CUSUM test is given by

$$\lim_{n \to \infty} P_{A_{\tau, h_C(n)}} \left(\frac{1}{n\, d_n\, |a_1|} \max_{1 \le k \le n-1} \sum_{i=1}^{k} \sum_{j=k+1}^{n} (X_j - X_i) \ge q_\alpha \right)$$

$$(53) \qquad = P \left(\sup_{0 \le \lambda \le 1} \left(B_H(\lambda) - \lambda B_H(1) + \frac{c_C}{|a_1|} \phi_\tau(\lambda) \right) \ge q_\alpha \right).$$

In the same way, we obtain the power of the Wilcoxon change-point test

$$\lim_{n \to \infty} P_{A_{\tau, h_W(n)}} \left(\frac{1}{n\, d_n\, |\int_{\mathbb{R}} J_1(x) dF(x)|} \max_{1 \le k \le n-1} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(1_{\{X_i \le X_j\}} - \frac{1}{2} \right) \ge q_\alpha \right)$$

$$(54) \qquad = P \left(\sup_{0 \le \lambda \le 1} \left(B_H(\lambda) - \lambda B_H(1) + \frac{c_W \int_{\mathbb{R}} f^2(x) dx}{|\int_{\mathbb{R}} J_1(x) dF(x)|} \phi_\tau(\lambda) \right) \ge q_\alpha \right).$$

Thus, if we want the two tests to have identical power, we have to choose c_C and c_W in such a way that

$$\frac{c_C}{|a_1|} \phi_\tau(\lambda) = \frac{c_W \int_{\mathbb{R}} f^2(x) dx}{|\int_{\mathbb{R}} J_1(x) dF(x)|} \phi_\tau(\lambda),$$

which again yields by (49) and (50),

$$(55) \qquad \frac{h_C(n)}{h_W(n)} = \frac{c_C}{c_W} = \frac{|a_1| \int_{\mathbb{R}} f^2(x) dx}{|\int_{\mathbb{R}} J_1(x) dF(x)|}.$$

This quantity gives the ratio of the height of a level shift that can be detected by a CUSUM test over the height that can be detected by a Wilcoxon change-point test, when both tests are assumed to have the same level α, the same power β and the shifts are taking place at the same time $[n\tau]$. In addition, we assume that the tests are based on the same number of observations n, which is supposed to be large.

Example 4.1. In the case of Gaussian data, i.e. when $G(\xi) = \xi$, we have $m = 1$, $a_1 = E(\xi_1 H_1(\xi_1)) = E(\xi_1^2) = 1$, $\int_{\mathbb{R}} f^2(x)dx = \int_{\mathbb{R}} \frac{1}{2\pi} e^{-x^2} dx = \frac{1}{2\sqrt{\pi}}$ and $\int_{\mathbb{R}} J_1(x)dF(x) = -\frac{1}{2\sqrt{\pi}}$; see Dehling, Rooch and Taqqu (2013), Relation (20). Thus we obtain

$$(56) \qquad \frac{c_C}{c_W} = \frac{1/2\sqrt{\pi}}{1/2\sqrt{\pi}} = 1.$$

Hence, both tests can asymptotically, as $n \to \infty$, detect level shifts of the same height.

The level shifts can be expressed in terms of

$$\psi(t) := P\left(\sup_{0 \leq \lambda \leq 1} (B_H(\lambda) - \lambda B_H(1) + t\,\phi_\tau(\lambda)) \geq q_\alpha\right),$$

viewed as a function of t, for fixed values of τ and α. The function ψ is monotonically increasing. We define the generalized inverse,

$$\psi^-(\beta) := \inf\{t \geq 0 : \psi(t) \geq \beta\}.$$

Thus, we get

$$(57) \qquad P\left(\sup_{0 \leq \lambda \leq 1} \left(B_H(\lambda) - \lambda B_H(1) + \psi^-(\beta)\,\phi_\tau(\lambda)\right) \geq q_\alpha\right) \geq \beta,$$

and, in fact, for given τ, α and β, $\psi^-(\beta)$ is the smallest number having this property.

We can now apply Theorem 2.1 and Theorem 3.1. By comparing (53) and (57), we can detect a level shift of size h at time $[n\tau]$ with a CUSUM test of level α and power β based on n observations, if $h_C(n) \sim \frac{d_n}{n} c_C$, where c_C satisfies $\frac{c_C}{|a_1|} = \psi^-(\beta)$. Hence we obtain that $h_C(n)$ has to satisfy

$$h_C(n) = \frac{d_n}{n}|a_1|\psi^-(\beta) = n^{-D/2}L(n)|a_1|\psi^-(\beta).$$

Similarly, by comparing (54) and (57), we get for the Wilcoxon change-point test that n has to satisfy

$$h_W(n) = n^{-D/2}L(n)\frac{|\int_{\mathbb{R}} J_1(x)dF(x)|}{\int_{\mathbb{R}} f^2(x)dx}\psi^-(\beta).$$

In the following theorem, we compute the asymptotic relative efficiency of the Wilcoxon change point test with respect to the CUSUM test.

Theorem 4.2. Let $(\xi_i)_{i \geq 1}$ be a stationary Gaussian process with mean zero, variance 1 and autocovariance function as in (4). Moreover, let $G : \mathbb{R} \to \mathbb{R}$ be a measurable function satisfying $E(G^2(\xi_1)) < \infty$, and such that $G(\xi_1)$ has a distribution function $F(x)$ with bounded density $f(x)$. Assume that the Hermite rank of $G(\xi_1)$ as well as the Hermite rank of the class of functions $(1_{\{G(\xi_1) \leq x\}} - F(x))$, $x \in \mathbb{R}$ are equal to 1. Moreover assume that $0 < D < 1$. Then

$$(58) \qquad ARE(T_W, T_C) = \left(\frac{|a_1|\int_{\mathbb{R}} f^2(x)dx}{|\int_{\mathbb{R}} J_1(x)dF(x)|}\right)^{2/D},$$

where T_C and T_W denote the CUSUM test and the Wilcoxon change-point test, respectively.

Proof. For abbreviation, we define

$$(59) \qquad b = \left(\frac{|\int_{\mathbb{R}} J_1(x)dF(x)|}{|a_1|\int_{\mathbb{R}} f^2(x)dx}\right)^{2/D}.$$

We will show that the Wilcoxon change-point test based on $b\,n$ observations has asymptotically the same power as the CUSUM test based on n observations. We will consider the local alternative

$$
(60) \qquad A_n^C = A_{\tau,h_n^C} = A_{\tau,c\frac{d_n}{n}}(n)
$$

for the CUSUM test, and the local alternative

$$
(61) \qquad A_n^W = A_{\tau,h_n^W} = A_{\tau,c\frac{b}{n}d_{n/b}}(n)
$$

for the Wilcoxon change-point test. Note that under A_{bn}^W the level shift is the same as under A_n^C. Further observe that, by (11),

$$
\begin{aligned}
h_n^W \sim c\,\frac{b}{n}d_{n/b} &= c\,\frac{b}{n}\kappa_1^{1/2}(n/b)^{1-D/2}L^{1/2}(n/b) \\
&= c\,\frac{b}{n}\kappa_1^{1/2}n^{1-D/2}L^{1/2}(n)b^{D/2-1}\frac{L^{1/2}(n/b)}{L^{1/2}(n)} \\
&= c\,\frac{d_n}{n}b^{D/2}\left(\frac{L(n/b)}{L(n)}\right)^{1/2} \\
(62) &\sim c\,\frac{d_n}{n}b^{D/2},
\end{aligned}
$$

where we have used the fact that $L(n)$ is a slowly varying function.

For the CUSUM test, we can apply Corollary 2.3 and we obtain under the local alternative A_n^C, that

$$
\frac{1}{|a_1|}D_n \xrightarrow{\mathcal{D}} \sup_{0\le\lambda\le1}\left\{(B_H(\lambda)-\lambda B_H(1))+\frac{1}{|a_1|}c\phi_\tau(\lambda)\right\}.
$$

For the Wilcoxon change-point test, we apply Corollary 3.3 with c replaced by $c\,b^{D/2}$, in view of (62). We thus obtain under the local alternative A_n^W,

$$
\begin{aligned}
\frac{1}{|\int_{\mathbb{R}}J_1(x)dF(x)|}W_n \;\xrightarrow{\mathcal{D}}\; &\sup_{0\le\lambda\le1}\left\{(B_H(\lambda)-\lambda B_H(1))+\frac{\int_{\mathbb{R}}f^2(x)dx}{|\int_{\mathbb{R}}J_1(x)dF(x)|}c\,b^{D/2}\phi_\tau(\lambda)\right\} \\
=\; &\sup_{0\le\lambda\le1}\left\{(B_H(\lambda)-\lambda B_H(1))+\frac{1}{|a_1|}c\,\phi_\tau(\lambda)\right\},
\end{aligned}
$$

by (62). Let q_α denote the upper α-quantile of the distribution of $\sup_{0\le\lambda\le1}\{B_H(\lambda)-\lambda B_H(1)\}$; then the test that rejects the null hypothesis when $\frac{1}{|a_1|}D_n \ge q_\alpha$ or when $\frac{1}{|\int_{\mathbb{R}}J_1(x)dF(x)|}W_n \ge q_\alpha$, respectively, have asymptotically level α. The power of these tests at the alternatives A_n^C and A_n^W, respectively, converges to

$$
P\left(\sup_{0\le\lambda\le1}\left\{(B_H(\lambda)-\lambda B_H(1))+\frac{1}{|a_1|}c\,\phi_\tau(\lambda)\right\}\ge q_\alpha\right).
$$

Note that this also holds for the power along any other sequence, such as bn. Since the level shift at the alternative A_n^C equals the level shift at the alternative A_{bn}^W, we have shown that the Wilcoxon change-point test requires $b\,n$ observations to yield the same performance as the CUSUM test with n observations. Thus $ARE(T_W,T_C)=1/b$, proving the theorem. $\qquad\square$

5. ARE OF THE WILCOXON CHANGE-POINT TEST AND THE CUSUM TEST FOR IID DATA

We have shown in Example 4.1 that in the case of LRD data, the ARE of the Wilcoxon change-point test and the CUSUM test is 1 for Gaussian data. In this section, we will

compare this surprising result with the case of i.i.d. Gaussian data. We will find that in this case, the ARE is $\frac{3}{\pi}$, i.e. the Wilcoxon change-point test is less efficient than the CUSUM test.

We consider i.i.d. observations $X_1, \ldots, X_n$ with $X_i \sim \mathcal{N}(0,1)$ and the U-statistic

$$U_k = \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j).$$

As kernel we will choose $h_C(x,y) = y - x$ and $h_W(x,y) = I_{\{x \leq y\}} - \frac{1}{2}$, in other words we consider

$$U_k^{(C)} = \sum_{i=1}^{k} \sum_{j=k+1}^{n} h_C(X_i, X_j) = \sum_{i=1}^{k} \sum_{j=k+1}^{n} (X_j - X_i),$$

$$U_k^{(W)} = \sum_{i=1}^{k} \sum_{j=k+1}^{n} h_W(X_i, X_j) = \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(I_{\{X_i \leq X_j\}} - \frac{1}{2} \right).$$

Both kernels h_C, h_W are antisymmetric, i.e. they satisfy $h(x,y) = -h(y,x)$, so in order to determine the limit behaviour of $U_k^{(C)}$ and $U_k^{(W)}$, we can apply the limit theorems of Csörgő and Horváth (1988).

Theorem 5.1. *Let $X_1, \ldots, X_n$ be i.i.d. random variables with $X_i \sim \mathcal{N}(0,1)$. Under the null hypothesis of no change in the mean, one has*

$$(63) \qquad \sup_{0 \leq \lambda \leq 1} \left| \frac{1}{n^{3/2}} U_{[\lambda n]}^{(C)} - BB_{1,n}(\lambda) \right| = o_P(1)$$

and

$$(64) \qquad \sup_{0 \leq \lambda \leq 1} \left| \frac{1}{n^{3/2} \sqrt{\frac{1}{12}}} U_{[\lambda n]}^{(W)} - BB_{2,n}(\lambda) \right| = o_P(1),$$

where $(BB_{i,n}(\lambda))_{0 \leq \lambda \leq 1}$, $i = 1, 2$, is a sequence of Brownian bridges with mean $E[BB_{i,n}(\lambda)] = 0$ and auto-covariance $E[BB_{i,n}(s) \, BB_{i,n}(t)] = \min(s,t) - st$.

Proof. By Theorem 4.1 of Csörgő and Horváth (1988), it holds under the null hypothesis H that

$$\sup_{0 \leq \lambda \leq 1} \left| \frac{1}{n^{3/2} \sigma} U_{[\lambda n]} - BB_n(\lambda) \right| = o_P(1)$$

where $(BB_n(\lambda))_{0 \leq \lambda \leq 1}$ is a sequence of Brownian bridges like $BB_{1,n}$ and $BB_{2,n}$ above and where $\sigma^2 = E[\tilde{h}^2(X_1)]$ with $\tilde{h}(t) = E[h(t, X_1)]$. The kernel h has to fulfill $E[h^2(X_1, X_2)] < \infty$ which is the case for $h_C(x,y) = y - x$ and $h_W(x,y) = I_{\{x \leq y\}} - \frac{1}{2}$ and Gaussian X_i. $\qquad \square$

Theorem 5.2. *Let $X_1, \ldots, X_n$ be i.i.d. random variables with $X_i \sim \mathcal{N}(0,1)$. Under the sequence of alternatives $A_{\tau, h_n}(n)$ and with $h_n = \frac{1}{\sqrt{n}} c$, where c is a constant, one has*

$$(65) \qquad \left(\frac{1}{n^{3/2}} U_{[\lambda n]}^{(C)} \right)_{0 \leq \lambda \leq 1} \to (BB_1(\lambda) + c \phi_\tau(\lambda))_{0 \leq \lambda \leq 1}$$

and

$$(66) \qquad \left(\frac{1}{n^{3/2} \sqrt{\frac{1}{12}}} U_{[\lambda n]}^{(W)} \right)_{0 \leq \lambda \leq 1} \to \left(BB_2(\lambda) + \frac{c}{2\sqrt{\pi} \cdot \sqrt{\frac{1}{12}}} \phi_\tau(\lambda) \right)_{0 \leq \lambda \leq 1}$$

in distribution, where $(BB_i(\lambda))_{0 \leq \lambda \leq 1}$ *is a Brownian bridge,* $i = 1, 2$.

Proof. First, we prove (65). Like for the case of LRD observations, we decompose the statistic, so that we obtain under the sequence of alternatives $A_{\tau, h_n}(n)$

$$\frac{1}{n^{3/2}} U_{[\lambda n]}^{(C)} = \frac{1}{n^{3/2}} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^{n} (\epsilon_j - \epsilon_i) + \frac{1}{n^{3/2}} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^{n} (\mu_j - \mu_i).$$

By Theorem 5.1, the first term on the right-hand side converges to a Brownian bridge $BB(\lambda)$. For the second term we have like in the proof for LRD observations

$$\frac{1}{n^{3/2}} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^{n} (\mu_j - \mu_i) \sim \sqrt{n} h_n \phi_\tau(\lambda),$$

and in order for the right-hand side to converge, we have to choose

$$(67) \qquad\qquad h_n = \frac{1}{\sqrt{n}} c.$$

Now let us prove (66). Again like for LRD observations, we decompose the statistic into one term that converges like under the null hypothesis and one term which converges to a constant. Under the sequence of alternatives $A_{\tau, h_n}(n)$ and for the case $\lambda \leq \tau$, this decomposition is

$$(68) \qquad \frac{1}{n^{3/2}} U_{[\lambda n]}^{(W)} = \frac{1}{n^{3/2}} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^{n} \left(I_{\{\epsilon_i \leq \epsilon_j\}} - \frac{1}{2} \right) + \frac{1}{n^{3/2}} \sum_{i=1}^{[\lambda n]} \sum_{j=[\tau n]+1}^{n} I_{\{\epsilon_j < \epsilon_i \leq \epsilon_j + h_n\}}.$$

The first term converges uniformly to a Brownian Bridge, like under the null hypothesis. We will show that, if the observations $\epsilon_i = G(\xi_i)$ are i.i.d. with c.d.f. F which has two bounded derivatives $F' = f$ and F'', the second term converges uniformly to $c\lambda(1-\tau) \int_{\mathbb{R}} f^2(x)\, dx$, which is $c\phi_\tau(\lambda) \int_{\mathbb{R}} f^2(x)\, dx$ for the case $\lambda \leq \tau$. In the case of standard normally distributed $G(\xi_i)$, i.e. for $F = \Phi$ and $f = \varphi$, this function is $c(2\sqrt{\pi})^{-1}\phi_\tau(\lambda)$. To this end, we consider the following sequence of Hoeffding decompositions for the sequence of kernels $h_n(x, y) = I_{\{y < x \leq y + h_n\}}$:

$$(69) \qquad\qquad h_n(x, y) = \vartheta_n + h_{1,n}(x) + h_{2,n}(y) + h_{3,n}(x, y)$$

Let $X, Y \sim F$ be i.i.d. random variables. Then we define

$$
\begin{aligned}
\vartheta_n := E[h_n(X, Y)] &= P(Y \leq X \leq Y + h_n) \\
&= \int_{\mathbb{R}} \left(\int_y^{y+h_n} f(x)\, dx \right) f(y)\, dy \\
&= \int_{\mathbb{R}} (F(y + h_n) - F(y)) f(y)\, dy \\
&= h_n \int_{\mathbb{R}} \frac{F(y + h_n) - F(y)}{h_n} f(y)\, dy \\
(70) \qquad &\sim h_n \int_{\mathbb{R}} f^2(y)\, dy,
\end{aligned}
$$

where in the last step we have used that $(F(y + h_n) - F(y))/h_n \to f(y)$ and the dominated convergence theorem. Moreover, we set

$$
\begin{aligned}
h_{1,n}(x) := E[h_n(x, Y)] - \vartheta_n &= E[I_{\{Y < x \leq Y + h_n\}}] - \vartheta_n \\
(71) \qquad\qquad &= F(x) - F(x - h_n) - \vartheta_n
\end{aligned}
$$

and

$$h_{2,n}(y) := E[h_n(X, y)] - \vartheta_n \quad = \quad E[I_{\{y < X \le y + h_n\}}] - \vartheta_n$$
$$= \quad F(y + h_n) - F(y) - \vartheta_n$$

and

$$h_{3,n}(x, y) \quad := \quad h_n(x, y) - h_{1,n}(x) - h_{2,n}(y) - \vartheta_n$$
$$= \quad I_{\{y < x \le y + h_n\}} - F(x) + F(x - h_n) + \vartheta_n - F(y + h_n) + F(y).$$

We will now show that

(72)
$$\sup_{0 \le \lambda \le \tau} \frac{1}{n^{3/2}} \left| \sum_{i=1}^{[\lambda n]} \sum_{j=[\tau n]+1}^{n} (h_{1,n}(\epsilon_i) + h_{2,n}(\epsilon_j) + h_{3,n}(\epsilon_i, \epsilon_j)) \right| \to 0$$

in probability, and from this it follows by the sequence of Hoeffding decompositions (69) that

$$\sup_{0 \le \lambda \le \tau} \frac{1}{n^{3/2}} \left| \sum_{i=1}^{[\lambda n]} \sum_{j=[\tau n]+1}^{n} (h_n(\epsilon_i, \epsilon_j) - \vartheta_n) \right| \to 0$$

i.e. that the second term in (68) converges uniformly to

$$\lim_{n \to \infty} \frac{1}{n^{3/2}} \sum_{i=1}^{[\lambda n]} \sum_{j=[\tau n]+1}^{n} \theta_n = \lim_{n \to \infty} \frac{1}{n^{3/2}} [\lambda n](n - [\tau n])\vartheta_n = \lambda(1 - \tau)c \int_{\mathbb{R}} f^2(x) \, dx$$

by (70) and (67).

We use the triangle inequality and show the uniform convergence to 0 for each of the three terms in (72) seperately. Since the parameter λ occurs only in the floor function value $[\lambda n]$, the supremum is in fact a maximum, and the $h_{1,n}(\epsilon_i)$ are i.i.d. random variables, so we can use Kolmogorov's inequality. We obtain for the first term in (72)

(73)
$$P\left(\sup_{0 \le \lambda \le \tau} \frac{n - [\tau n]}{n^{3/2}} \left| \sum_{i=1}^{[\lambda n]} h_{1,n}(\epsilon_i) \right| > s \right) \le \frac{1}{s^2} \frac{n^2(1 - \tau)^2}{n^3} \sum_{i=1}^{[\tau n]} \text{Var}[h_{1,n}(\epsilon_i)].$$

Now consider an independent copy ϵ of the ϵ_i and the Taylor expansion of F around the value of ϵ,

$$F(\epsilon + h_n) = F(\epsilon) + F'(\epsilon)h_n + \frac{F''(\epsilon + \delta)}{2} h_n^2,$$

where the last term is the Lagrange remainder and thus $\epsilon + \delta$ is between ϵ and $\epsilon + h_n$. Then we obtain

$$\frac{1}{h_n^2} \text{Var}[h_{1,n}(\epsilon)] = \text{Var}\left[\frac{F(\epsilon) - F(\epsilon - h_n)}{h_n} \right]$$
$$= \text{Var}\left[f(\epsilon) + F''(\epsilon + \delta)\frac{h_n}{2} \right]$$
$$= \text{Var}\left[f(\epsilon) \right] + \text{Var}\left[F''(\epsilon + \delta)\frac{h_n}{2} \right]$$
$$+ 2 \left(E[f(\epsilon)F''(\epsilon + \delta)h_n] - E[f(\epsilon)] \, E[F''(\epsilon + \delta)h_n] \right),$$

and since $f = F'$ and F'' are bounded by assumption, we get $\text{Var}[h_{1,n}(\epsilon)] = O(h_n^2)$. Since $h_n \to 0$, the right-hand side of (73) converges to 0 as n increases.

In the same manner, we obtain

$$(74) \qquad P\left(\sup_{0\le\lambda\le\tau}\frac{[\lambda n]}{n^{3/2}}\left|\sum_{j=[\tau n]+1}^{n}h_{2,n}(\epsilon_j)\right|>s\right)\le\frac{1}{s^2}\frac{n^2\lambda^2}{n^3}\sum_{j=1}^{n}\mathrm{Var}[h_{2,n}(\epsilon_j)]\to 0.$$

Finally, we have to show that

$$(75) \qquad \sup_{0\le\lambda\le\tau}\frac{1}{n^{3/2}}\left|\sum_{i=1}^{[\lambda n]}\sum_{j=[\tau n]+1}^{n}h_{3,n}(\epsilon_i,\epsilon_j)\right|\to 0$$

in probability. We set temporarily $l:=[\lambda n]$ and $T:=[\tau n]$ and obtain from Markov's inequality

$$P\left(\max_{0\le l\le T}\frac{1}{n^{3/2}}\left|\sum_{i=1}^{l}\sum_{j=T+1}^{n}h_{3,n}(\epsilon_i,\epsilon_j)\right|>s\right)\le\frac{1}{s^2}E\left[\max_{0\le l\le T}\frac{1}{n^{3/2}}\sum_{i=1}^{l}\sum_{j=T+1}^{n}h_{3,n}(\epsilon_i,\epsilon_j)\right]^2.$$

Now for any collection of random variables $Y_1,\ldots,Y_k$, one has $E[\max\{Y_1^2,\ldots Y_k^2\}]\le\sum_{i=1}^{k}EY_i^2$, so that

$$\frac{1}{s^2}E\left[\max_{0\le l\le T}\frac{1}{n^{3/2}}\sum_{i=1}^{l}\sum_{j=T+1}^{n}h_{3,n}(\epsilon_i,\epsilon_j)\right]^2\le\frac{1}{s^2}\frac{1}{n^3}\sum_{l=1}^{T}E\left[\sum_{i=1}^{l}\sum_{j=T+1}^{n}h_{3,n}(\epsilon_i,\epsilon_j)\right]^2$$

$$=\frac{1}{s^2}\frac{1}{n^3}\sum_{l=1}^{T}\sum_{i=1}^{l}\sum_{j=T+1}^{n}\mathrm{Var}\left[h_{3,n}(\epsilon_i,\epsilon_j)\right],$$

where in the last step we have used that $h_{3,n}(\epsilon_i,\epsilon_j)$ are uncorrelated by Hoeffding's decomposition. Now for two i.i.d. random variables ϵ,η, we have, like above with the Taylor expansion of F:

$$\mathrm{Var}\left[h_{3,n}(\epsilon,\eta)\right]=\mathrm{Var}\left[I_{\{\eta<\epsilon\le\eta+h_n\}}-F(\epsilon)+F(\epsilon-h_n)+\vartheta_n-F(\eta+h_n)+F(\eta)\right]$$

$$=\mathrm{Var}\left[I_{\{\eta<\epsilon\le\eta+h_n\}}-h_n\left(f(\epsilon)+O_P(h_n)\right)+h_n\left(f(\eta)+O_P(h_n)\right)\right]$$

$$=\mathrm{Var}\left[I_{\{\eta<\epsilon\le\eta+h_n\}}\right]+\mathrm{Var}\left[h_n\left(f(\epsilon)+f(\eta)+O_P(h_n)\right)\right]$$

$$\qquad+2\mathrm{Cov}\left[I_{\{\eta<\epsilon\le\eta+h_n\}},h_n\left(f(\epsilon)+f(\eta)+O_P(h_n)\right)\right]$$

$$\le(\vartheta_n-\vartheta_n^2)+h_n^2 O(1)+2\sqrt{(\vartheta_n-\vartheta_n^2)\cdot h_n^2 O(1)}$$

$$=O(h_n),$$

using (70). We have just shown that

$$P\left(\max_{0\le l\le T}\frac{1}{n^{3/2}}\left|\sum_{i=1}^{l}\sum_{j=T+1}^{n}h_{3,n}(\epsilon_i,\epsilon_j)\right|>s\right)\le\frac{1}{s^2}O(h_n),$$

which proves (75). So we have proven (68) for the case $\lambda\le\tau$. The proof for $\lambda>\tau$ is similar. $\square$

Now the stage is set to calculate the ARE of the Wilcoxon test based on $U_{[\lambda n]}^{(W)}$ and the CUSUM test based on $U_{[\lambda n]}^{(C)}$, as defined in the section about the ARE in the LRD case. Let q_α denote the upper α-quantile of the distribution of $\sup_{0\le\lambda\le 1}BB(\lambda)$. By Theorem 5.2, the power of the tests is given respectively by

$$(76) \qquad P\left(\sup_{0\le\lambda\le 1}\left(BB(\lambda)+c_C\phi_\tau(\lambda)\right)\ge q_\alpha\right)$$

and

(77)
$$P\left(\sup_{0\leq\lambda\leq1}\left(BB(\lambda)+c_W\frac{1}{\sigma\cdot2\sqrt{\pi}}\phi_\tau(\lambda)\right)\geq q_\alpha\right)$$

where $\sigma^2 = 1/12$ and we assume that

$$h_W(n) = \frac{c_W}{\sqrt{n}}, \qquad h_C(n) = \frac{c_C}{\sqrt{n}}.$$

Thus if we want both tests to have identical power, we must ensure that $c_C = c_W/(\sigma\cdot2\sqrt{\pi})$, in other words

$$\frac{h_C(n)}{h_W(n)} = \frac{c_C}{c_W} = \frac{1}{\sigma\cdot2\sqrt{\pi}}.$$

Now we define, as in the proof for LRD observations, the probability

$$\psi(t) := P\left(\sup_{0\leq\lambda\leq1}\left(BB(\lambda)+t\,\phi_\tau(\lambda)\right)\geq q_\alpha\right),$$

for whose generalized inverse ψ^- holds

(78)
$$P\left(\sup_{0\leq\lambda\leq1}\left(BB(\lambda)+\psi^-(\beta)\,\phi_\tau(\lambda)\right)\geq q_\alpha\right)\geq\beta.$$

Now, comparing (78) and (76), we conclude that we can detect a level shift of size h at time $[n\tau]$ with the CUSUM test of (asymptotic) level α and power β based on n observations, if $h_C(n) = \frac{c_C}{\sqrt{n}}$ and where c_C satisfies $c_C = \psi^-(\beta)$; hence we obtain that $h_C(n)$ has to satisfy

$$h_C(n) = \frac{1}{\sqrt{n}}\psi^-(\beta).$$

In the same manner, we get for the Wilcoxon test the conditions $h_W(n) = \frac{c_W}{\sqrt{n}}$ and $c_W/(\sigma2\sqrt{\pi}) = \psi^-(\beta)$ and thus

$$h_W(n) = \frac{\sigma2\sqrt{\pi}}{\sqrt{n}}\psi^-(\beta).$$

Solving these two equations for n again and denoting the resulting numbers of observations by n_C and n_W, respectively, we obtain

$$n_C = \left(\frac{1}{h_C}\psi^-(\beta)\right)^2$$

$$n_W = \left(\frac{2\sigma\sqrt{\pi}}{h_W}\psi^-(\beta)\right)^2.$$

To obtain $ARE(T_W, T_C)$, we equate h_W and h_C. We then obtain the following theorem.

Theorem 5.3. *Let $X_1,\ldots,X_n$ be i.i.d. random variables with $X_i \sim \mathcal{N}(0,1)$. Then*

(79)
$$ARE(T_W, T_C) = \lim_{h\to0}\frac{n_C}{n_W} = (2\sigma\sqrt{\pi})^{-2} = \frac{3}{\pi},$$

where T_C, T_W denote the one-sided CUSUM-test, respectively the one-sided Wilcoxon test, for the test problem (H, A_{τ,h_n}).

6. Simulation Results

We have proven that for Gaussian data, the CUSUM test and the Wilcoxon change-point test show asymptotically the same performance, i.e. that their ARE is 1. For Pareto(3,1) distributed data, we obtain, using (58) and numerical integration, an ARE of approximately $(2.68)^{2/D}$. Now we will illustrate these findings by a simulation study.

6.1. **Gaussian data.** We consider realizations $\xi_1, \ldots, \xi_n$ of a fGn process with Hurst parameter $H = 0.7$ ($D = 0.6$), using the `fArma` package in `R`, and create observations

$$X_i = \begin{cases} G(\xi_i) & \text{for } i = 1, \ldots, [n\lambda] \\ G(\xi_i) + h & \text{for } i = [n\lambda] + 1, \ldots, n \end{cases},$$

by applying a transformation G which is (with respect to the standard normal measure) normalized and square-integrable: $E[G(\xi)] = 0$, $E[G^2(\xi)] = 1$ for $\xi \sim \mathcal{N}(0,1)$. As a first step, we choose $G(t) = t$ in order to obtain Gaussian observations $X_1, \ldots, X_n$ (later we will choose a function G such that we obtain Pareto distributed data). In other words, we consider data which follow the local alternative

$$A_{\lambda,h} : \begin{cases} \mu = E[X_i] = 0 & \text{for } i = 1, \ldots, [n\tau] \\ \mu = E[X_i] = h & \text{for } i = [n\tau] + 1, \ldots, n, \end{cases}$$

as in (8). In contrast to the simulations by Dehling, Rooch and Taqqu (2013), we choose a sample size $n = 2,000$ instead of $n = 500$. We let both the break point $k = [\tau n]$ and the level shift $h := \mu_{k+1} - \mu_k$ vary; specifically, we choose $k = 100, 200, 600, 1000$ (which corresponds to $\tau = 0.05, 0.1, 0.3, 0, 5$) and we let $h = 0.5, 1, 2$. For each of these situations, we will compare the power of the CUSUM test and the power of the Wilcoxon change-point test in the test problem $(H, A_{\lambda,h})$: We have repeated each simulation $10,000$ times and counted, how often the respective test (correctly) rejected the null hypothesis.

Since our theoretical considerations yield an ARE of 1, we expect that both tests detect jumps equally well – that means that both tests, set on the same level, detect jumps of the same height and at the same position in the same number of observations with the same relative frequency. And indeed, we can clearly see in Table 1 that the power of both tests approximately coincides at many points; differences can be spotted only when the break occurs early in the data.

	relative jump position τ					relative jump position τ			
	0.05	0.1	0.3	0.5		0.05	0.1	0.3	0.5
h=0.5	0.074	0.153	0.767	0.874	h=0.5	0.072	0.143	0.765	0.876
h=1	0.153	0.694	1.000	1.000	h=1	0.128	0.602	1.000	1.000
h=2	0.828	1.000	1.000	1.000	h=2	0.321	1.000	1.000	1.000

Table 1. Power of the CUSUM test (left) and of the Wilcoxon change-point test (right), for $n = 2000$ observations of fGn with LRD parameter $H = 0.7$, different break points $[\tau n]$ and different level shifts h. Both tests have asymptotically level 5%. The calculations are based on 10,000 simulation runs.

6.2. **Heavy tailed data.** We consider again realizations $\xi_1, \ldots, \xi_n$ of a fGn process with Hurst parameter $H = 0.7$ ($D = 0.6$) and create observations

$$X_i = \begin{cases} G(\xi_i) & \text{for } i = 1, \ldots, [n\lambda] \\ G(\xi_i) + h & \text{for } i = [n\lambda] + 1, \ldots, n \end{cases},$$

by applying the transformation

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

In this case, the first Hermite coefficient of G, obtained by numerical integration, equals $a_1 \approx -0.6784$. This transformation G produces observations $X_i = G(\xi_i)$ which follow a standardized $\text{Pareto}(3, 1)$ distribution with mean zero and variance 1. The probability density function of X_i is given by

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

To the second sample of observations, $X_{[\tau n]+1}, \ldots, X_n$, we again add a constant h, but this time we choose

(80) $$h = h_n = c\frac{d_n}{n} = cn^{-D/2}$$

as in (18). We let the break point $k = [\tau n]$ vary; here, we choose $\tau = 0.05, 0.1, 0.3, 0.5$. We let also the sample size vary; we will give details below. To these data, we have applied the CUSUM test and the Wilcoxon change-point test, and under $10,000$ simulation runs we counted how often the respective test (rightly) rejected the null hypothesis.

Now our theoretical considerations, see (58), predict for this situation

$$ARE = \lim_{n \to \infty} \frac{n_C}{n_W} = \left(\frac{|a_1| \int_{\mathbb{R}} f^2(x)\, dx}{|\int_{\mathbb{R}} J_1(x) f(x)\, dx|} \right)^{2/D} \approx (2.67754)^{2/0.6} \approx 26.655.$$

This means that the CUSUM test needs approximately 26.66 times as many observations as the Wilcoxon test in order to detect the same jump on the same level with the same probability. In order to find this behaviour, we have analysed the power of the Wilcoxon test for $n_W = 10, 50, 100, 200$ observations and the power of the CUSUM test for $n_C = 266, 1332, 2666, 5330$ observations.

In order to be able to compare the two tests, we need to have identical level shifts when applying the Wilcoxon test to a sample of size n_W and the CUSUM test to a sample of size $n_C = 26.655\, n_W$. This can be achieved by choosing the constants c in (80) accordingly, namely taking $c_C = 2.67754\, c_W$. In this way, we obtain

$$h_{n_C}^C = c_C n_C^{-D/2} = 2.67754 c_W (26.655 n_W)^{-D/2} = c_W n_W^{-D/2} = h_{n_W}^W.$$

We ran simulations for two different choices of c_W, namely $c_W = 1$ and $c_W = 2$; see Table 3 and Table 4 for the results.

Here, we have to face a problem which was already encountered by Dehling, Rooch and Taqqu (2013). For the heavy-tailed Pareto data, the convergence of the CUSUM test statistic towards its limit is so slow that the asymptotic quantiles of the limit distribution are not appropriate as critical values to define the domain of rejection of the test: In finite sample situations, the observed level of the test is not 5% – as it should be when using the 5%-quantile of the asymptotic limit distribution. In order to remedy this, we used as critical value for the test, the finite sample 5% quantiles of the distribution of the CUSUM test

statistic under the null hypothesis, using a Monte Carlo simulation; see Table 6 in Dehling, Rooch and Taqqu (2013). Here, we have performed the same steps, but for sample sizes $n = n_C = 266, 1332, 2666, 5330$. The results are given in Table 2. Note that this problem does not arise when using the Wilcoxon change-point test, since the Wilcoxon test is distribution free under the null hypothesis.

n	266	1332	2666	5330	∞
$q_{\text{emp},0.05}$	0.73	0.66	0.64	0.63	0.59

TABLE 2. 5%-quantiles of the finite sample distribution of the CUSUM test under the null hypothesis for Pareto(3,1)-transformed fGn with LRD parameter $H = 0.7$ and different sample sizes $n = n_C$.

The simulation results are shown in Table 3 (for $c_W = 1$) and Table 4 (for $c_W = 2$). Indeed, for a fixed jump position τ, the power of the CUSUM test (for $n = n_C = 266, 1332, 2666, 5330$ observations) and of the Wilcoxon test (for $n = n_W = 10, 50, 100, 200$ observations) coincide. They are not fully equal, but we conjecture this is due to the small sample size which conflicts with the asymptotic character of our results. But it becomes clear: The CUSUM test needs quite a number of observations more to detect the same jump on the same level with the same probability – as predicted by our calculation around 25 times as many.

n	h	relative jump position τ				n	h	relative jump position τ			
		0.05	0.1	0.3	0.5			0.05	0.1	0.3	0.5
266	0.50	0.049	0.049	0.066	0.088	10	0.50	0.036	0.025	0.033	0.079
1332	0.31	0.050	0.052	0.083	0.110	50	0.31	0.049	0.051	0.093	0.120
2666	0.25	0.052	0.055	0.092	0.127	100	0.25	0.050	0.053	0.102	0.134
5330	0.20	0.051	0.054	0.099	0.130	200	0.20	0.051	0.055	0.103	0.134

TABLE 3. Power of the CUSUM test (left) and of the Wilcoxon change-point test (right), at different break points $[\tau n]$, different sample sizes n, and different jump heights h, for Pareto(3,1) distributed data. Both tests have asymptotically level 5% (CUSUM test is performed with empirical quantiles). The calculations are based on 10,000 simulation runs.

n	h	relative jump position τ				n	h	relative jump position τ			
		0.05	0.1	0.3	0.5			0.05	0.1	0.3	0.5
266	1.00	0.049	0.054	0.162	0.259	10	1.00	0.033	0.024	0.039	0.197
1332	0.62	0.052	0.062	0.236	0.345	50	0.62	0.049	0.055	0.199	0.283
2666	0.50	0.055	0.069	0.272	0.390	100	0.50	0.051	0.063	0.225	0.316
5330	0.41	0.054	0.074	0.287	0.402	200	0.41	0.054	0.066	0.242	0.338

TABLE 4. Power of the CUSUM test (left) and of the Wilcoxon change-point test (right) at different break points $[\tau n]$, different sample sizes n, and different jump heights h, for Pareto(3,1) distributed data. Both tests have asymptotically level 5% (CUSUM test is performed with empirical quantiles). The calculations are based on 10,000 simulation runs.

References

[1] Miklós Csörgő and Lajos Horváth (1988): Invariance Principle for Changepoint Problems, *Journal of Multivariate Analysis* **27**, 151–168.

[2] Miklós Csörgő and Lajos Horváth (1997): *Limit Theorems in Change-Point Analysis.* J. Wiley & Sons, Chichester.

[3] Herold Dehling and Murad S. Taqqu (1989): The Empirical Process of Some Long-Range Dependent Sequences with an Application to U-Statistics. *The Annals of Statistics* **17**, 1767–1783.

[4] Herold Dehling, Aeneas Rooch and Murad S. Taqqu (2013): Nonparametric Change-Point Tests for Long-Range Dependent Data. *Scandinavian Journal of Statistics* **40**, 153–173.

[5] Vladas Pipiras and Murad S. Taqqu (2010): Regularization and Integral Representations of Hermite Processes. *Statistics and Probability Letters* **80**, 697–705.

[6] Galen R. Shorack and Jon A. Wellner (1986): *Empirical Processes with Applications to Statistics.* John Wiley & Sons, New York.

[7] Murad S. Taqqu (1979): Convergence of Integrated Processes of Arbitrary Hermite Rank. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50**, 53–83.

Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany
E-mail address: herold.dehling@rub.de

Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany
E-mail address: aeneas.rooch@rub.de

Department of Mathematics, 111 Cummington St., Boston University, Boston MA 02215
E-mail address: murad@bu.edu