SFB 823

Robust discrimination between long-range dependence and a change in mean

Carina Gerstenberger

Discussion Paper

Nr. 10/2018

Robust Discrimination between Long-Range Dependence and a Change in Mean

Carina Gerstenberger*

In this paper we introduce a robust to outliers Wilcoxon change-point testing procedure, for distinguishing between short-range dependent time series with a change in mean at unknown time and stationary long-range dependent time series. We establish the asymptotic distribution of the test statistic under the null hypothesis for L_1 near epoch dependent processes and show its consistency under the alternative. The Wilcoxon-type testing procedure similarly as the CUSUM-type testing procedure of Berkes, Horváth, Kokoszka and Shao (2006), requires estimation of the location of a possible change-point, and then using pre- and post-break subsamples to discriminate between short and long-range dependence. A simulation study examines the empirical size and power of the Wilcoxon-type testing procedure in standard cases and with disturbances by outliers. It shows that in standard cases the Wilcoxon-type testing procedure behaves equally well as the CUSUM-type testing procedure but outperforms it in presence of outliers.

KEYWORDS: Wilcoxon change-point test statistic; change-point; near epoch dependence; long-range dependence

1 Introduction

Since the pioneering work of Hurst (1951), Mandelbrot and Van Ness (1968) and Mandelbrot and Wallis (1968), the phenomenon of long-range dependence or Hust effect has been observed in many data sets, e.g. in hydrology, geophysics and economics. A lively debate also rages over the observed Hurst effect is due to long-range dependence or nonstationarity. Bhattacharya *et al.* (1983) showed that the Hurst effect detected by R/S statistics can be explained not only by long-range dependence, but by presence of a deterministic trend in short-range dependent data. Giraitis *et al.* (2001) showed that some modified R/S statistics reject the hypothesis of short-range dependence for long-range dependence but also for short-range dependent data in presence of a trend or change-points. The phenomenon of spurious long-range dependence has also been discussed in many other papers, see e.g. Granger and Hyung (2004).
A first attempt for distinguishing between long-range dependence and short-range dependence with a monotonic trend was made by Künsch (1986), who showed that the

Date: April 4, 2018.
*Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany

periodogram in these two cases behaves differently. A test allowing to distinguish between a stationary long-range dependent process and short-range dependent process with a change in mean was introduced by Berkes *et al.* (2006) and is based on the CUSUM statistic

$$C_{m,n}(k) = \sum_{i=m}^{k} X_i - \frac{k-m+1}{n} \sum_{i=1}^{n} X_i, \qquad m \leq k \leq n. \tag{1}$$

It is well known that the CUSUM statistic is sensitive to outliers since it sums up the observations. In this paper we introduce a new robust to outliers testing procedure, which is based on the Wilcoxon change-point test statistic

$$W_{m,n}(k) = \sum_{i=m}^{k} \sum_{j=k+1}^{n} (1_{\{X_i \leq X_j\}} - 1/2), \qquad m \leq k \leq n. \tag{2}$$

Dehling *et al.* (2013b, 2015) used this test statistic for testing for changes in the mean of long-range dependent and short-range dependent processes respectively. In both papers the simulation studies point out that the Wilcoxon test statistic (2) is more robust to outliers than the CUSUM statistic (1). Recently, Gerstenberger (2018) showed that Wilcoxon-type change-point location estimator for a change in mean of short-range dependent data based on test statistic (2) is also robust against outliers.
The new Wilcoxon-type testing procedure suggested in this paper is based on the idea of Berkes *et al.* (2006). Firstly, given a sample $X_1, \ldots, X_n$, one estimates the location $\hat{k}$ of a possible change in mean. Then the test statistic is defined as the maximum of the Wilcoxon change-point statistic (2) applied to the subsamples $X_1, \ldots, X_{\hat{k}}$ and $X_{\hat{k}+1}, \ldots, X_n$.

Wilcoxon-type testing procedure

Assuming that sample $X_1, \ldots, X_n$ is given, we want to test the hypothesis

$\boldsymbol{H_0}$: $X_i = Y_i + \mu_i$, $i = 1, \ldots, n$ is generated by a stationary zero mean short-range dependent process (Y_j) and has a change in mean $\mu_1 = \ldots = \mu_{k^*} \neq \mu_{k^*+1} = \ldots = \mu_n$ at unknown time k^*,

against the alternative

$\boldsymbol{H_1}$: $X_1, \ldots, X_n$ is a sample from a stationary long-range dependent process.

To construct the test statistic, first, we estimate the location k^* of a change-point by a Wilcoxon-type change-point location estimator

$$\hat{k} = \min\Big\{k : \max_{1 \leq l < n} \big|W_{1,n}(l)\big| = \big|W_{1,n}(k)\big|\Big\}, \tag{3}$$

which is defined as the smallest k for which $|W_{1,n}(k)|$ attains its maximum.

Next we divide the sample $X_1, \ldots, X_n$ into subsamples $X_1, \ldots, X_{\hat{k}}$ and $X_{\hat{k}+1}, \ldots, X_n$, and set

$$T(X_1, \ldots, X_n) = n^{-3/2} \max_{1 \le k \le n} \big|W_{1,n}(k)\big|.$$

Then we compute $T(X_1, \ldots, X_{\hat{k}})$ and $T(X_{\hat{k}+1}, \ldots, X_n)$, and denote

$$T_{n,1} := T(X_1, \ldots, X_{\hat{k}}) = \hat{k}^{-3/2} \max_{1 \le k \le \hat{k}} \big|W_{1,\hat{k}}(k)\big|, \tag{4}$$

$$T_{n,2} := T(X_{\hat{k}+1}, \ldots, X_n) = (n - \hat{k})^{-3/2} \max_{\hat{k} < k \le n} \big|W_{\hat{k}+1,n}(k)\big|. \tag{5}$$

Finally, we define the test statistic

$$M_n = \max\{T_{n,1}, T_{n,2}\}. \tag{6}$$

We show that $T(X_1, \ldots, X_n)$ allows to discriminate whether the sample has been generated by a short or long-range dependent stationary process. Hence, if we split the sample at time $\hat{k}$, which is close to the true change-point k^*, since $\hat{k}/k^* \to_p 1$ asymptotically we can assume that $X_1, \ldots, X_{\hat{k}}$ and $X_{\hat{k}+1}, \ldots, X_n$ are samples from a stationary sequence with a constant mean, see Lemma 4.1 in Section 4. Subsequently, M_n can be used to test if the samples $X_1, \ldots, X_{\hat{k}}$ and $X_{\hat{k}+1}, \ldots, X_n$ have been generated by a short-range or long-range dependent stationary process.

The outline of the paper is as follows. Section 2 specifies assumptions allowing to establish asymptotic distribution of M_n under H_0 and consistency under H_1. Section 3 compares finite sample performance of the Wilcoxon-type and the CUSUM-type testing procedure. All proofs are given in Section 4.

2 Definitions, assumptions and main results

In this section we present main assumptions, definitions and main results.
Throughout the paper, C denotes a generic non-negative constant, which may vary from time to time. The notation $a_n \sim b_n$ means that sequences a_n and b_n of real numbers have property $a_n/b_n \to c$, as $n \to \infty$, where $c \neq 0$. $\xrightarrow{d}$ and $\to_p$ stand for convergence in distribution and probability, respectively. By $\overset{d}{=}$ we denote equality in distribution. $\|g\|_\infty = \sup_x |g(x)|$ denotes the supremum norm of a function g.

Null hypothesis: short-range dependence with a change in mean

Under the null hypothesis we assume the random variables $X_1, \ldots, X_n$ follow the change-point model

$$X_i = \begin{cases} Y_i + \mu & , 1 \le i \le k^* \\ Y_i + \mu + \Delta_n & , k^* < i \le n, \end{cases} \tag{7}$$

where k^* denotes the unknown location of the change-point in the mean and (Y_j) is a zero-mean stationary short-range dependent process.

To cover a wide range of processes, we assume that the underlying process (Y_j) can be written as $Y_j = f(Z_j, Z_{j-1}, Z_{j-2}, \ldots)$, $j \in \mathbb{Z}$, where $f : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is a measurable function, and (Z_j) is an absolutely regular (weakly dependent) process.

Definition 2.1. *A stationary process (Z_j) is called absolutely regular (or β-mixing) if*

$$\beta_k = \sup_{n\geq 1} \mathrm{E} \sup_{A\in\mathcal{G}_1^n} \left|\mathrm{P}\left(A|\mathcal{G}_{n+k}^{\infty}\right) - \mathrm{P}\left(A\right)\right| \to 0, \tag{8}$$

as $k \to \infty$, where $\mathcal{G}_k^m$ is the σ-field generated by random variables $Z_k, \ldots, Z_m$, $k < m$.

Absolute regularity or β-mixing implies the weaker property of α-mixing, see e.g. Bradley (2007).
In addition, we will assume that (Y_j) satisfies near epoch dependence condition, i.e. Y_j depends on the near past of (Z_j).

Definition 2.2. *A stationary process (Y_j) is L_1 near epoch dependent (L_1 NED) on some stationary process (Z_j) with approximation constants a_k, $k \geq 0$, if*

$$\mathrm{E}\,|Y_1 - \mathrm{E}(Y_1|\mathcal{G}_{-k}^k)| \leq a_k, \qquad k = 0, 1, 2, \ldots \tag{9}$$

where $\mathcal{G}_{-k}^k$ is the σ-field generated by random variables $Z_{-k}, \ldots, Z_k$ and $a_k \to 0$ as $k \to \infty$.

Notice that a linear process or AR process might not be absolutely regular, but it would be L_1 near epoch dependent; see Example 2.1 in Gerstenberger (2018) for linear processes and Hansen (1991) for GARCH(1,1) processes. More examples of L_1 NED processes can be found in Borovkova *et al.* (2001), who also discuss more general L_r NED processes, $r \geq 1$.

We need further additional assumptions on the distribution function F of Y_1, the mixing coefficients β_k in (8) and a_k in (9).

Assumption 1. *The process (Y_j) in (7) is L_1 NED on some absolutely regular process (Z_j) with mixing coefficients β_k and approximation constants a_k such that*

$$\sum_{k=1}^{\infty} k^2(\beta_k + \sqrt{a_k}) < \infty. \tag{10}$$

Moreover, Y_1 has a continuous distribution function F with bounded second derivative, and variables $Y_1 - Y_k$, $k \geq 1$ satisfy

$$\mathrm{P}(x \leq Y_1 - Y_k \leq y) \leq C|y - x|, \tag{11}$$

for all $x \leq y$, where C does not depend on k and x, y.

We suppose that both, the unknown change-point k^* and the magnitude of change Δ_n in (7), depend on the sample size n.

Assumption 2. *a) The change-point $k^* = [n\theta]$, where $0 < \theta < 1$ is fixed, is proportional to the sample size n.*

b) The magnitude of change Δ_n in (7) depends on n, and is such that

$$\Delta_n \to 0, \qquad n\Delta_n^2 \to \infty, \qquad n \to \infty.$$

An important step of our testing procedure is the estimation of the location k^* of the change-point in mean. Gerstenberger (2018) showed that under Assumptions 1 and 2 the Wilcoxon-type change-point location estimator $\hat{k}$ in (3) is consistent,

$$\Delta_n^2 \big|\hat{k} - k^*\big| = O_P(1), \qquad \text{as } n \to \infty. \tag{12}$$

Alternative: long-range dependence

Under alternative H_1, the sample $X_1, \ldots, X_n$ is generated by a stationary long-range dependent process:

$$X_i = G(\xi_i) + \mu, \qquad i = 1, \ldots, n, \tag{13}$$

where μ is the unknown mean and (ξ_j) is a stationary long memory Gaussian process with $\mathrm{E}(\xi_1) = 0$, $\mathrm{Var}(\xi_1) = 1$ and (non-summable) auto-covariances $\gamma_k = \mathrm{Cov}(\xi_1, \xi_{1+k}) \sim k^{2d-1}c_0$, where $c_0 > 0$ and $d \in (0, 1/2)$. Furthermore, we assume that $G : \mathbb{R} \to \mathbb{R}$ is a measurable, strictly monotone function such that $\mathrm{E}(G(\xi_1)) = 0$.

Main results

The following theorem derives the limit distribution of the test procedure under the null hypothesis H_0. Below, $B(t) = W(t) - tW(1)$ denotes a standard Brownian bridge, where $W(t)$ is a standard Brownian motion.

Theorem 2.1. *Let (X_j) follow the model in (7). Then, under Assumptions 1 and 2,*

$$M_n = \max\{T_{n,1}, T_{n,2}\} \xrightarrow{d} \sigma \max\left\{ \sup_{0\le t\le 1} \big|B^{(1)}(t)\big|, \sup_{0\le t\le 1} \big|B^{(2)}(t)\big| \right\} =: \sigma Z \tag{14}$$

where $B^{(1)}$ and $B^{(2)}$ are two independent Brownian bridges,

$$\sigma^2 = \sum_{k=-\infty}^{\infty} \mathrm{Cov}\left(F(Y_0), F(Y_k)\right), \tag{15}$$

and F denotes the distribution function of Y_1.

Since the limit distribution of M_n depends on the long-run variance σ^2, to calculate the critical values for the test, we need to estimate the long-run variance; see Section 3.

We will compare performance of our test with the CUSUM-type test by Berkes *et al.* (2006) defined as

$$\tilde{M}_{C,n} = \max\{\tilde{T}_C(X_1, \ldots, X_{\tilde{k}_C}), \tilde{T}_C(X_{\tilde{k}_C+1}, \ldots, X_n)\}, \tag{16}$$

where

$$\tilde{T}_C(X_1,\ldots,X_n) = (\hat{s}_n\sqrt{n})^{-1}\max_{1\le k\le n}\left|C_{1,n}(k)\right|,$$

is based on the CUSUM statistic $C_{1,n}(k)$ in (1). $\tilde{k}_C = \min\Big\{k : \max_{1\le l\le n}\left|C_{1,n}(l)\right| = \left|C_{1,n}(k)\right|\Big\}$ is a CUSUM-type estimator of k^* and $\hat{s}_n^2$ is a long-run variance estimator of $\sigma_c^2 = \sum_{k=-\infty}^{\infty}\mathrm{Cov}\,(Y_0,Y_k)$ given in (21). Berkes *et al.* (2006) showed that under their assumptions under the null hypothesis, $\tilde{M}_{C,n}\xrightarrow{d} Z$.

The next theorem establishes consistency of the test M_n, i.e. that the test will detect long-range dependence with probability tending to 1.

Theorem 2.2. *Let (X_j) be as in (13). Then, as $n\to\infty$,*

$$M_n \to_p \infty.$$

Proofs of Theorem 2.1 and 2.2 are given in Section 4.

3 Simulation Study

In this simulation study we compare the finite sample performance (size and power) of the Wilcoxon-type testing procedure M_n in (6) with the CUSUM-type testing procedure $\tilde{M}_{C,n}$ of Berkes *et al.* (2006), given in (16).

Simulation set up
To calculate the *empirical size* we generate the sample of random variables $X_1,\ldots,X_n$ using the change-point model

$$X_i = \begin{cases} Y_i+\mu & ,1\le i\le k^* \\ Y_i+\mu+\Delta & ,k^*< i\le n, \end{cases} \tag{17}$$

where $Y_i = \rho Y_{i-1}+\epsilon_i$ is an AR(1) process with $\rho = 0.4$ and standard normal innovations ϵ_i. We set $k^* = [n\theta]$, $\theta = 0.25, 0.5, 0.75$ and $\Delta = 0.5, 1, 2$.
To evaluate the *empirical power* of the test we generate a sample $X_1,\ldots,X_n$ of fractional Gaussian noise (fGn)

$$X_i = W_H(i+1) - W_H(i), \tag{18}$$

where $W_H(t)$, $H = d+1/2\in(1/2,1)$ is a fractional Brownian motion, see e.g. Mandelbrot and Van Ness (1968). The sequence (X_j) is a long-range dependent process: $\mathrm{Cov}(X_1,X_{1+k})\sim k^{2d-1}c_0$ with long-range dependence parameter $d\in(0,1/2)$. We consider $d = 0.1, 0.2, 0.3, 0.4$.
To analyse the robustness of Wilcoxon and CUSUM testing procedures to *outliers*, we replace observations $X_{[0.2n]}, X_{[0.4n]}, X_{[0.6n]}, X_{[0.8n]}$ in the sample $(X_1,\ldots,X_n)$ (under the null hypothesis or alternative) by outliers $50X_{[0.2n]}, 50X_{[0.4n]}, 50X_{[0.6n]}$ and $50X_{[0.8n]}$.

We consider sample sizes $n = 200, 500, 1000, 2000, 5000$. All simulation results are based on $10,000$ replications.

Critical values

To analyse the empirical size and power, we need to know the critical values for the tests M_n and $\tilde{M}_{C,n}$.
By Theorem 2.1, under the null hypothesis,

$$M_n = \max\left\{T_{n,1}, T_{n,2}\right\} \xrightarrow{d} \sigma Z.$$

Hence, if $\hat{\sigma}^2(X_1, \ldots, X_k)$ is a consistent estimator for the long-run variance σ^2 based on the sample $X_1, \ldots, X_k$, then

$$\hat{M}_n = \max\left\{\frac{T_{n,1}}{\hat{\sigma}(X_1, \ldots, X_{\hat{k}})}, \frac{T_{n,2}}{\hat{\sigma}(X_{\hat{k}+1}, \ldots, X_n)}\right\} \xrightarrow{d} Z.$$

The same asymptotics holds for the CUSUM test: $\tilde{M}_{C,n} \xrightarrow{d} Z$, see Corollary 2.1 of Berkes *et al.* (2006). Thus, the critical value c_α for a given significance level α is obtained by solving

$$\mathrm{P}\left(Z > c_\alpha\right) = \alpha. \tag{19}$$

Since $B^{(1)}$ and $B^{(2)}$ are independent Brownian bridges, (19) reduces to

$$\mathrm{P}\left(\sup_{0\le t\le 1}\left|B^{(1)}(t)\right| \le c_\alpha\right) = (1-\alpha)^{1/2}, \tag{20}$$

where $\sup_{0\le t\le 1}\left|B^{(1)}(t)\right|$ has the well-known Kolmogorov-Smirnov distribution, and its quantiles can be found in statistical tables. For $\alpha = 5\%$ (20) implies $c_{5\%} = 1.478$.

Estimation of long-run variance

The selection of a long-run variance estimate $\hat{\sigma}$ in $\hat{M}_n$ has a strong impact on the size and power properties of the tests in finite samples.

To estimate the long-run variance $\sigma_c^2 = \sum_{k=-\infty}^{\infty} \mathrm{Cov}\left(Y_0, Y_k\right)$ in $\tilde{M}_{C,n}$ in (16), Berkes *et al.* (2006) suggested to use the Bartlett estimator

$$\hat{s}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 + 2\sum_{j=1}^{q(n)}\left(1 - \frac{j}{q+1}\right)\frac{1}{n}\sum_{i=1}^{n-j}\left(X_i - \bar{X}_n\right)\left(X_{i+j} - \bar{X}_n\right), \tag{21}$$

where $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$, with the bandwidth $q\left(n\right) = C\log_{10}\left(n\right)$. Table 1 reports the empirical size (for $\theta = 0.5$, $\Delta = 1$) and power (for $d = 0.4$) in % at significance level 5% of $\tilde{M}_{C,n}$ test, with $\hat{s}_n^2$ as in (21) computed with bandwidth $15\log_{10}\left(n\right)$. It shows that $\tilde{M}_{C,n}$ with Bartlett estimator $\hat{s}_n^2$ is too conservative and has low power against the

n =	500	1000	2000	5000
emp. size	0.05	0.87	2.48	3.79
power	0.30	7.62	27.44	60.51

Table 1: Empirical size and power of $\tilde{M}_{C,n}$ test using the Bartlett estimator.

alternative, which has also been pointed out by Baek and Pipiras (2012) and Preuß *et al.* (2017).
In our simulation study to improve the performance of $\tilde{M}_{C,n}$ test we proceed as follows. To estimate σ_C^2, instead of $\hat{s}_n^2$, we use the non-overlapping subsampling estimator of σ_C^2 by Carlstein (1986), with block length l_n,

$$\hat{\sigma}_C^2 = \frac{1}{[n/l_n]} \sum_{i=1}^{[n/l_n]} \frac{1}{l_n} \Big(\sum_{j=(i-1)l_n+1}^{il_n} X_j - \frac{l_n}{n} \sum_{j=1}^{n} X_j \Big)^2, \tag{22}$$

which yields better size and power balance for $\tilde{M}_{C,n}$, as seen from Tables 2 and 4. This estimator has also been used by Dehling *et al.* (2015) for a CUSUM-type test for changes in the mean of a short-range dependent process.
In turn, for our test $\hat{M}_n$ to estimate σ we shall use the Carlstein type estimator for long-run variance proposed by Dehling *et al.* (2013a),

$$\hat{\sigma}_W = \frac{1}{[n/l_n]} \sqrt{\frac{\pi}{2}} \sum_{i=1}^{[n/l_n]} \frac{1}{\sqrt{l_n}} \Big| \sum_{j=(i-1)l_n+1}^{il_n} F_n(X_j) - \frac{l_n}{n} \sum_{j=1}^{n} F_n(X_j) \Big|, \tag{23}$$

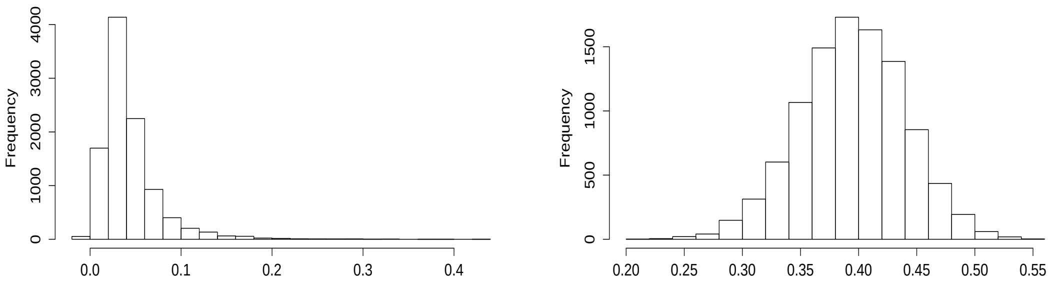
where $F_n(x) = n^{-1} \sum_{i=1}^n 1_{\{X_i \le x\}}$. Note that $\hat{\sigma}_W$ estimates σ, not σ^2.
The Carlstein estimator $\hat{\sigma}_C^2$ as well as the estimator $\hat{\sigma}_W$ (23) are subsampling type estimators and require to choose a suitable block length l_n. The choice of l_n is widely discussed in the literature. For AR(1)-processes Carlstein (1986) suggests to use

$$l_n = \max\big\{ \lceil n^{1/3} (2\rho/(1-\rho^2))^{2/3} \rceil, 1 \big\}, \tag{24}$$

where ρ denotes the autocorrelation coefficient at lag 1. In our simulation study we use this block length with ρ estimated by the sample autocorrelation coefficient $\hat{\rho}$ since it yields good results for the empirical size and power.
In the presence of outliers, we need to robustify further the choice of the block length. Since the sample autocorrelation is highly sensitive to outliers, we use in (24) a robust estimator of ρ proposed by Ma and Genton (2000),

$$\hat{\rho}_Q = \frac{Q_{n-1}^2(u+v) - Q_{n-1}^2(u-v)}{Q_{n-1}^2(u+v) + Q_{n-1}^2(u-v)},$$

where $Q_n = 2.21914\{|X_i - X_j|; i < j\}_{(k)}$, which is the $k = \binom{n}{2}/4$-th order statistic of the $\binom{n}{2}$ interpoint distances, is a robust scale estimator introduced by Rousseeuw and Croux



(a) sample autocorrelation (b) MG-estimator

Figure 1: Histogram of $\hat{\rho}(1)$ and $\hat{\rho}_Q(1)$ based on 10,000 replications. X_i is generated by an AR(1) process with outliers, $\epsilon_i \sim N(0,1)$, $\rho = 0.4$ and $n = 500$.

(1993), $u = (X_1, \dots, X_{n-1})$ and $v = (X_2, \dots, X_n)$. Figure 1 contains the histogram of estimates $\hat{\rho}$ and $\hat{\rho}_Q$ based on 10,000 replications of sample $X_1, \dots, X_{500}$ with outliers, generated by an AR(1) model with $\rho = 0.4$ and i.i.d. standard normal innovations. For a further discussion on robust estimation of autocorrelation function see Dürre *et al.* (2015).

Simulation results

Table 2 reports the empirical size at the 5% significance level based on 10,000 replications of $\tilde{M}_{C,n}$ and $\hat{M}_n$ tests, for the model (17) without outliers. The empirical size of $\hat{M}_n$ and $\tilde{M}_{C,n}$ slightly exceed the 5% level for large sample size n for $\theta = 0.5$ and $\Delta = 0.5, 1, 2$. The size of the tests is more distorted if the change-point is located close to the beginning or end of the sample, i.e. for $\theta = 0.25, 0.75$. We also consider the situation of no change, i.e. $\Delta = 0$, for which the empirical size of both testing procedures is close to the nominal size. Empirical sizes of $\hat{M}_n$ and $\tilde{M}_{C,n}$ are comparable in the absence of outliers.
Table 3 reports the empirical size of $\hat{M}_n$ and $\tilde{M}_{C,n}$ in presence of outliers. While test $\hat{M}_n$ is robust to the outliers, the test $\tilde{M}_{C,n}$ becomes too conservative.
Tables 4 and 5 report the empirical power of test $\tilde{M}_{C,n}$ and $\hat{M}_n$, for X_i in (18) without outliers and with outliers, respectively. Table 4 shows that the power of both tests increases with increasing sample size and dependence parameter d (except power of $\hat{M}_n$ for $n = 200$, $d = 0.4$). It shows that in absence of outliers $\hat{M}_n$ and $\tilde{M}_{C,n}$ have similar power properties.
Table 5 shows that the empirical size of $\hat{M}_n$ is practically not affected by the outliers, whereas $\tilde{M}_{C,n}$ suffers a loss of power.
Since the nominator of the CUSUM-type test is based on partial sums, outliers in the data have strong impact on the test statistic $\tilde{M}_{C,n}$ and hence, one should expect that it over rejects the true hypothesis H_0. Since presence of outliers increases the long-run variance estimate in (22) in the denominator of the test, this leads to additional

$\theta =$	0.25		0.5		0.75		0.5	
	$\tilde{M}_{C,n}$	$\hat{M}_n$	$\tilde{M}_{C,n}$	$\hat{M}_n$	$\tilde{M}_{C,n}$	$\hat{M}_n$	$\tilde{M}_{C,n}$	$\hat{M}_n$
n=	$\Delta = 1$						$\Delta = 0$	
200	3.79	3.52	3.90	3.41	4.46	3.92	3.48	2.78
500	8.35	7.71	5.12	4.28	8.47	8.10	4.36	3.89
1000	9.83	9.44	5.11	4.68	10.10	9.49	4.61	4.11
2000	9.45	9.37	5.96	5.23	9.87	9.76	5.10	4.64
5000	8.28	7.77	6.26	5.59	8.51	8.01	5.18	4.91
n=	$\Delta = 2$						$\Delta = 0.5$	
200	5.08	4.68	4.18	3.69	5.85	5.12	3.63	3.03
500	7.32	8.03	5.49	4.67	7.07	7.43	4.54	4.10
1000	7.67	8.05	5.38	4.79	7.15	7.38	4.82	4.46
2000	7.11	7.16	6.03	5.31	6.88	7.15	5.57	4.90
5000	6.30	6.12	6.15	5.58	6.45	6.29	6.01	5.46

Table 2: Empirical size of $\tilde{M}_{C,n}$ and $\hat{M}_n$ tests at the 5% significance level, 10,000 replications. X_i follows the model (17) without outliers.

$\theta =$	0.5			
	$\tilde{M}_{C,n}$	$\hat{M}_n$	$\tilde{M}_{C,n}$	$\hat{M}_n$
n=	$\Delta = 1$		$\Delta = 2$	
200	1.21	3.21	0.60	3.61
500	0.92	4.30	0.56	4.64
1000	0.86	4.72	0.62	4.77
2000	1.35	5.31	0.94	5.26
5000	2.67	5.71	1.95	5.58

Table 3: Empirical size of $\tilde{M}_{C,n}$ and $\hat{M}_n$ tests at the 5% significance level, 10,000 replications. X_i follows the model (17) with outliers.

d =	0.1		0.2		0.3		0.4	
n=	$\tilde{M}_{C,n}$	$\hat{M}_n$	$\tilde{M}_{C,n}$	$\hat{M}_n$	$\tilde{M}_{C,n}$	$\hat{M}_n$	$\tilde{M}_{C,n}$	$\hat{M}_n$
200	7.68	5.90	12.28	9.99	14.11	11.50	12.53	9.35
500	14.12	11.53	25.31	22.84	31.52	28.33	32.03	28.42
1000	20.22	16.95	35.37	32.64	46.41	43.11	50.22	46.06
2000	26.67	23.90	49.17	45.95	61.92	58.68	67.50	63.52
5000	35.05	32.68	64.44	61.27	79.67	77.48	85.12	82.63

Table 4: Empirical power of $\tilde{M}_{C,n}$ and $\hat{M}_n$ tests at the 5% significance level, 10,000 replications. X_i follows the model (18) without outliers.

d =	0.1		0.2		0.3		0.4	
n=	$\tilde{M}_{C,n}$	$\hat{M}_n$	$\tilde{M}_{C,n}$	$\hat{M}_n$	$\tilde{M}_{C,n}$	$\hat{M}_n$	$\tilde{M}_{C,n}$	$\hat{M}_n$
200	1.63	6.06	2.53	10.06	2.65	11.88	3.62	9.69
500	2.76	11.71	5.02	22.95	7.26	28.60	8.69	28.37
1000	4.10	17.13	10.40	32.60	16.91	43.11	21.96	46.18
2000	8.46	23.88	23.07	45.90	37.05	58.71	47.00	63.68
5000	18.76	32.66	46.78	61.55	68.99	77.54	78.65	82.68

Table 5: Empirical power of $\tilde{M}_{C,n}$ and $\hat{M}_n$ tests at the 5% significance level, 10,000 replications. X_i follows the model (18) with outliers.

reduction of size and a loss in power.

In general, we conclude that Wilcoxon test $\hat{M}_n$ allows discrimination between long-range dependence and short-range dependence with a change in mean that is robust to outliers. In absence of outliers it performs equally well as CUSUM test $\tilde{M}_{C,n}$, but outperforms it in presence of outliers.

4 Proofs

This section contains the proofs of Theorem 2.1, Theorem 2.2 and auxiliary lemmas.

4.1 Proof of Theorem 2.1

Suppose that $X_1, \ldots, X_n$ follow the model in (7) and Assumptions 1 and 2 are satisfied. Throughout the proofs without loss of generality, we assume $\mu = 0$ and $\Delta_n > 0$.

Proof of Theorem 2.1. We divide the proof into two steps, as in the proof of Theorem 2.1 in Berkes *et al.* (2006).
First, in Lemma 4.1 below we show that with $\hat{k}$ as in (3),

$$T_n(X_1, \ldots, X_{\hat{k}}) = T_n(Y_1, \ldots, Y_{\hat{k}}) + o_P(1)$$

and

$$T_n(X_{\hat{k}+1}, \ldots, X_n) = T_n(Y_{\hat{k}+1}, \ldots, Y_n) + o_P(1).$$

Subsequently, in Lemma 4.2 below we prove that

$$\Big(T_n(Y_1, \ldots, Y_{\hat{k}}), T_n(Y_{\hat{k}+1}, \ldots, Y_n)\Big) \xrightarrow{d} \sigma(Z^{(1)}, Z^{(2)}),$$

where $Z^{(i)} = \sup_{0 \le t \le 1} |B^{(i)}(t)|$, $i = 1, 2$. Then, the claim (14) of Theorem 2.1 follows by the continuous mapping theorem. □

Before proceeding to Lemma 4.1, similarly to the notation $W_{m,n}(k)$ in (2), we define

$$U_{m,n}(k) = \sum_{i=m}^{k} \sum_{j=k+1}^{n} (1_{\{Y_i \le Y_j\}} - 1/2), \qquad m \le k \le n. \tag{25}$$

Lemma 4.1. *Let $X_1, \dots, X_n$ follow the model in (7), and Assumptions 1 and 2 be satisfied. Let $\hat{k}$ be defined as in (3). Then,*

$$n^{-3/2} \max_{1 \le k \le \hat{k}} \left| W_{1,\hat{k}}(k) \right| = n^{-3/2} \max_{1 \le k \le \hat{k}} \left| U_{1,\hat{k}}(k) \right| + o_P(1) \tag{26}$$

$$n^{-3/2} \max_{\hat{k} < k \le n} \left| W_{\hat{k}+1,n}(k) \right| = n^{-3/2} \max_{\hat{k} < k \le n} \left| U_{\hat{k}+1,n}(k) \right| + o_P(1). \tag{27}$$

Proof. We have to distinguish between two cases, $\hat{k} \le k^*$ and $\hat{k} > k^*$, where $k^* = [n\theta]$. If $\hat{k} \le k^*$, then by (7), $X_i = Y_i$, $i = 1, \dots, \hat{k}$, and hence, $W_{1,\hat{k}}(k) = U_{1,\hat{k}}(k)$, $k = 1, \dots, \hat{k}$. In turn, $X_i = Y_i$ for $i = \hat{k}+1, \dots, k^*$, and $X_i = Y_i + \Delta_n$ for $i = k^*+1, \dots, n$. Since $1_{\{Y_i + \Delta_n \le Y_j + \Delta_n\}} = 1_{\{Y_i \le Y_j\}}$, $W_{\hat{k}+1,n}(k)$ can be decomposed into two terms,

$$W_{\hat{k}+1,n}(k) = \begin{cases} U_{\hat{k}+1,n}(k) + \sum_{i=\hat{k}+1}^{k} \sum_{j=k^*+1}^{n} 1_{\{Y_j < Y_i \le Y_j + \Delta_n\}}, & \hat{k} < k \le k^* \\ U_{\hat{k}+1,n}(k) + \sum_{i=\hat{k}+1}^{k^*} \sum_{j=k+1}^{n} 1_{\{Y_j < Y_i \le Y_j + \Delta_n\}}, & k^* < k \le n. \end{cases}$$

If $\hat{k} > k^*$, similar argument yields, $W_{\hat{k}+1,n}(k) = U_{\hat{k}+1,n}(k)$, for $k = \hat{k}+1, \dots, n$ and

$$W_{1,\hat{k}}(k) = \begin{cases} U_{1,\hat{k}}(k) + \sum_{i=1}^{k} \sum_{j=k^*+1}^{\hat{k}} 1_{\{Y_j < Y_i \le Y_j + \Delta_n\}}, & 1 \le k \le k^* \\ U_{1,\hat{k}}(k) + \sum_{i=1}^{k^*} \sum_{j=k+1}^{\hat{k}} 1_{\{Y_j < Y_i \le Y_j + \Delta_n\}}, & k^* < k \le \hat{k}. \end{cases} \tag{28}$$

Proof of (26). For $\hat{k} \le k^*$, equation (26) holds trivially, since $W_{1,\hat{k}}(k) = U_{1,\hat{k}}(k)$, $k = 1, \dots, \hat{k}$.
For $\hat{k} > k^*$, equation (28) yields,

$$\left| W_{1,\hat{k}}(k) - U_{1,\hat{k}}(k) \right| \le \sum_{i=1}^{k^*} \sum_{j=k^*+1}^{\hat{k}} 1_{\{Y_j < Y_i \le Y_j + \Delta_n\}} =: I_{1,\hat{k}}(k^*),$$

for all $1 \le k \le \hat{k}$. Hence, using Lemma 4.3 *i)*,

$$\left| n^{-3/2} \max_{1 \le k \le \hat{k}} \left| W_{1,\hat{k}}(k) \right| - n^{-3/2} \max_{1 \le k \le \hat{k}} \left| U_{1,\hat{k}}(k) \right| \right| \le n^{-3/2} I_{1,\hat{k}}(k^*).$$

Thus, property (26) holds if $n^{-3/2} I_{1,\hat{k}}(k^*) = o_P(1)$.
By Lemma 4.5 below, $n^{-3/2} I_{1,\hat{k}}(k^*) = n^{-3/2} k^* (\hat{k} - k^*) \Theta_{\Delta_n} + o_P(1)$, where $\Theta_{\Delta_n} = \mathrm{E}\left(1_{\{Y_2' < Y_1' \le Y_2' + \Delta_n\}}\right)$ and Y_1' and Y_2' are independent copies of Y_1. The distribution function F of Y_1 has bounded second derivative. Hence, as $n \to \infty$,

$$\begin{aligned} \Theta_{\Delta_n} &= \mathrm{E}\, 1_{\{Y_2' < Y_1' \le Y_2' + \Delta_n\}} = \mathrm{P}\left(Y_2' < Y_1' \le Y_2' + \Delta_n\right) \\ &= \int_{\mathbb{R}} \left(F(y + \Delta_n) - F(y)\right) dF(y) = \Delta_n \left(\int_{\mathbb{R}} f^2(y)\, dy + o(1) \right). \end{aligned} \tag{29}$$

Furthermore, by (12), $\Delta_n^2|\hat{k}-k^*| = O_P(1)$ and by Assumption 2, $k^*/n \sim \theta$ and $n\Delta_n^2 \to \infty$, as $n \to \infty$. This yields

$$n^{-3/2}k^*|\hat{k}-k^*|\Theta_{\Delta_n} \leq C\frac{\Delta_n^2|\hat{k}-k^*|}{n^{1/2}\Delta_n} = o_P(1).$$

This completes the proof of (26). The proof of (27) follows using similar argument. □

Lemma 4.2. *Let (Y_j) satisfy Assumption 1 and let Assumption 2 hold. Then,*

$$\Big(T(Y_1,\ldots,Y_{\hat{k}}), T(Y_{\hat{k}+1},\ldots,Y_n)\Big) \xrightarrow{d} \Big(\sigma \sup_{0\leq t\leq 1}\left|B^{(1)}(t)\right|, \sigma \sup_{0\leq t\leq 1}\left|B^{(2)}(t)\right|\Big), \tag{30}$$

where $B^{(1)}$ and $B^{(2)}$ are independent Brownian bridges, and σ is given in (15).

Proof. To prove Lemma 4.2 we will use the idea of the proof of Theorem 3 of Dehling *et al.* (2015).
Recall that $T(Y_1,\ldots,Y_{\hat{k}}) = \hat{k}^{-3/2}\max_{1\leq k\leq \hat{k}}|U_{1,\hat{k}}(k)|$ and similarly $T(Y_{\hat{k}+1},\ldots,Y_n) = (n-\hat{k})^{-3/2}\max_{\hat{k}<k\leq n}|U_{\hat{k}+1,n}(k)|$. Note that the terms $U_{1,\hat{k}}(k)$ and $U_{\hat{k}+1,n}(k)$ defined in (25) can be written as a second order U-statistic

$$U_{a,b}(k) = \sum_{i=a}^{k}\sum_{j=k+1}^{b}\left(h(Y_i,Y_j)-\Theta\right), \qquad a\leq k<b,$$

with kernel function $h(x,y) = 1_{\{x\leq y\}}$ and constant $\Theta = \operatorname{E} h(Y_1',Y_2') = 1/2$, where Y_1' and Y_2' are independent copies of Y_1.
By applying Hoeffding's decomposition of U-statistics to $U_{a,b}(k)$, the kernel function h can be written as the sum

$$h(x,y) = \Theta + h_1(x) + h_2(y) + g(x,y), \tag{31}$$

where $h_1(x) = \operatorname{E} h(x,Y_2') - \Theta = 1/2 - F(x)$,

$$h_2(y) = \operatorname{E} h\left(Y_1',y\right) - \Theta = F(y) - 1/2, \qquad g(x,y) = h(x,y) - h_1(x) - h_2(y) - \Theta.$$

Therefore,

$$U_{a,b}(k) = \sum_{i=a}^{k}\sum_{j=k+1}^{b}\left(h_1(Y_i) + h_2(Y_j) + g(Y_i,Y_j)\right) =: s_{a,b}(k) + v_{a,b}(k),$$

where

$$s_{a,b}(k) = (b-k)\sum_{i=a}^{k}h_1(Y_i) + (k-a+1)\sum_{j=k+1}^{b}h_2(Y_j), \qquad v_{a,b}(k) = \sum_{i=a}^{k}\sum_{j=k+1}^{b}g(Y_i,Y_j).$$

Note that

$$v_{a,b}(k)=\sum_{i=1}^{k}\sum_{j=1}^{b}g\left(Y_i,Y_j\right)-\sum_{i=1}^{k}\sum_{j=1}^{k}g\left(Y_i,Y_j\right)-\sum_{i=1}^{a-1}\sum_{j=1}^{b}g\left(Y_i,Y_j\right)+\sum_{i=1}^{a-1}\sum_{j=1}^{k}g\left(Y_i,Y_j\right).$$

Thus, Lemma 4.4 below yields

$$n^{-3/2}\max_{a\le k\le b}\left|v_{a,b}(k)\right|\le 4n^{-3/2}\max_{1\le k\le n}\max_{1\le l\le n}\left|\sum_{i=1}^{k}\sum_{j=1}^{l}g\left(Y_i,Y_j\right)\right|=o_P(1).$$

Furthermore, by Lemma 4.3 *ii)*,

$$\max_{a\le k\le b}\left|U_{a,b}(k)\right|=\max_{a\le k\le b}\left|s_{a,b}(k)\right|+\max_{a\le k\le b}\left|v_{a,b}(k)\right|=\max_{a\le k\le b}\left|s_{a,b}(k)\right|+o_P(n^{3/2}).$$

It remains to show that

$$\hat{k}^{-3/2}\max_{1\le k\le\hat{k}}\left|s_{1,\hat{k}}(k)\right|\xrightarrow{d}\sigma\sup_{0\le t\le 1}\left|B^{(1)}\left(t\right)\right|,$$
$$(n-\hat{k})^{-3/2}\max_{\hat{k}<k\le n}\left|s_{\hat{k}+1,n}(k)\right|\xrightarrow{d}\sigma\sup_{0\le t\le 1}\left|B^{(2)}\left(t\right)\right|,$$

where $B^{(1)}$ and $B^{(2)}$ are independent Brownian bridges. By Slutsky's Lemma this implies (30). Note that $h_1(x)=-h_2(x)$. Hence,

$$\begin{aligned}s_{1,\hat{k}}(k)&=(\hat{k}-k)\sum_{i=1}^{k}h_1(Y_i)+k\sum_{j=k+1}^{\hat{k}}h_2(Y_j)\\&=\hat{k}n^{1/2}\Big\{\frac{1}{n^{1/2}}\sum_{i=1}^{k}h_1\left(Y_i\right)-\frac{k}{\hat{k}}\frac{1}{n^{1/2}}\sum_{i=1}^{\hat{k}}h_1\left(Y_i\right)\Big\}=:\hat{k}n^{1/2}\Gamma_k^{(1)}\end{aligned}$$

and

$$\begin{aligned}&s_{\hat{k}+1,n}(k)=(n-k)\sum_{i=\hat{k}+1}^{k}h_1(Y_i)+(k-\hat{k})\sum_{j=k+1}^{n}h_1(Y_j)\\&=(n-\hat{k})n^{1/2}\Big\{\frac{1}{n^{1/2}}\sum_{i=\hat{k}+1}^{k}h_1\left(Y_i\right)-\frac{k-\hat{k}}{n-\hat{k}}\frac{1}{n^{1/2}}\sum_{i=\hat{k}+1}^{n}h_1\left(Y_i\right)\Big\}\\&=(n-\hat{k})n^{1/2}\Big\{\frac{1}{n^{1/2}}\Big(\sum_{i=1}^{k}h_1\left(Y_i\right)-\sum_{i=1}^{\hat{k}}h_1\left(Y_i\right)\Big)-\frac{k-\hat{k}}{n-\hat{k}}\frac{1}{n^{1/2}}\Big(\sum_{i=1}^{n}h_1\left(Y_i\right)-\sum_{i=1}^{\hat{k}}h_1\left(Y_i\right)\Big)\Big\}\\&=:(n-\hat{k})n^{1/2}\Gamma_k^{(2)}.\end{aligned}$$

Corollary 4.1 below implies convergence of finite dimensional distribution of the partial sum process,

$$\left(\frac{1}{n^{1/2}}\sum_{i=1}^{[nt]}h_1\left(Y_i\right)\right)_{0\le t\le 1}\xrightarrow{d}\left(\sigma W\left(t\right)\right)_{0\le t\le 1},$$

where $W(t)$ is a Brownian motion and σ as in (15). By the Skorokhod-Wichura-Dudley representation (see e.g., Shorack and Wellner (2009), Theorem 4 on page 47) there exists a series of Brownian motions $W_n(t)$, $t \in [0,1]$, such that

$$\sup_{0\le t\le 1}\left|n^{-1/2}\sum_{i=1}^{[nt]} h_1(Y_i) - \sigma W_n(t)\right| = o_P(1).$$

Set

$$\Gamma^{(1)}_{W,k} = W_n\Big(\frac{k}{n}\Big) - \frac{k}{\hat{k}} W_n\Big(\frac{\hat{k}}{n}\Big), \quad \Gamma^{(2)}_{W,k} = \Big(W_n\Big(\frac{k}{n}\Big) - W_n\Big(\frac{\hat{k}}{n}\Big)\Big) - \frac{k-\hat{k}}{n-\hat{k}}\Big(W_n(1) - W_n\Big(\frac{\hat{k}}{n}\Big)\Big).$$

Thus,

$$\max_{1\le k\le \hat{k}} \big|\Gamma^{(1)}_k - \sigma\Gamma^{(1)}_{W,k}\big| = o_P(1), \qquad \max_{\hat{k}<k\le n} \big|\Gamma^{(2)}_k - \sigma\Gamma^{(2)}_{W,k}\big| = o_P(1).$$

Consistency of $\hat{k}$ in (12), $\Delta_n^2|\hat{k}-k| = O_P(1)$, and Assumption 2, $n\Delta_n^2 \to \infty$, as $n \to \infty$, yield

$$\left|\frac{\hat{k}}{n} - \theta\right| = o_P(1).$$

Therefore, by the continuity of Brownian motion W_n and using the continuous mapping theorem, $\big|W_n\big(\hat{k}/n\big) - W_n(\theta)\big| = o_P(1)$. Hence,

$$\max_{1\le k\le \hat{k}} \big|\Gamma^{(1)}_{W,k}\big| = \sup_{0\le t\le \theta}\left|W_n(t) - \frac{t}{\theta}W_n(\theta)\right| + o_P(1)$$

and

$$\begin{aligned}\max_{\hat{k}<k\le n} \big|\Gamma^{(2)}_{W,k}\big| &= \sup_{\theta<t\le 1}\left|\Big(W_n(t) - W_n(\theta)\Big) - \frac{t-\theta}{1-\theta}\Big(W_n(1) - W_n(\theta)\Big)\right| + o_P(1)\\ &\overset{d}{=} \sup_{\theta<t\le 1}\left|W_n(t-\theta) - \frac{t-\theta}{1-\theta}W_n(1-\theta)\right|,\end{aligned}$$

since Brownian motions have stationary increments and $W_n(0) = 0$. Finally,

$$(\hat{k}/n)^{-1/2}\max_{1\le k\le \hat{k}} \big|\Gamma^{(1)}_k\big| = \frac{\sigma}{\theta^{1/2}}\sup_{0\le t\le \theta}\left|W_n(t) - \frac{t}{\theta}W_n(\theta)\right| + o_P(1) \overset{d}{=} \sigma \sup_{0\le t\le 1}\big|B^{(1)}(t)\big|,$$

since Brownian motions are scale invariant, i.e. $\theta^{-1/2}W_n(t) \overset{d}{=} W_n(t/\theta)$, and

$$\begin{aligned}((n-\hat{k})/n)^{-1/2}\max_{\hat{k}<k\le n} \big|\Gamma^{(2)}_k\big| &\overset{d}{=} \frac{\sigma}{(1-\theta)^{1/2}}\sup_{\theta<t\le 1}\left|W_n(t-\theta) - \frac{t-\theta}{1-\theta}W_n(1-\theta)\right|\\ &\overset{d}{=} \frac{\sigma}{(1-\theta)^{1/2}}\sup_{0<t\le 1-\theta}\left|W_n(t) - \frac{t}{1-\theta}W_n(1-\theta)\right| \overset{d}{=} \sigma\sup_{0\le t\le 1}\big|B^{(2)}(t)\big|.\end{aligned}$$

The increments of Brownian motions are independent, thus $B^{(1)}$ and $B^{(2)}$ are independent. This proves the lemma. □

Concept of 1-continuity

Before we state the auxiliary results, we recall the concept of 1-*continuity*, which was introduced by Borovkova *et al.* (2001).

To study the asymptotic behaviour of the Wilcoxon test

$$W_{1,n}(k) = \sum_{i=1}^{k} \sum_{j=k+1}^{n} (1_{\{X_i \le X_j\}} - 1/2)$$

we need to show that the function $h(x,y) = 1_{\{x \le y\}}$ is 1-continuous. Then the variables $(h(Y_i, Y_j))$ retain some characteristics of the variables (Y_i, Y_j).

Definition 4.1. *(Borovkova et al. (2001))*
We say that the kernel $h(x,y)$ is 1-continuous with respect to a distribution of a stationary process (Y_j) if there exists a function $\phi(\epsilon)$, $\epsilon \ge 0$ such that $\phi(\epsilon) \to 0$, $\epsilon \to 0$, and for all $\epsilon > 0$ and $k \ge 1$

$$\begin{aligned} &\mathrm{E}\left(\left|h\left(Y_1, Y_k\right) - h\left(Y_1', Y_k\right)\right| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) \le \phi(\epsilon), \\ &\mathrm{E}\left(\left|h\left(Y_k, Y_1\right) - h\left(Y_k, Y_1'\right)\right| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) \le \phi(\epsilon), \end{aligned} \tag{32}$$

and

$$\begin{aligned} &\mathrm{E}\left(\left|h\left(Y_1, Y_2'\right) - h\left(Y_1', Y_2'\right)\right| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) \le \phi(\epsilon), \\ &\mathrm{E}\left(\left|h\left(Y_2', Y_1\right) - h\left(Y_2', Y_1'\right)\right| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) \le \phi(\epsilon), \end{aligned} \tag{33}$$

where Y_2' is an independent copy of Y_1 and Y_1' is any random variable that has the same distribution as Y_1.

For a univariate function $g(x)$, the 1-continuity property is defined as follows.

Definition 4.2. *The function $g(x)$ is 1-continuous with respect to a distribution of a stationary process (Y_j) if there exists a function $\phi(\epsilon)$, $\epsilon \ge 0$ such that $\phi(\epsilon) \to 0$, $\epsilon \to 0$, and for all $\epsilon > 0$*

$$\mathrm{E}\left(\left|g\left(Y_1\right) - g\left(Y_1'\right)\right| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) \le \phi(\epsilon), \tag{34}$$

where Y_1' is any random variable that has the same distribution as Y_1.

The following remark states functions $h(x,y) = 1_{\{x \le y\}}$, $h_1(x)$, $h_2(x)$ and $g(x,y)$ appearing in the Hoeffding decomposition (31) are 1-continuous functions.

Remark 4.1. Let (Y_j) be a stationary process, Y_1 has continuous distribution function F with bounded second derivative and the variables $Y_1 - Y_k$, $k \ge 1$ satisfy (11).

i) The function $h(x,y) = 1_{\{x \le y\}}$ is 1-continuous function (i.e. satisfies (32) and (33)) with respect to the distribution of (Y_j) with function $\phi(\epsilon) = C\epsilon$, for some $C > 0$, see e.g. Corollary 4.1 of Gerstenberger (2018).

ii) Lemma 2.15 of Borovkova *et al.* (2001) yields that if a general function $h(x,y)$ satisfies (32) and (33) with some function $\phi(\epsilon)$ then $\mathrm{E}\, h(x, Y_2')$, where Y_2' is an independent copy of Y_1, satisfies the condition in (34) with the same function $\phi(\epsilon)$. Hence, $h_1(x) = \mathrm{E}\, h(x, Y_2') - 1/2$ and $h_2(x) = \mathrm{E}\, h(Y_2', y) - 1/2$ are 1-continuous.

iii) The function $g(x,y) = h(x,y) - h_1(x) - h_2(x) - 1/2$ is 1-continuous (satisfies (32) and (33)), since h and h_1 satisfy (32), (33) and (34) with $\phi(\epsilon) = C\epsilon$, for some $C > 0$. In particular,

$$
\begin{aligned}
&\mathrm{E}\left(|g(Y_1, Y_k) - g(Y_1', Y_k)| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) \\
&\le \mathrm{E}\left(|h(Y_1, Y_k) - h(Y_1', Y_k)| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) + \mathrm{E}\left(|h_1(Y_1) - h_1(Y_1')| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) \\
&\le 2\phi(\epsilon)
\end{aligned}
$$

and similarly, $\mathrm{E}\left(|g(Y_k, Y_1) - g(Y_k, Y_1')| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) \le 2\phi(\epsilon)$.

Auxiliary results

The following lemma yields maximum inequalities used in the proofs of Lemma 4.1 and Lemma 4.2.

Lemma 4.3. *Let (a_k) and (b_k) be two sequences of real numbers and $c_k = a_k + b_k$. Then,*

i) $\big|\max_k |a_k| - \max_k |b_k|\big| \le \max_k |a_k - b_k|$

ii) $\max_k |a_k| - \max_k |b_k| \le \max_k |c_k| \le \max_k |a_k| + \max_k |b_k|$.

Proof. We start with the *proof of i).* Assume that $\max_k |a_k| \ge \max_k |b_k|$ and define $\tilde{k} = \arg\max_k |a_k|$. Note that $|b_{\tilde{k}}| \le \max_k |b_k|$. Then,

$$
\begin{aligned}
\big|\max_k |a_k| - \max_k |b_k|\big| = \max_k |a_k| - \max_k |b_k| &\le |a_{\tilde{k}}| - |b_{\tilde{k}}| \le \big||a_{\tilde{k}}| - |b_{\tilde{k}}|\big| \le |a_{\tilde{k}} - b_{\tilde{k}}| \\
&\le \max_k |a_k - b_k|.
\end{aligned}
$$

For $\max_k |a_k| \le \max_k |b_k|$ the proof follows a similar argument using $\tilde{k} = \arg\max_k |b_k|$.

Proof of ii). It is obvious that $\max_k |c_k| = \max_k |a_k + b_k| \le \max_k |a_k| + \max_k |b_k|$. Define $\tilde{k} = \arg\max_k |a_k|$. Then

$$
\max_k |a_k + b_k| \ge \max_k (|a_k| - |b_k|) \ge |a_{\tilde{k}}| - |b_{\tilde{k}}| \ge \max_k |a_k| - \max_k |b_k|,
$$

which finishes the proof. □

The following lemma derives the functional central limit theorem for partial sum processes of $(h_1(Y_j))$.

Corollary 4.1. *Suppose that the assumptions of Lemma* 4.2 *hold. Then,*

$$\left(\frac{1}{n^{1/2}}\sum_{i=1}^{[nt]} h_1\left(Y_i\right)\right)_{0\le t\le 1} \xrightarrow{d} \left(\sigma W\left(t\right)\right)_{0\le t\le 1},$$

where $W(t)$ is a Brownian motion and σ is given in (15).

Proof. Wooldridge and White (1988) in Corollary 3.2 established a functional central limit theorem for partial sum process $\sum_{i=1}^{k}\tilde{Y}_i$, $k\ge 1$, for a process $(\tilde{Y}_j)$ which is L_2 NED on a strongly mixing process $(\tilde{Z}_j)$. Therefore, Corollary 4.1 is proved, by showing that $(h_1(Y_j))$ is L_2 NED on a strongly mixing process.
By Proposition 2.11 of Borovkova *et al.* (2001), if (Y_j) is L_1 NED on a stationary absolutely regular process (Z_j) with approximation constants a_k and $g(x)$ is 1-continuous with function ϕ, then $(g(Y_j))$ is also L_1 NED on (Z_j) with approximation constants $a'_k = \phi\left(\sqrt{2a_k}\right)+2\sqrt{2a_k}||g||_\infty$. By Remark 4.1 ii), $h_1(x)=1/2-F(x)$ is 1-continuous function with $\phi(\epsilon)=C\epsilon$. Thus, the processes $(h_1(Y_j))$ is L_1 NED processes with approximation constants $a'_k = C\sqrt{a_k}\ge\phi\left(\sqrt{2a_k}\right)+2\sqrt{2a_k}||h_1||_\infty$.
Observe that the variables $\eta_k := h_1(Y_1)-\mathrm{E}(h_1(Y_1)|\mathcal{G}_{-k}^{k})$ satisfy the L_1 NED condition (9) with a'_k. To show L_2 NED for $(h_1(Y_j))$ note that by definition of h_1, $\mathrm{E}\,h_1(Y_1)=0$ and $|h_1(Y_1)|\le C<\infty$. Thus,

$$\mathrm{E}\,\eta_k^2 \le \mathrm{E}\left(|\eta_k|\cdot\left(|h_1(Y_1)|+|\,\mathrm{E}(h_1(Y_1)|\mathcal{G}_{-k}^{k})|\right)\right)\le C\,\mathrm{E}\,|\eta_k|\le Ca'_k.$$

The last inequality holds, because by L_1 NED of $(h_1(Y_j))$, $\mathrm{E}\,|h_1(Y_1)-\mathrm{E}(h_1(Y_1)|\mathcal{G}_{-k}^{k})|\le a'_k$. Therefore, the process $(h_1(Y_j))$ is also L_2 NED on (Z_j) with approximation constant $a'_k = Ca_k^{1/2}$. Moreover, absolute regularity of (Z_j) implies the process (Z_j) is also strong mixing. Assumption (10) yields $a'_k = O(k^{-1/2})$ and $\beta_k = O(k^{-2})$. Thus, $(h_1(Y_j))$ satisfies the conditions of Corollary 3.2 of Wooldridge and White (1988) which implies

$$\left(\frac{1}{n^{1/2}}\sum_{i=1}^{[nt]} h_1\left(Y_i\right)\right)_{0\le t\le 1} \xrightarrow{d} \left(\sigma W\left(t\right)\right)_{0\le t\le 1},$$

where $W(t)$ is a Brownian motion and $\sigma^2=\sum_{k=-\infty}^{\infty}\mathrm{Cov}(F(Y_1),F(Y_k))$. □

Next we show that the contribution of $g(x,y)$ of the Hoeffding decomposition (31) is negligible.

Lemma 4.4. *Suppose that the assumptions of Lemma* 4.2 *hold. Then,*

$$n^{-3/2}\max_{1\le k\le n}\max_{1\le l\le n}\Big|\sum_{i=1}^{k}\sum_{j=1}^{l} g(Y_i,Y_j)\Big| = o_P(1). \tag{35}$$

Proof. We first prove for $1 \le q \le p \le n$, $1 \le h \le l \le n$,

$$\mathrm{E}\left(\left|n^{-3/2} \sum_{i=q+1}^{p} \sum_{j=h+1}^{l} g(Y_i, Y_j)\right|^2\right) \le \frac{C}{n^3}(p-q)(l-h). \tag{36}$$

Proof of (36) Lemma 1 of Dehling *et al.* (2015) showed if f is a 1-continuous bounded degenerate kernel function and $\phi_f(\epsilon)$ satisfies

$$\sum_{k=1}^{\infty} k(\beta(k) + \sqrt{a_k} + \phi_f(a_k)) < \infty, \tag{37}$$

then

$$\mathrm{E}\Big(\sum_{i=1}^{k} \sum_{j=k+1}^{n} f(Y_i, Y_j)\Big)^2 \le Ck(n-k), \qquad 1 \le k \le n. \tag{38}$$

The proof of Lemma 1 in Dehling *et al.* (2015) shows that (38) can be extended to (36). Hence, to complete the proof, we need to verify that $g(x, y)$ satisfies the assumptions of Lemma 1 of Dehling *et al.* (2015).
By the Hoeffding decomposition (31), $g(x, y) = h(x, y) + F(x) - F(y) - 1/2$. Note that $\mathrm{E}\, F(Y_1) = 1/2$, thus $\mathrm{E}\, g(x, Y_1) = \mathrm{E}\, g(Y_1, y) = 0$, i.e. $g(x, y)$ is a degenerate kernel. Furthermore, $g(x, y)$ is bounded, since $h(x, y) = 1_{\{x \le y\}}$ and $F(x)$ are bounded. By Remark 4.1 iii) $g(x, y)$ is 1-continuous with $\phi(\epsilon) = C\epsilon$, the latter satisfies (37) because of condition (10). This completes the proof of (36).

Proof of (35) To prove the lemma, we use Theorem 10.2 of Billingsley (1999), which states that if the increments of partial sums $S_i = \sum_{j=1}^{i} \zeta_i$ of random variables ζ_i, $i = 1, 2, \ldots$ are bounded in probability, in particular if there exist $\alpha > 1$, $\beta > 0$ and non-negative numbers $u_{n,1}, \ldots, u_{n,n}$ such that

$$\mathrm{P}\left(|S_j - S_i| \ge \epsilon\right) \le \frac{1}{\epsilon^{\beta}} \left(\sum_{l=i+1}^{j} u_{n,l}\right)^{\alpha},$$

for $\epsilon > 0$, $0 \le i \le j \le n$, then for all $\epsilon > 0$, $n \ge 2$,

$$\mathrm{P}\left(\max_{1 \le k \le n} |S_k| \ge \epsilon\right) \le \frac{K}{\epsilon^{\beta}} \left(\sum_{l=1}^{n} u_{n,l}\right)^{\alpha},$$

where $K > 0$ depends only on α and β.
Denote

$$G_n(l) = n^{-3/2} \max_{1 \le k \le n} \Big|\sum_{i=1}^{k} \sum_{j=1}^{l} g(Y_i, Y_j)\Big|,$$

with $G_n(0) = 0$ and define random variables $\zeta_i = G_n(i) - G_n(i-1)$, where $\zeta_0 = 0$. Note that $S_i = \sum_{j=1}^{i} \zeta_i = G_n(i)$ and by Lemma 4.3 *i)* for $1 \le h \le l \le n$,

$$\begin{aligned} \mathrm{P}\left(|S_l - S_h| \ge \epsilon\right) &\le \mathrm{P}\Big(n^{-3/2} \max_{1\le k\le n} \Big| \sum_{i=1}^{k}\sum_{j=1}^{l} g(Y_i, Y_j) - \sum_{i=1}^{k}\sum_{j=1}^{h} g(Y_i, Y_j)\Big| \ge \epsilon\Big) \\ &= \mathrm{P}\Big(n^{-3/2} \max_{1\le k\le n} \Big| \sum_{i=1}^{k}\sum_{j=h+1}^{l} g(Y_i, Y_j)\Big| \ge \epsilon\Big). \end{aligned}$$

Let us now define

$$\tilde{S}_k = \sum_{i=1}^{k} \Big(n^{-3/2} \sum_{j=h+1}^{l} g(Y_i, Y_j)\Big).$$

Note that for $1 \le q \le p \le n$,

$$\big|\tilde{S}_p - \tilde{S}_q\big| = n^{-3/2} \Big| \sum_{i=q+1}^{p}\sum_{j=h+1}^{l} g(Y_i, Y_j)\Big|.$$

By Markov inequality and (36),

$$\mathrm{P}\left(\big|\tilde{S}_p - \tilde{S}_q\big| \ge \epsilon\right) \le \frac{1}{\epsilon^2}\,\mathrm{E}\left(\big|\tilde{S}_p - \tilde{S}_q\big|^2\right) \le \frac{1}{\epsilon^2}\frac{C}{n^3}(p-q)(l-h) \le \frac{1}{\epsilon^2}\Big(\sum_{t=q+1}^{p} u_{n,t}\Big)^{4/3},$$

where $u_{n,t} = \frac{C^{3/4}}{n^{9/4}}(l-h)$. Hence, $\tilde{S}_i$ satisfies assumption of Theorem 10.2 of Billingsley (1999) with $\beta = 2$, $\alpha = 4/3$. Thus, for any fixed $\epsilon > 0$,

$$\mathrm{P}\Big(\max_{1\le k\le n} \big|\tilde{S}_k\big| \ge \epsilon\Big) \le \frac{K}{\epsilon^2}\Big(\sum_{t=1}^{n} \frac{C^{3/4}}{n^{9/4}}(l-h)\Big)^{4/3} \le \frac{1}{\epsilon^2}\Big((l-h)\frac{C^{3/4}}{n^{5/4}}\Big)^{4/3}$$

and moreover

$$\mathrm{P}\left(|S_l - S_h| \ge \epsilon\right) \le \mathrm{P}\Big(\max_{1\le k\le n} \big|\tilde{S}_k\big| \ge \epsilon\Big) \le \frac{1}{\epsilon^2}\Big(\sum_{t=h+1}^{l} u_{n,t}\Big)^{4/3},$$

where $u_{n,t} = \frac{C^{3/4}}{n^{5/4}}$. Therefore, S_i satisfies assumption of Theorem 10.2 of Billingsley (1999) with $\beta = 2$, $\alpha = 4/3$. Finally, for any fixed $\epsilon > 0$, as $n \to \infty$,

$$\begin{aligned} &\mathrm{P}\Big(n^{-3/2} \max_{1\le l\le n}\max_{1\le k\le n} \Big| \sum_{i=1}^{k}\sum_{j=1}^{l} g(Y_i, Y_j)\Big| \ge \epsilon\Big) \\ &\qquad = \mathrm{P}\Big(\max_{1\le l\le n} \big|S_l\big| \ge \epsilon\Big) \le \frac{K}{\epsilon^2}\Big(\sum_{t=1}^{n} \frac{C^{3/4}}{n^{5/4}}\Big)^{4/3} \le \frac{K}{\epsilon^2}\frac{1}{n^{1/3}} \to 0, \end{aligned}$$

which proves the lemma. □

In the following we state auxiliary results to deal with the terms

$$\tilde{U}_{1,\hat{k}}(k^*) := \sum_{i=1}^{k^*} \sum_{j=k^*+1}^{\hat{k}} 1_{\{Y_j < Y_i \le Y_j + \Delta_n\}}, \qquad \hat{k} \ge k^*,$$

and

$$\tilde{U}_{\hat{k}+1,n}(k^*) := \sum_{i=\hat{k}+1}^{k^*} \sum_{j=k^*+1}^{n} 1_{\{Y_j < Y_i \le Y_j + \Delta_n\}}, \qquad \hat{k} < k^*$$

appearing in the proof of Lemma 4.1.
Note that the terms $\tilde{U}_{1,\hat{k}}(k^*)$ and $\tilde{U}_{\hat{k}+1,n}(k^*)$ can be written as a second order U-statistic

$$\tilde{U}_{a,b}(k) = \sum_{i=a}^{k} \sum_{j=k+1}^{b} h_n(Y_i, Y_j), \qquad a \le k < b,$$

with kernel function $h_n(x, y) = 1_{\{y < x \le y + \Delta_n\}}$.
Applying Hoeffding's decomposition of U-statistics (Hoeffding (1948)) to $\tilde{U}_{a,b}(k)$, decomposes the kernel function h_n into the sum

$$h_n(x, y) = \Theta_{\Delta_n} + h_{1,n}(x) + h_{2,n}(y) + g_n(x, y), \tag{39}$$

with $\Theta_{\Delta_n} = \mathrm{E}\left(1_{\{Y_2' < Y_1' \le Y_2' + \Delta_n\}}\right)$,

$$\begin{aligned} h_{1,n}(x) &= \mathrm{E}\, h_n\left(x, Y_2'\right) - \Theta_{\Delta_n} = F(x) - F(x - \Delta_n) - \Theta_{\Delta_n}, \\ h_{2,n}(y) &= \mathrm{E}\, h_n\left(Y_1', y\right) - \Theta_{\Delta_n} = F(y + \Delta_n) - F(y) - \Theta_{\Delta_n}, \\ g_n(x, y) &= h_n(x, y) - h_{1,n}(x) - h_{2,n}(y) - \Theta_{\Delta_n}, \end{aligned}$$

where Y_1' and Y_2' are independent copies of Y_1.

Lemma 4.5. *Suppose that the assumptions of Lemma* 4.1 *hold. Then,*

$$n^{-3/2}\left|\tilde{U}_{1,\hat{k}}(k^*) - k^*(\hat{k} - k^*)\Theta_{\Delta_n}\right| = o_P(1) \tag{40}$$

and

$$n^{-3/2}\left|\tilde{U}_{\hat{k}+1,n}(k^*) - (k^* - \hat{k})(n - k^*)\Theta_{\Delta_n}\right| = o_P(1), \tag{41}$$

where $\Theta_{\Delta_n} = \mathrm{E}\left(1_{\{Y_2' < Y_1' \le Y_2' + \Delta_n\}}\right)$ *and* Y_1' *and* Y_2' *are independent copies of* Y_1 .

Proof. Let us start with the proof of (40). The Hoeffding decomposition (39) yields

$$\begin{aligned} \tilde{U}_{1,\hat{k}}(k^*) - k^*(\hat{k} - k^*)\Theta_{\Delta_n} &= \sum_{i=1}^{k^*} \sum_{j=k^*+1}^{\hat{k}} \left(h_{1,n}(Y_i) + h_{2,n}(Y_j) + g_n(Y_i, Y_j)\right) \\ &= (\hat{k} - k^*) \sum_{i=1}^{k^*} h_{1,n}(Y_i) + k^* \sum_{j=k^*+1}^{\hat{k}} h_{2,n}(Y_j) + \sum_{i=1}^{k^*} \sum_{j=k^*+1}^{\hat{k}} g_n(Y_i, Y_j). \end{aligned}$$

Therefore,

$$\begin{aligned}
&n^{-3/2}\Big|\tilde{U}_{1,\hat{k}}(k^*) - k^*(\hat{k}-k^*)\Theta_{\Delta_n}\Big| \\
&\leq n^{-3/2}\Big|(\hat{k}-k^*)\sum_{i=1}^{k^*} h_{1,n}(Y_i) + k^* \sum_{j=k^*+1}^{\hat{k}} h_{2,n}(Y_j)\Big| + n^{-3/2}\Big|\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{\hat{k}} g_n(Y_i,Y_j)\Big|.
\end{aligned}$$

Note that the indicator function $h_n(x,y) = 1_{\{y<x\leq y+\Delta_n\}}$ is bounded. Furthermore, by (29), $\Theta_{\Delta_n} \sim C\Delta_n$, thus

$$\begin{aligned}
|h_{1,n}(x)| &\leq |F(x) - F(x-\Delta_n) - \Theta_{\Delta_n}| \leq C\Delta_n + \Theta_{\Delta_n} \leq C\Delta_n, \\
|h_{2,n}(x)| &\leq |F(x+\Delta_n) - F(x) - \Theta_{\Delta_n}| \leq C\Delta_n + \Theta_{\Delta_n} \leq C\Delta_n,
\end{aligned} \tag{42}$$

where $C > 0$ is a constant. Hence, $g_n(x,y) = h_n(x,y) - h_{1,n}(x) - h_{2,n}(y) - \Theta_{\Delta_n}$ is bounded. Since $\mathrm{E}\, h_{1,n}(Y_1) = 0$ and $\mathrm{E}\, h_{2,n}(Y_1) = 0$, $g_n(x,y)$ is a degenerate kernel, i.e. $\mathrm{E}\, g_n(x,Y_1) = \mathrm{E}\, g_n(Y_1,y) = 0$. $h_n(x,y)$ satisfies (32) and (33) with $\phi_{h_n}(\epsilon) = C\epsilon$, see e.g. Corollary 4.1 of Gerstenberger (2018). Then, with similar argument as in Remark 4.1, $h_{1,n}$ and $h_{2,n}$ are 1-continuous and therefore, $g_n(x,y)$ is 1-continuous with function $\phi_{g_n}(\epsilon) = C\epsilon$ satisfying (37). Hence, $g_n(x,y)$ satisfies the conditions on $g(x,y)$ in Lemma 4.4, which yields

$$n^{-3/2}\Big|\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{\hat{k}} g_n(Y_i,Y_j)\Big| \leq 2 \max_{1\leq k\leq n}\max_{1\leq k\leq n} n^{-3/2}\Big|\sum_{i=1}^{k}\sum_{j=1}^{l} g_n(Y_i,Y_j)\Big| = o_P(1).$$

Thus, it remains to show $n^{-3/2}\big|(\hat{k}-k^*)\sum_{i=1}^{k^*} h_{1,n}(Y_i) + k^*\sum_{j=k^*+1}^{\hat{k}} h_{2,n}(Y_j)\big| = o_P(1)$. By (42), we receive the following inequality

$$\begin{aligned}
&n^{-3/2}\Big|(\hat{k}-k^*)\sum_{i=1}^{k^*} h_{1,n}(Y_i) + k^* \sum_{j=k^*+1}^{\hat{k}} h_{2,n}(Y_j)\Big| \\
&\qquad\qquad \leq n^{-3/2}C(\hat{k}-k^*)k^*\Delta_n = C\frac{k^*}{n}\frac{\Delta_n^2|\hat{k}-k^*|}{n^{1/2}\Delta_n} = o_P(1),
\end{aligned}$$

where we used the consistency of $\hat{k}$ in (12), $\Delta_n^2|\hat{k}-k^*| = O_P(1)$, and Assumption 2, $k^*/n \sim \theta$ and $n\Delta_n^2 \to \infty$ as $n \to \infty$. This completes the proof of (40).
The proof of (41) follows using similar argument. $\square$

4.2 Proof of Theorem 2.2

Under the alternative we consider observations $X_1, \ldots, X_n$ with $X_i = G(\xi_i) + \mu$, $i = 1, \ldots, n$. Note that the indicator function $1_{\{x\leq y\}}$ is invariant under strictly increasing functions, i.e. $1_{\{G(\xi_i)\leq G(\xi_j)\}} = 1_{\{\xi_i\leq\xi_j\}}$, if G is strictly increasing. For G being a strictly decreasing function, observe that $1_{\{G(\xi_i)\le G(\xi_j)\}} = 1 - 1_{\{\xi_i\le\xi_j\}}$. Therefore, for G being strictly monotone,

$$\Big|\sum_{i=1}^{k}\sum_{j=k+1}^{n}(1_{\{X_i\le X_j\}}-1/2)\Big| = \Big|\sum_{i=1}^{k}\sum_{j=k+1}^{n}(1_{\{\xi_i\le\xi_j\}}-1/2)\Big|.$$

Thus, to prove Theorem 2.2 it is sufficient to consider $T_{n,1}$ and $T_{n,2}$ in (4), (5) applied to the stationary Gaussian process (ξ_j), i.e. $T_{n,1}(\xi_1,\ldots,\xi_{\hat{k}})$ and $T_{n,2}(\xi_{\hat{k}+1},\ldots,\xi_n)$, instead of $T_{n,1}(X_1,\ldots,X_{\hat{k}})$ and $T_{n,2}(X_{\hat{k}+1},\ldots,X_n)$.

Before we prove that the test M_n tends to infinity in probability under the alternative, we will consider the limit distribution of $T_{n,1}(\xi_1,\ldots,\xi_{\hat{k}})$ and $T_{n,2}(\xi_{\hat{k}+1},\ldots,\xi_n)$ in Lemma 4.7, using a different normalization $n^{d+3/2}c_d$, where $c_d^2 = \frac{c_0}{d(2d+1)}$, $c_0>0$. Note that in the following we always assume $d\in(0,1/2)$. By $(W_H(t))_{0\le t\le 1}$ we denote a fractional Brownian motion process with Hurst parameter $H=d+1/2$, that is a mean zero Gaussian process with auto-covariances $\mathrm{Cov}(W_H(t),W_H(s)) = (t^{2H}+s^{2H}-|t-s|^{2H})/2$.

Lemma 4.6. *Assume that the assumptions of Theorem* 2.2 *hold. Then, for* $0\le s\le t\le 1$,

$$\frac{1}{n^{d+3/2}c_d}\sum_{i=1}^{[ns]}\sum_{j=[nt]+1}^{n}(1_{\{\xi_i\le\xi_j\}}-1/2) \xrightarrow{d} \frac{1}{2\sqrt{\pi}}\Big(s(W_H(1)-W_H(t))-(1-t)W_H(s)\Big),$$

where W_H, $H=d+1/2$ *is a standard fractional Brownian motion,* $c_d^2=\frac{c_0}{d(2d+1)}$, $c_0>0$ *and* $d\in(0,1/2)$.

In the proof of Lemma 4.6 we apply the empirical process non-central limit theorem of Dehling and Taqqu (1989), which uses the Hermite expansion of $1_{\{G(\xi)\le x\}}-F(x)$. Before proceeding to the proof, we will have a brief look at this concept.

Hermite expansion: Since function $g(\xi)=1_{\{G(\xi)\le x\}}-F(x)$ is a measurable function with $\mathrm{E}\,g(\xi)=0$ and $\mathrm{E}\,g^2(\xi)<\infty$, $\xi\sim N(0,1)$, i.e. $g\in L^2(\mathbb{R},N)$, we could represent g by its Hermite expansion

$$g(\xi)=\sum_{i=1}^{\infty}\frac{J_k(x)}{k!}H_k(\xi),$$

where the equality means convergence in the L^2 sense. The k-th order Hermite polynomial is given by

$$H_k(\xi)=(-1)^k e^{\xi^2/2}\frac{d^k}{d\xi^k}e^{-\xi^2/2},$$

and the coefficients are given by $J_k(x)=\mathrm{E}(1_{\{G(\xi)\le x\}}H_k(\xi))$, with $J_1(x)=\mathrm{E}(\xi_1 1_{\{\xi_1\le x\}}) = -\varphi(x)$, where $\varphi(x)$ denotes the standard normal density function. The Hermite rank is defined as $m=\min\{k\ge 0: J_k\ne 0\}$, the smallest k for which the term in the Hermite expansion is not zero. Since $J_1(x)\ne 0$ for some $x\in\mathbb{R}$, we have Hermite rank $m=1$.

Hermite process: The limit process $Z_m(t)$ in Theorem 1.1 of Dehling and Taqqu (1989) is called m-th order Hermite process and is defined e.g. in Taqqu (1978). If $m = 1$, $Z_1(t)$ is the standard Gaussian fractional Brownian motion.

Proof of Lemma 4.6. Dehling *et al.* (2013b) have shown in their Theorem 1 that

$$\left(\frac{1}{n^{d+3/2}c_d}\sum_{i=1}^{[ns]}\sum_{j=[ns]+1}^{n}(1_{\{X_i\le X_j\}}-1/2)\right)_{0\le s\le 1} \xrightarrow{d} \left(\frac{1}{m!}(Z_m(s)-sZ_m(1))\int_{\mathbb{R}}J_m(x)dF(x)\right)_{0\le s\le 1}$$

for $X_i = G(\xi_i)$, where $G : \mathbb{R} \to \mathbb{R}$ is a measurable function (that might not be strictly monotone), F is the continuous distribution of X_i, m is the Hermite rank of the class functions $1_{\{G(\xi_i)\le x\}} - F(x)$, and $J_m(x)$, H_m and $(Z_m(s))_{s\in[0,1]}$ are given above.

Following the proof of Theorem 1 of Dehling *et al.* (2013b) we will show

$$\left(\frac{1}{n^{d+3/2}c_d}\sum_{i=1}^{[ns]}\sum_{j=[nt]+1}^{n}(1_{\{X_i\le X_j\}}-1/2)\right)_{0\le s\le t\le 1} \xrightarrow{d} \left(\frac{1}{m!}\big((1-t)Z_m(s)-s(Z_m(1)-Z_m(t)\big)\int_{\mathbb{R}}J_m(x)dF(x)\right)_{0\le s\le t\le 1}. \quad (43)$$

Since F is a continuous distribution function, $\int_{\mathbb{R}} F(x)dF(x) = 1/2$. Denote $F_k(x) = \frac{1}{k}\sum_{i=1}^{k} 1_{\{X_i\le x\}}$ and $F_{k+1,n}(x) = \frac{1}{n-k}\sum_{i=k+1}^{n} 1_{\{X_i\le x\}}$. Then,

$$\sum_{i=1}^{[ns]}\sum_{j=[nt]+1}^{n}(1_{\{X_i\le X_j\}}-1/2) = [ns](n-[nt])\Big(\int_{\mathbb{R}}\big(F_{[ns]}(x)-F(x)\big)dF_{[nt]+1,n}(x)\Big) + [ns](n-[nt])\Big(\int_{\mathbb{R}}F(x)d\big(F_{[nt]+1,n}-F\big)(x)\Big).$$

Integration by parts yields,

$$\int_{\mathbb{R}}F(x)d\big(F_{[nt]+1,n}-F\big)(x) = -\int_{\mathbb{R}}\big(F_{[nt]+1,n}-F\big)(x)dF(x).$$

Hence,

$$\sum_{i=1}^{[ns]}\sum_{j=[nt]+1}^{n}(1_{\{X_i\le X_j\}}-1/2) = [ns](n-[nt])\int_{\mathbb{R}}(F_{[ns]}(x)-F(x))dF_{[nt]+1,n}(x) - [ns](n-[nt])\int_{\mathbb{R}}(F_{[nt]+1,n}(x)-F(x))dF(x).$$

With the same argument as used in Dehling *et al.* (2013b), we show that

$$\frac{[ns](n-[nt])}{n^{d+3/2}c_d}\int_{\mathbb{R}}(F_{[ns]}(x)-F(x))dF_{[nt]+1,n}(x)-\frac{(1-t)}{m!}\int_{\mathbb{R}}J_m(x)Z_m(s)dF(x)\to 0 \tag{44}$$

$$\frac{[ns](n-[nt])}{n^{d+3/2}c_d}\int_{\mathbb{R}}(F_{[nt]+1,n}(x)-F(x))dF(x)-\frac{s}{m!}\int_{\mathbb{R}}J_m(x)(Z_m(1)-Z_m(t))dF(x)\to 0, \tag{45}$$

almost surely, uniformly in $0<s\leq t<1$.

Let us start with (44). We can write

$$\begin{aligned}
&\frac{[ns](n-[nt])}{n^{d+3/2}c_d}\int_{\mathbb{R}}(F_{[ns]}(x)-F(x))dF_{[nt]+1,n}(x)-\frac{(1-t)}{m!}\int_{\mathbb{R}}J_m(x)Z_m(s)dF(x)\\
&=\frac{(n-[nt])}{n}\int_{\mathbb{R}}\frac{[ns]}{n^{d+1/2}c_d}(F_{[ns]}(x)-F(x))dF_{[nt]+1,n}(x)-(1-t)\int_{\mathbb{R}}J_m(x)\frac{Z_m(s)}{m!}dF(x)\\
&=\frac{(n-[nt])}{n}\int_{\mathbb{R}}\Big(\frac{[ns]}{n^{d+1/2}c_d}(F_{[ns]}(x)-F(x))-J_m(x)\frac{Z_m(s)}{m!}\Big)dF_{[nt]+1,n}(x)\\
&+\frac{(n-[nt])}{n}\int_{\mathbb{R}}J_m(x)\frac{Z_m(s)}{m!}d\big(F_{[nt]+1,n}-F\big)(x)\\
&+\Big(\frac{(n-[nt])}{n}-(1-t)\Big)\int_{\mathbb{R}}J_m(x)\frac{Z_m(s)}{m!}dF(x).
\end{aligned} \tag{46}$$

The empirical process non-central limit theorem of Dehling and Taqqu (1989) yields

$$\Big(d_n^{-1}[ns]\big(F_{[ns]}(x)-F(x)\big)\Big)_{x\in[-\infty,\infty],s\in[0,1]}\xrightarrow{d}\Big(J(x)Z(s)\Big)_{x\in[-\infty,\infty],s\in[0,1]},$$

where $J(x)=J_m(x)$, $Z(x)=Z_m(x)/m!$ and $d_n^2\sim n^{2d+1}c_d^2$.
Dehling *et al.* (2013b) argue that applying the Skorohod-Dudley-Wichura representation yields almost sure convergence, i.e.

$$\sup_{s,x}\left|d_n^{-1}[ns]\big(F_{[ns]}(x)-F(x)\big)-J(x)Z(x)\right|\to 0\qquad\text{a.s.} \tag{47}$$

Thus, the first term on the right-hand side of (46) converges to 0 almost surely, uniformly in $0<s\leq t<1$.
Furthermore, we note that

$$\begin{aligned}
&\frac{(n-[nt])}{n}\int_{\mathbb{R}}J(x)Z(s)d\big(F_{[nt]+1,n}-F\big)(x)\\
&\qquad=Z(s)\Big[\frac{(n-[nt])}{n}\int_{\mathbb{R}}J(x)dF_{[nt]+1,n}(x)-\frac{(n-[nt])}{n}\int_{\mathbb{R}}J(x)dF(x)\Big]\\
&\qquad=Z(s)\Big[\frac{1}{n}\sum_{i=[nt]+1}^{n}J(X_i)-\frac{(n-[nt])}{n}\operatorname{E}(J(X_i))\Big]\\
&\qquad=Z(s)\frac{1}{n}\sum_{i=1}^{n}\big(J(X_i)-\operatorname{E}(J(X_i))\big)-Z(s)\frac{1}{n}\sum_{i=1}^{[nt]}\big(J(X_i)-\operatorname{E}(J(X_i))\big).
\end{aligned}$$

By the ergodic theorem, $\frac{1}{n}\sum_{i=1}^{[nt]}\big(J(X_i)-\mathrm{E}(J(X_i))\big)\to 0$ almost surely for all $0\le t\le 1$. Therefore, the second term on the right-hand side of (46) converges to 0 almost surely, uniformly in $0<s\le t<1$.
Also the third term on the right-hand side of (46) converges to 0, since, as $n\to\infty$, $\big((n-[nt])/n-(1-t)\big)\to 0$, and $\int_{\mathbb{R}} J_m(x)\frac{Z_m(s)}{m!}dF(x)$ is bounded. This finishes the proof of (44).
Note that

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

and hence,

$$(n-[nt])\big(F_{[nt]+1,n}(x)-F(x)\big)=n\big(F_n(x)-F(x)\big)-[nt]\big(F_{[nt]}(x)-F(x)\big).$$

Then the proof of (45) follows using again (47). Thus, (43) is shown.
Note that this result holds for $X_i=G(\xi_i)$, but in our lemma we consider $X_i=\xi_i$, where (ξ_j) is a stationary mean zero Gaussian process with auto-covariances $\gamma_k\sim k^{2d-1}c_0$, $d\in(0,1/2)$. In this case, $J_1(x)=-\varphi(x)$, where $\varphi(x)$ denotes the standard normal density function and $\int_{\mathbb{R}} J_1(x)dF(x)=-\frac{1}{2\sqrt{\pi}}$, since F is the normal distribution function. Furthermore, $J_1(x)\neq 0$ for all x and hence, we have Hermite rank $m=1$. Therefore, $(Z_1(s))$ denotes the standard fractional Brownian motion process $(W_H(s))$. Thus, the limit in (43) equals

$$\frac{1}{2\sqrt{\pi}}\Big(s(W_H(1)-W_H(t))-(1-t)W_H(s)\Big),$$

which proves the lemma. □

Lemma 4.7. *Assume that the assumptions of Theorem* 2.2 *hold. Then,*

$$\Bigg[\frac{1}{n^{d+3/2}c_d}\max_{1\le k\le\hat{k}}\Big|\sum_{i=1}^{k}\sum_{j=k+1}^{\hat{k}}(1_{\{\xi_i\le\xi_j\}}-1/2)\Big|,\frac{1}{n^{d+3/2}c_d}\max_{\hat{k}<k\le n}\Big|\sum_{i=\hat{k}+1}^{k}\sum_{j=k+1}^{n}(1_{\{\xi_i\le\xi_j\}}-1/2)\Big|\Bigg]$$

$$\xrightarrow{d}\Bigg[\frac{\zeta}{2\sqrt{\pi}}\sup_{0\le t\le\zeta}\big|W_H(t)-\frac{t}{\zeta}W_H(\zeta)\big|,$$
$$\frac{1-\zeta}{2\sqrt{\pi}}\sup_{\zeta\le t\le 1}\Big|W_H(t)-W_H(\zeta)-\frac{t-\zeta}{1-\zeta}(W_H(1)-W_H(\zeta))\Big|\Bigg],$$

where $c_d^2=\frac{c_0}{d(2d+1)}$, $c_0>0$, $d\in(0,1/2)$, W_H *is a standard fractional Brownian motion,* $H=d+1/2$ *and*

$$\zeta=\inf\Big\{t\ge 0:\sup_{0\le s\le 1}|W_H(s)-sW_H(1)|=|W_H(t)-tW_H(1)|\Big\}. \qquad (48)$$

Proof. Denote for $0 \le s \le t \le 1$

$$\tilde{U}_n(s,t) = \frac{1}{n^{d+3/2}c_d} \sum_{i=1}^{[ns]} \sum_{j=[nt]+1}^{n} (1_{\{\xi_i \le \xi_j\}} - 1/2),$$

$$\tilde{W}_H(s,t) = -\frac{1}{2\sqrt{\pi}}\big((1-t)W_H(s) - s(W_H(1) - W_H(t))\big)$$

and note that by Lemma 4.6, $(\tilde{U}_n(s,t))_{s,t} \xrightarrow{d} (\tilde{W}_H(s,t))_{s,t}$. Furthermore, we denote

$$\tilde{U}_{n,1}(t) = \frac{1}{n^{d+3/2}c_d} \max_{1\le k \le nt} \Big| \sum_{i=1}^{k} \sum_{j=k+1}^{nt} (1_{\{\xi_i \le \xi_j\}} - 1/2)\Big|,$$

$$\tilde{U}_{n,2}(t) = \frac{1}{n^{d+3/2}c_d} \max_{nt < k \le n} \Big| \sum_{i=nt+1}^{k} \sum_{j=k+1}^{n} (1_{\{\xi_i \le \xi_j\}} - 1/2)\Big|,$$

$$\tilde{W}_{H,1}(t) = \frac{t}{2\sqrt{\pi}} \sup_{0\le s\le t} \big|W_H(s) - \frac{s}{t}W_H(t)\big|,$$

$$\tilde{W}_{H,2}(t) = \frac{1-t}{2\sqrt{\pi}} \sup_{t\le s\le 1} \big|(W_H(s) - W_H(t)) - \frac{1-s}{1-t}(W_H(1) - W_H(t))\big|.$$

Since

$$\sum_{i=1}^{k} \sum_{j=k+1}^{nt} (1_{\{\xi_i \le \xi_j\}} - 1/2) = \sum_{i=1}^{k} \sum_{j=k+1}^{n} (1_{\{\xi_i \le \xi_j\}} - 1/2) - \sum_{i=1}^{k} \sum_{j=nt+1}^{n} (1_{\{\xi_i \le \xi_j\}} - 1/2),$$

we can write $\tilde{U}_{n,1}(t) = \sup_{0\le s\le t} |\tilde{U}_n(s,s) - \tilde{U}_n(s,t)|$ and with a similar argument $\tilde{U}_{n,2}(t) = \sup_{t\le s\le 1} |\tilde{U}_n(s,s) - \tilde{U}_n(t,s)|$. Note that $\tilde{W}_{H,1}(t) = \sup_{0\le s\le t} |\tilde{W}_H(s,s) - \tilde{W}_H(s,t)|$ and $\tilde{W}_{H,2}(t) = \sup_{t\le s\le 1} |\tilde{W}_H(s,s) - \tilde{W}_H(t,s)|$. Thus, the same continuous mapping transforms $\tilde{U}_n(s,t)$ into the vector $(\hat{k}/n, \tilde{U}_{n,1}(t), \tilde{U}_{n,2}(t))$ and $\tilde{W}_H(s,t)$ into $(\zeta, \tilde{W}_{H,1}(t), \tilde{W}_{H,2}(t))$, where ζ is given in (48). Hence, by the continuous mapping theorem and Lemma 4.6

$$\Big(\hat{k}/n, \tilde{U}_{n,1}(t), \tilde{U}_{n,2}(t)\Big) \xrightarrow{d} \Big(\zeta, \tilde{W}_{H,1}(t), \tilde{W}_{H,2}(t)\Big).$$

Applying the mapping $(z, x(t), y(t)) \mapsto (x(z), y(z))$ to both vectors finishes the proof. □

Proof of Theorem 2.2. By Lemma 4.7,

$$T_{n,1} = \hat{k}^{-3/2} \max_{1\le k\le \hat{k}} \Big| \sum_{i=1}^{k} \sum_{j=k+1}^{\hat{k}} (1_{\{\xi_i \le \xi_j\}} - 1/2)\Big|$$

$$= \frac{n^{d+3/2}c_d}{\hat{k}^{3/2}} \frac{1}{n^{d+3/2}c_d} \max_{1\le k\le \hat{k}} \Big| \sum_{i=1}^{k} \sum_{j=k+1}^{\hat{k}} (1_{\{\xi_i \le \xi_j\}} - 1/2)\Big| = \frac{n^{d+3/2}c_d}{\hat{k}^{3/2}} O_P(1).$$

Similar argument yields $T_{n,2} = \frac{n^{d+3/2}c_d}{(n-\hat{k})^{3/2}} O_P(1)$. Thus, to prove Theorem 2.2 it remains to show $\frac{n^{d+3/2}c_d}{\hat{k}^{3/2}} \rightarrow_p \infty$ and $\frac{n^{d+3/2}c_d}{(n-\hat{k})^{3/2}} \rightarrow_p \infty$. The proof of Lemma 4.7 yields $\hat{k}/n \xrightarrow{d} \zeta$, where ζ is given in (48), and hence, $(n/\hat{k})^{3/2} = O_P(1)$ and $(n/(n-\hat{k}))^{3/2} = O_P(1)$. Since $d > 0$, $n^d \rightarrow \infty$ as $n \rightarrow \infty$. Thus, $T_{n,1} \rightarrow_p \infty$ and $T_{n,2} \rightarrow_p \infty$. This finishes the proof of Theorem 2.2. □

Acknowledgement

The author would like to thank Herold Dehling, Liudas Giraitis and Isabel Garcia for valuable discussions. The research was supported by the Collaborative Research Centre 823 *Statistical modelling of nonlinear dynamic processes* and the Konrad-Adenauer-Stiftung.

References

Baek, C. and Pipiras, V. (2012). Statistical tests for a single change in mean against long-range dependence. *J. Time Series Anal.* **33** 131-151.

Berkes, I., Horváth, L., Kokoszka, P., and Shao, Q. (2006). On discriminating between long-range dependence and changes in mean. *Ann. Statist.* **34** 1140-1165.

Bhattacharya, R., Gupta, V., and Waymire, E. (1983). The Hurst effect under trends. *J. Appl. Probab.* **20** 649-662.

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York.

Borovkova, S., Burton, R. and Dehling, H. (2001). Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Trans. Amer. Math. Soc.* **353** 4261-4318.

Bradley, R.C. (2007). *Introduction to Strong Mixing Conditions.* Kendrick Press, Heber City.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.* **14** 1171-1179.

Dehling, H., Fried, R., Garcia Arboleda, I. and Wendler, M. (2015). Change-point detection under dependence based on two-sample U-statistics. In: Dawson, D., Kulik, R., Jaye, M. O., Szyszkowicz, B., Zhao, Y.(Eds.) *Asymptotic laws and methods in stochastics. Fields Institute Communication* **76** 195-220.

Dehling, H., Fried, R., Sharipov, O., Vogel, D. and Wornowizki, M. (2013a). Estimation of the variance of partial sums of dependent processes. *Statist. Probab. Lett.* **83** 141-147.

Dehling, H., Rooch, A. and Taqqu, M. S. (2013b). Non-parametric change-point tests for long-range dependent data. *Scand. J. Stat.* **40** 153-173.

Dehling, H., and Taqqu, M. S. (1989). The Empirical Process of some Long-Range Dependent Sequences with an Application to U-Statistics. *Ann. Statist.* **17** 1767-1783.

Dürre, A., Fried, R. and Liboschik, T. (2015). Robust estimation of (partial) autocorrelation. *Wiley Interdiscip. Rev. Comput. Stat.* **7** 205-222.

Gerstenberger, C. (2018). Robust Wilcoxon-Type Estimation of Change-Point Location Under Short-Range Dependence. *J. Time Series Anal.* **1** 90-104

Giraitis, L., Kokoszka, P. and Leipus, R. (2001). Testing for long memory in the presence of a general trend. *J. Appl. Probab.* **38** 1033-1054.

Granger, C. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *J. Empirical Finance* **11** 399-421.

Hansen, B. E. (1991). GARCH(1,1) processes are near epoch dependent. *Econom. Lett.* **36** 181-186.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293-325.

Hurst, H. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.* **116** 770-808.

Künsch, H. (1986). Discrimination between monotonic trends and long-range dependence. *J. Appl. Probab.* **4** 1025-1030.

Ma, Y. and Genton, M. (2000). Highly robust estimation of the autocovariance function. *J. Time Series Anal.* **21** 663-684.

Mandelbrot, B. and Van Ness, J. (1968). Fractional Brownian motions, fractional noises and applications. *Soc. Ind. Appl. Math.* **10** 422-437.

Mandelbrot, B. and Wallis, J. (1968). Noah, Joseph, and Operational Hydrology. *WaterResour.Res* **4** 909-918.

Preuß, P., Sen, K. and Dette, H. (2017). Detecting long-range dependence in non-stationary time series. *Electron. J. Stat.* **1** 1600-1659

Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc.* **88** 1273-1283.

Shorack, G.R. and Wellner, J.A. (2009). *Empirical Processes with Applications to Statistics*, Wiley, New York.

Taqqu, M. (1978). A representation for self-similar processes. *Stochastic Process. Appl.* **7** 55-64.

Wooldridge, J. M. and White, H. (1988). Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* **4** 210–230.