Gerstenberger, Carina



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

Journal of Time Series Analysis

**Provided in Cooperation with:**
John Wiley & Sons

*Suggested Citation:* Gerstenberger, Carina (2021) : Robust discrimination between long-range dependence and a change in mean, Journal of Time Series Analysis, ISSN 1467-9892, John Wiley & Sons, Ltd, Oxford, UK, Vol. 42, Iss. 1, pp. 34-62, https://doi.org/10.1111/jtsa.12554

JOURNAL OF TIME SERIES ANALYSIS
*J. Time Ser. Anal.* **42**: 34–62 (2021)
Published online 26 August 2020 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12554

ORIGINAL ARTICLE

# ROBUST DISCRIMINATION BETWEEN LONG-RANGE DEPENDENCE AND A CHANGE IN MEAN

CARINA GERSTENBERGER[a]

[a] *Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, Germany*

In this article 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 I., Horváth L., Kokoszka P. and Shao Q. 2006. *Ann.Statist*. 34:1140–1165), 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. We also apply both testing procedure to hydrologic data.

*Received 04 April 2018; Accepted 04 July 2020*

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

MOS subject classification: 62M10; 62F03; 62F05; 62G35.

## 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, for example in hydrology, geophysics and economics. A lively debate also rages over the observed Hurst effect is due to long-range dependence or non-stationarity. 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 articles, see, for example, Granger and Hyung (2004), Diebold and Inoue (2001), Chang and Perron (2016), Iacone *et al.* (2019). Also the problem of detecting change points if the data might be long-range dependent has been widely discussed, see among others Lavielle and Moulines (2000), Iacone *et al.* (2013a), Iacone *et al.* (2013b), McCloskey and Perron (2013) and Iacone *et al.* (2017).

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 periodogram in these two cases behaves differently. Another test against spurious long-range dependence was given by Qu (2011). A test allowing to distinguish between a stationary long-range dependent process and short-range dependent process with a change in mean was

* Correspondence to: Carina Gerstenberger, Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany.
E-mail: carina.gerstenberger@rub.de

introduced by Berkes *et al.* (2006) and is based on the CUSUM statistic

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

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

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

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 articles 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 article is based on the idea of Berkes *et al.* (2006). First, given a sample $X_1, \dots, X_n$, 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 $X_1, \dots, X_{\hat{k}}$ and $X_{\hat{k}+1}, \dots, X_n$.

### 1.1. Wilcoxon-type Testing Procedure

Assuming that sample $X_1, \dots, X_n$ is given, we want to test the hypothesis
$\boldsymbol{H_0}$: $X_i = Y_i + \mu_i$, $i = 1, \dots, n$ is generated by a stationary zero mean short-range dependent process $(Y_j)$ and has a change in mean $\mu_1 = \cdots = \mu_{k^*} \neq \mu_{k^*+1} = \cdots = \mu_n$ at unknown time $k^*$,
against the alternative
$\boldsymbol{H_1}$: $X_1, \dots, X_n$ is a sample from a stationary long-range dependent process.

Note that during the article stationary means strictly stationary.

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

$$\hat{k} = \min \left\{ k : \max_{1 \le l < n} \left| W_{1,n}(l) \right| = \left| W_{1,n}(k) \right| \right\}, \tag{3}$$

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

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

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

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

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

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

Finally, we define the test statistic

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




DOI: 10.1111/jtsa.12554 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

We show that $T(X_1, \dots, X_n)$ 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 $X_1, \dots, X_{\hat{k}}$ and $X_{\hat{k}+1}, \dots, X_n$ are samples from a stationary sequence with a constant mean. Subsequently, $M_n$ can be used to test if the samples $X_1, \dots, X_{\hat{k}}$ and $X_{\hat{k}+1}, \dots, X_n$ have been generated by a short-range or long-range dependent stationary process.

The outline of the article 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. An application to hydrologic data is given in Section 4. All proofs are given in Section 5.

## 2. DEFINITIONS, ASSUMPTIONS AND MAIN RESULTS

We present main assumptions, definitions and main results.

Throughout the article, $C$ 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 $\stackrel{d}{=}$ we denote equality in distribution. $\|g\|_\infty = \sup_x |g(x)|$ denotes the supremum norm of a function $g$.

### 2.1. Null Hypothesis: Short-Range Dependence with a Change in Mean

Under the null hypothesis we assume the random variables $X_1, \dots, X_n$ 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, $\Delta_n$ denotes the unknown magnitude of change (see Assumption 2) and $(Y_j)$ is a zero-mean strictly stationary short-range dependent process.

To cover a wide range of processes, we assume that the underlying process $(Y_j)$ can be written as $Y_j = f(Z_j, Z_{j-1}, Z_{j-2}, \dots), j \in \mathbb{Z}$, 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 $\beta$-mixing) if

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

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

Absolute regularity or $\beta$-mixing implies the weaker property of $\alpha$-mixing, see for example, Bradley (2007).

In addition, we will assume that $(Y_j)$ satisfies near epoch dependence condition, that is, $Y_j$ 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 $a_k$, $k \ge 0$, if

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

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

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

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, $r \geq 1$. The concept of $L_1$ near epoch dependence only assumes existence of the first moment $\mathrm{E}\,|Y_1|$. Therefore, we can allow heavy-tailed distributions.

We need further additional assumptions on the distribution function $F$ of $Y_1$, the mixing coefficients $\beta_k$ 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 $\beta_k$ 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 $\Delta_n$ 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 $\Delta_n$ 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,

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

### 2.2. 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 $\mu$ is the unknown mean and $(\xi_j)$ 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 $\mathrm{E}(G(\xi_1)) = 0$.

### 2.3. 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.

  
DOI: 10.1111/jtsa.12554 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

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

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

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 $\sigma^2$, 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

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

where

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

is based on the CUSUM statistic $C_{1,n}(k)$ in (1). $\tilde{k}_C = \min\left\{k : \max_{1\le l\le n}\left|C_{1,n}(l)\right| = \left|C_{1,n}(k)\right|\right\}$ is a CUSUM-type estimator of $k^*$ and $\hat{s}_n^2$ is a long-run variance estimator of $\sigma_c^2 = \sum_{k=-\infty}^{\infty} \mathrm{Cov}\left(Y_0, Y_k\right)$ 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 $M_n$, that is, 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 $n \to \infty$,

$$M_n \to_p \infty.$$

Under the alternative in (13) we do not consider the long memory Gaussian process itself, but a function of it. This concept also allows non-Gaussianity. We restrict the result of Theorem 2.2 to strictly monotone functions due to simplicity of the proof. However, the result can also be expanded to more general functions $G(\cdot)$. In this case the dependence structure of $(G(\xi_i))$ is in general not clear. Proposition 1.2 of Rooch (2012) yields that under slight assumptions if $\gamma_k \sim c_0 k^{2d-1}$, $c_0 > 0$, $d \in (0, 1/2)$ then $\mathrm{Cov}(G(\xi_i), G(\xi_{i+k})) \sim (c_0/m!)k^{(2d-1)m}$, where $m$ is the Hermite rank of $G$ (see Section 5.2 for more details about Hermite rank). Therefore, for $-1 < (2d-1)m < 0$, the process $(G(\xi_i))$ is still long-range dependent.

Proofs of Theorems 2.1 and 2.2 are given in Section 5.

## 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).

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

### 3.1. Simulation Set Up

To calculate the *empirical size* we generate the sample of random variables $X_1, \dots, X_n$ 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 $\rho = 0.4$. The innovations $\epsilon_i$ are generated from a standard normal distribution and a Student's $t$-distribution with $\nu = 1$ degree of freedom. We set $k^* = [n\theta]$, $\theta = 0.25, 0.5, 0.75$ and $\Delta = 0.5, 1, 2$.

Note that $t_1$-distributed innovations do not satisfy the $L_1$ NED condition, since $L_1$ NED requires the existence of $\mathrm{E}\,|Y_1|$. However, $t_1$-distributed innovations are included in the simulation study, since it proofs the functionality of Wilcoxon-type testing procedure even in the case of extremely heavy tails.

To evaluate the *empirical power* of the test we generate a sample $X_1, \dots, X_n$ 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 for example, 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 $(X_1, \dots, X_n)$ (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.

### 3.2. 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 $\hat{\sigma}^2(X_1, \dots, X_k)$ is a consistent estimator for the long-run variance $\sigma^2$ based on the sample $X_1, \dots, X_k$, then

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

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 $c_\alpha$ for a given significance level $\alpha$ 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

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




DOI: 10.1111/jtsa.12554 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

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 $c_{5\%} = 1.478$.

### 3.3. 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 $q(n) = C\log_{10}(n)$. Table I 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 $15\log_{10}(n)$. It shows that $\tilde{M}_{C,n}$ with Bartlett estimator $\hat{s}_n^2$ is too conservative and has low power against the alternative, which has also been pointed out by Baek and Pipiras (2012) and Preuß *et al.* (2017).

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

| $n =$ | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|
| Emp. size | 0.05 | 0.87 | 2.48 | 3.79 |
| Power | 0.30 | 7.62 | 27.44 | 60.51 |

In our simulation study to improve the performance of $\tilde{M}_{C,n}$ test we proceed as follows. To estimate $\sigma_C^2$, instead of $\hat{s}_n^2$, we use the non-overlapping subsampling estimator of $\sigma_C^2$ by Carlstein (1986), with block length $l_n$,

$$\hat{\sigma}_C^2 = \frac{1}{\lceil n/l_n\rceil}\sum_{i=1}^{\lceil n/l_n\rceil}\frac{1}{l_n}\left(\sum_{j=(i-1)l_n+1}^{il_n} X_j - \frac{l_n}{n}\sum_{j=1}^{n} X_j\right)^2, \tag{22}$$

which yields better size and power balance for $\tilde{M}_{C,n}$, as seen from Tables II and IV. 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 $\sigma$ we shall use the Carlstein type estimator for long-run variance proposed by Dehling *et al.* (2013a),

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

where $F_n(x) = n^{-1}\sum_{i=1}^{n} 1_{\{X_i\le x\}}$. Note that $\hat{\sigma}_W$ estimates $\sigma$, not $\sigma^2$.

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

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

where $\rho$ denotes the autocorrelation coefficient at lag 1. In our simulation study we use this block length with $\rho$ estimated by the sample autocorrelation coefficient $\hat{\rho}$ since it yields good results for the empirical size and power.

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

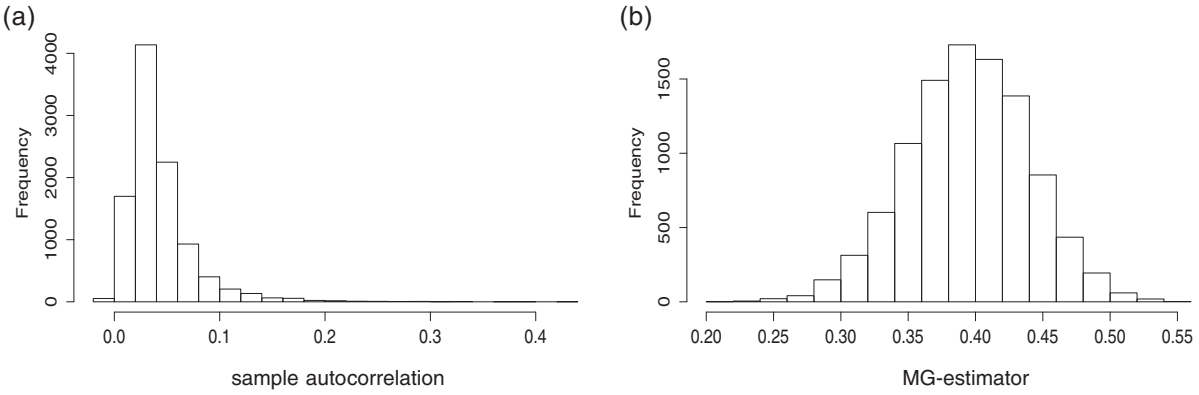



Figure 1. Histogram of $\hat{\rho}(1)$ and $\hat{\rho}_Q(1)$ based on 10,000 replications. $X_i$ is generated by an AR(1) process with outliers, $\epsilon_i \sim \mathrm{N}(0,1)$, $\rho = 0.4$ and $n = 500$

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 $\rho$ 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 $Q_n(x) = 2.21914\{|X_i - X_j|; i < j\}_{(k)}$, $x = (X_1, \ldots, X_n)$, 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 (1993), $u = (X_1, \ldots, X_{n-1})$ and $v = (X_2, \ldots, X_n)$. Figure 1 contains the histogram of estimates $\hat{\rho}$ and $\hat{\rho}_Q$ based on 10,000 replications of sample $X_1, \ldots, X_{500}$ 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).

### 3.4. Simulation Results

Table II 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, that is, for $\theta = 0.25, 0.75$. We also consider the situation of no change, that is, $\Delta = 0$, 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.

Note that in Table II both tests do not tend to 5% as it is expected. This is due to a very slow convergence to the limit process. In simulation studies with really large sample size $n > 10{,}000$ the empirical size of both tests is tending to 5%. Since $\tilde{M}_{C,n}$ and $\hat{M}_n$ are both suffering from this slow convergence, they are still comparable to each other.

Table III reports the empirical size of $\hat{M}_n$ and $\tilde{M}_{C,n}$ in presence of outliers and $t_1$-distributed innovations. While test $\hat{M}_n$ is robust to the outliers and just slightly affected by the heavy-tailed innovations, the test $\tilde{M}_{C,n}$ becomes much too conservative.

Tables IV and V 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 IV 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.

  
DOI: 10.1111/jtsa.12554 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Table II. 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 and $\epsilon_i \sim \mathrm{N}(0,1)$

| $\theta =$ | 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 III. 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 $\epsilon_i \sim \mathrm{N}(0,1)$ without and with outliers, and $\epsilon_i \sim t_1$. We consider $\Delta = 1$ and $\theta = 0.5$

| | $\epsilon_i \sim \mathrm{N}(0,1)$ | | $\epsilon_i \sim t_1$ | | $\epsilon_i \sim \mathrm{N}(0,1)$ with outliers | |
|---|---|---|---|---|---|---|
| | $\tilde{M}_{C,n}$ | $\hat{M}_n$ | $\tilde{M}_{C,n}$ | $\hat{M}_n$ | $\tilde{M}_{C,n}$ | $\hat{M}_n$ |
| $n =$ | | | | | | |
| 1000 | 5.11 | 4.68 | 0.83 | 2.92 | 0.56 | 4.82 |
| 2000 | 5.96 | 5.23 | 1.22 | 3.74 | 1.17 | 5.56 |
| 5000 | 6.26 | 5.59 | 1.03 | 4.57 | 2.28 | 5.41 |

Table IV. 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 | |
|---|---|---|---|---|---|---|---|---|
| | $\tilde{M}_{C,n}$ | $\hat{M}_n$ | $\tilde{M}_{C,n}$ | $\hat{M}_n$ | $\tilde{M}_{C,n}$ | $\hat{M}_n$ | $\tilde{M}_{C,n}$ | $\hat{M}_n$ |
| $n =$ | | | | | | | | |
| 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 V 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.

Let us have a closer look on what happens in the case of outliers. There are different steps in the testing procedures that might be affected by outliers: the estimation of the time of change, the estimation of the long-run variance and the test statistic itself. The impact of outliers on a CUSUM and Wilcoxon based change-point estimator has already been discussed in Gerstenberger (2018). It is shown that the Wilcoxon-type estimator is nearly not affected by outliers whereas the CUSUM-type estimator has trouble in detecting the correct time of change. Therefore, if this would be the only problem in the CUSUM-type testing procedure, we should expect $\tilde{M}_{C,n}$ to reject the hypothesis more often due to splitting the data at the spuriously estimated change-point. However as we have seen in Table III this is not the case. Let us now have a closer look at the CUSUM statistic $C_{1,n}(k)$ and the Wilcoxon statistic $W_{1,n}(k)$. We generated a series of random variables $Y_1, \dots, Y_n$, $n = 1000$ following the AR(1) process given in (17), but without a change in mean. In Figure 2 the solid line shows in (a) $n^{-1/2}|C_{1,n}(k)|$, $k = 1, \dots, 1000$ and in (b) $n^{-3/2}|W_{1,n}(k)|$, $k = 1, \dots, 1000$, both applied to $Y_1, \dots, Y_{1000}$. Then we disturbed the same variables $Y_1, \dots, Y_n$ with outliers as described above. The dashed lines in both figures show the results for $n^{-1/2}|C_{1,n}(k)|$ and $n^{-3/2}|W_{1,n}(k)|$ applied to the variables including outliers. We see again that the Wilcoxon statistic is not affected

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

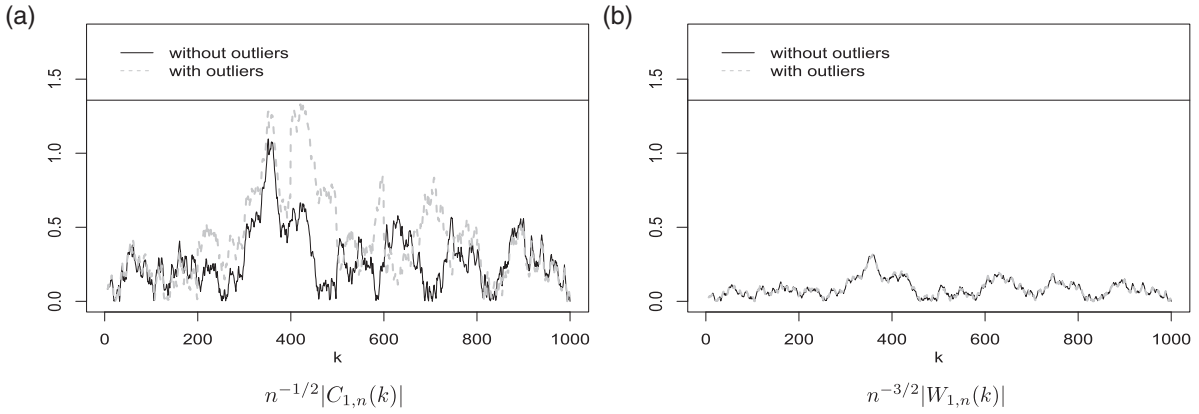



Figure 2. Values of $n^{-1/2}|C_{1,n}(k)|$ and $n^{-3/2}|W_{1,n}(k)|$ for $k = 1, \dots, 1000$. $Y_i = \rho Y_{i-1} + \epsilon_i$ is an AR(1) process with $\rho = 0.4$ and standard normal innovations $\epsilon_i$. For the dashed lines ($Y$) is disturbed by outliers

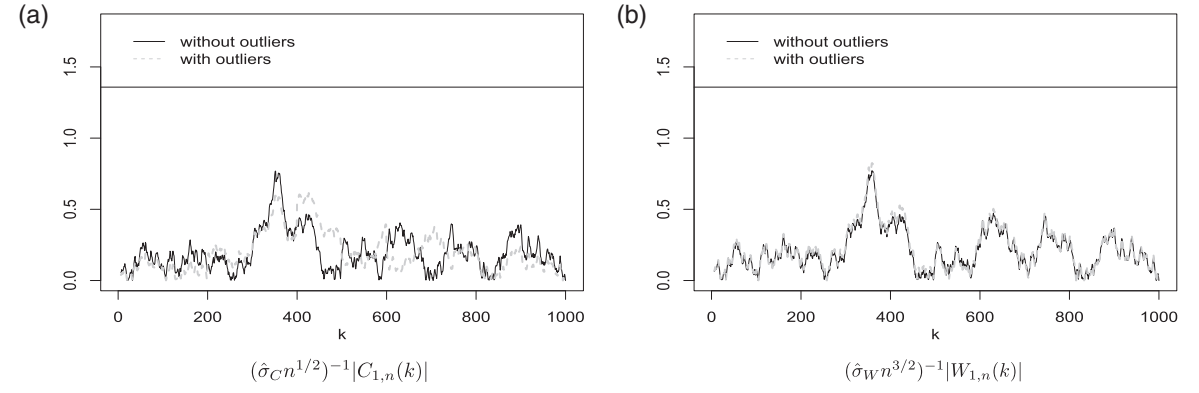



Figure 3. Values of $(\hat{\sigma}_C n^{1/2})^{-1}|C_{1,n}(k)|$ and $(\hat{\sigma}_W n^{3/2})^{-1}|W_{1,n}(k)|$ for $k = 1, \dots, 1000$. $Y_i = \rho Y_{i-1} + \epsilon_i$ is an AR(1) process with $\rho = 0.4$ and standard normal innovations $\epsilon_i$. For the dashed lines ($Y$) is disturbed by outliers

by the outliers. However as expected, the CUSUM statistic has larger values in the outlier scenario and therefore it has a larger maximum. However again, this should lead to a more often rejection of the hypothesis. So why do the simulation results show more conservatism for the CUSUM-type testing procedure in the outlier scenario? This is due to the long-run variance estimation. If we have a look at the value for the estimator given in (22) applied to the example we see that the value for the data with outliers ($\hat{\sigma}_C^2 = 4.63$) is much higher than the value for data without outliers ($\hat{\sigma}_C^2 = 2.04$). This reduces the values for the CUSUM-testing procedure for outlier scenario, since we divide by the estimate of the long-run variance, see Figure 3(a). This leads to reduction of size and a loss in power. For the Wilcoxon-type testing procedure we can observe that the value of $\hat{\sigma}_W$ given in (23) is in both cases nearly the same ($\hat{\sigma}_W = 0.38$ with outliers and $\hat{\sigma}_W = 0.41$ without), see Figure 3(b).

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.

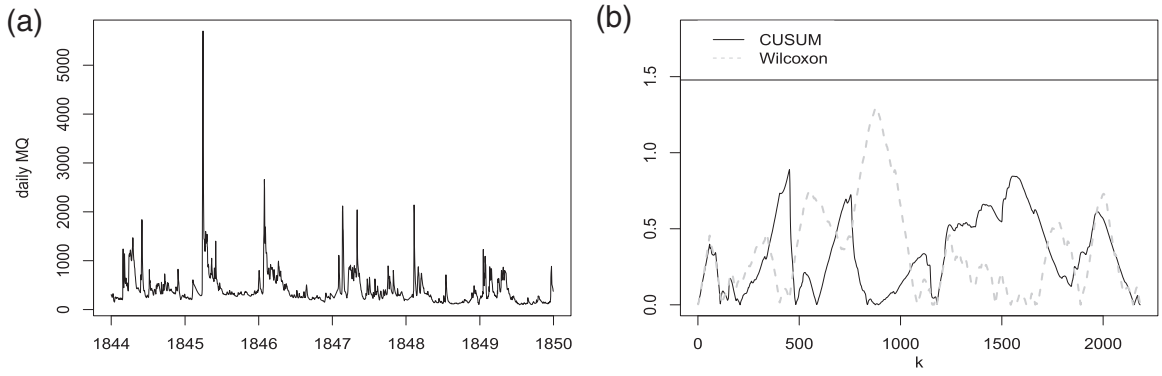
## 4. DATA EXAMPLE

In the following data example we consider a hydrologic time series. In particular, we consider the mean daily discharges (MQ) of the river Elbe in Dresden, Germany. The data cover the time from 01.01.1844 to 31.12.1849 ($n = 2191$) and are shown in Figure 4(a). It is well known that daily MQ are strongly correlated, see Figure 5

  
DOI: 10.1111/jtsa.12554 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Table V. 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

| $d =$ | 0.1 | | 0.2 | | 0.3 | | 0.4 | |
|---|---|---|---|---|---|---|---|---|
| | $\tilde{M}_{C,n}$ | $\hat{M}_n$ | $\tilde{M}_{C,n}$ | $\hat{M}_n$ | $\tilde{M}_{C,n}$ | $\hat{M}_n$ | $\tilde{M}_{C,n}$ | $\hat{M}_n$ |
| $n =$ | | | | | | | | |
| 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 |




Daily MQ of the river Elbe in Dresden, Germany.

Values for CUSUM and Wilcoxon testing procedures.

Figure 4. Mean daily discharge (MQ) of the river Elbe in Dresden, Germany, from 1844 to 1849 (a). In (b) we see the corresponding pointwise values for the CUSUM and Wilcoxon type testing procedure

for the sample autocorrelation function. Hence, testing for dependency should result in long-range dependence. In the year 1845 there was a big flood in Dresden, which appears in Figure 4(a) as an outlier. The time series also contains some smaller outliers after 1845.

We calculated the CUSUM testing procedure $\tilde{M}_{C,n}$ and the Wilcoxon testing procedure $\hat{M}_n$ for each time point $k = 1, \dots, 2191$. That means we divide the sample at the estimated time of change $\hat{k}$ and consider $(\hat{\sigma}_C \hat{k}^{1/2})^{-1} |C_{1,\hat{k}}(k)|$ for $k = 1, \dots, \hat{k}$ and $(\hat{\sigma}_C (n-\hat{k})^{1/2})^{-1} |C_{\hat{k}+1,n}(k)|$ for $k = \hat{k}+1, \dots, n$ for the CUSUM test and $(\hat{\sigma}_W \hat{k}^{3/2})^{-1} |W_{1,\hat{k}}(k)|$ and $(\hat{\sigma}_W (n-\hat{k})^{3/2})^{-1} |W_{\hat{k}+1,n}(k)|$ respectively, for the Wilcoxon test. The results are shown in Figure 4(b). The vertical line in the plot refers to the critical value $c_{5\%} = 1.478$.

Although the data seem to be long-range dependent both testing procedures have a maximum value less than the critical value, where the CUSUM test has a much smaller value $\tilde{M}_{C,n} = 0.89$ than the Wilcoxon test $\hat{M}_n = 1.30$. This seems to be in line with the conclusion of the simulation section that the CUSUM test loses power due to the affect of outliers on the long-run variance estimation. Even though the Wilcoxon test would also not reject, the value is close to the critical value.

## 5. PROOFS

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

### 5.1. Proof of Theorem 2.1

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

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

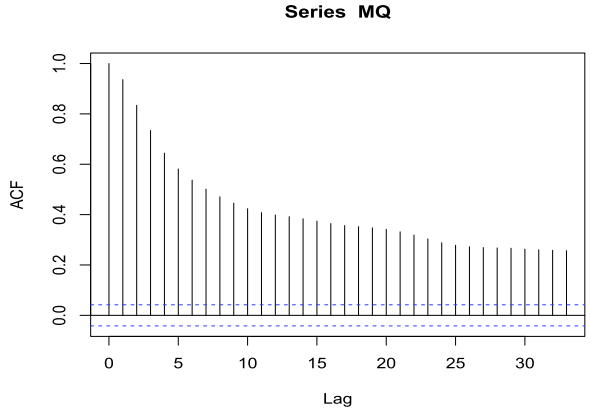



Figure 5. Sample autocorrelation function of the daily MQ of the river Elbe in Dresden [Color figure can be viewed at wileyonlinelibrary.com]

Before we can state the proof of Theorem 2.1, we need to consider Lemma 5.1 and Lemma 5.2, which proofs can be found in Sections 5.1.2 and 5.1.3 respectively.

**Lemma 5.1.** 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,

$$n^{-3/2} \max_{1\le k\le \hat{k}} \left|W_{1,\hat{k}}(k)\right| = n^{-3/2} \max_{1\le k\le \hat{k}} \Big| \sum_{i=1}^{k} \sum_{j=k+1}^{\hat{k}} (1_{\{Y_i\le Y_j\}} - 1/2)\Big| + o_P(1)$$

$$n^{-3/2} \max_{\hat{k}<k\le n} \left|W_{\hat{k}+1,n}(k)\right| = n^{-3/2} \max_{\hat{k}<k\le n} \Big| \sum_{i=\hat{k}+1}^{k} \sum_{j=k+1}^{n} (1_{\{Y_i\le Y_j\}} - 1/2)\Big| + o_P(1).$$

**Lemma 5.2.** Let $\left(Y_j\right)$ satisfy Assumption 1 and let Assumption 2 hold. Then,

$$\Big(T(Y_1, \dots, Y_{\hat{k}}), T(Y_{\hat{k}+1}, \dots, Y_n)\Big) \xrightarrow{d} \Big(\sigma \sup_{0\le t\le 1} \left|B^{(1)}(t)\right|, \sigma \sup_{0\le t\le 1} \left|B^{(2)}(t)\right|\Big), \tag{25}$$

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

*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 5.1 we show that with $\hat{k}$ as in (3),

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

and

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

Subsequently, in Lemma 5.2 we prove that

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




DOI: 10.1111/jtsa.12554 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

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. □

*5.1.1. Auxiliary Results*

We state auxiliary results needed to prove Lemmas 5.1 and 5.2 in Sections 5.1.2 and 5.1.3 respectively.

*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 5.1.** (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 $k \ge 1$

$$\begin{aligned} &\mathrm{E}\left(\left|h\left(Y_1, Y_k\right) - h\left(Y_1', Y_k\right)\right| 1_{\{|Y_1 - Y_1'|\le\epsilon\}}\right) \le \phi(\epsilon), \\ &\mathrm{E}\left(\left|h\left(Y_k, Y_1\right) - h\left(Y_k, Y_1'\right)\right| 1_{\{|Y_1 - Y_1'|\le\epsilon\}}\right) \le \phi(\epsilon), \end{aligned} \tag{26}$$

and

$$\begin{aligned} &\mathrm{E}\left(\left|h\left(Y_1, Y_2'\right) - h\left(Y_1', Y_2'\right)\right| 1_{\{|Y_1 - Y_1'|\le\epsilon\}}\right) \le \phi(\epsilon), \\ &\mathrm{E}\left(\left|h\left(Y_2', Y_1\right) - h\left(Y_2', Y_1'\right)\right| 1_{\{|Y_1 - Y_1'|\le\epsilon\}}\right) \le \phi(\epsilon), \end{aligned} \tag{27}$$

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 5.2.** 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 $\epsilon > 0$

$$\mathrm{E}\left(\left|g\left(Y_1\right) - g\left(Y_1'\right)\right| 1_{\{|Y_1 - Y_1'|\le\epsilon\}}\right) \le \phi(\epsilon), \tag{28}$$

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

Note that the term $W_{1,n}(k)$ can be written as a second order $U$-statistic

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

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

with kernel function $h(x,y) = 1_{\{x\le y\}}$ and constant $\Theta = \mathrm{E}\, h(Y_1', Y_2') = 1/2$, where $Y_1'$ and $Y_2'$ are independent copies of $Y_1$.

By applying Hoeffding's decomposition of $U$-statistics (Hoeffding (1948)) to $U_{a,b}(k)$, the kernel function $h$ can be written as the sum

$$h(x,y) = \Theta + h_1(x) + h_2(y) + g(x,y), \tag{29}$$

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

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

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

**Remark 5.1.** 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 (26) and (27)) with respect to the distribution of $(Y_j)$ with function $\phi(\epsilon) = C\epsilon$, for some $C > 0$, see for example, 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 (26) and (27) 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 (28) with the same function $\phi(\epsilon)$. Hence, $h_1(x) = \mathrm{E}\, h\left(x, Y_2'\right) - 1/2$ and $h_2(y) = \mathrm{E}\, h\left(Y_2', y\right) - 1/2$ are 1-continuous.

(iii) The function $g(x,y) = h(x,y) - h_1(x) - h_2(x) - 1/2$ is 1-continuous (satisfies (26) and (27)), since $h$ and $h_1$ satisfy (26), (27) and (28) with $\phi(\epsilon) = C\epsilon$, for some $C > 0$. In particular,

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

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

**Lemma 5.3.** Suppose that the assumptions of Lemma 5.2 hold. Then,

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

where $W(t)$ is a Brownian motion and $\sigma$ 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, Lemma 5.3 is proved, by showing that $(h_1(Y_j))$ is $L_2$ NED on a strongly mixing process.




DOI: 10.1111/jtsa.12554 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

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 $\phi$, 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 5.1 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 $a'_k = C\sqrt{a_k} \geq \phi\left(\sqrt{2a_k}\right) + 2\sqrt{2a_k}||h_1||_\infty$.

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 $|h_1(Y_1)| \leq C < \infty$. Thus,

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

The last inequality holds, because by $L_1$ NED of $(h_1(Y_j))$, $\mathrm{E}\,|h_1(Y_1) - \mathrm{E}(h_1(Y_1)|\mathcal{G}^k_{-k})| \leq a'_k$. 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 proves the lemma. □

Next we show that the contribution of $g(x, y)$ of the Hoeffding decomposition (29) is negligible.

**Lemma 5.4.** Suppose that the assumptions of Lemma 5.2 hold. Then,

$$n^{-3/2} \max_{1\leq k\leq n} \max_{1\leq l\leq n} \Big| \sum_{i=1}^{k} \sum_{j=1}^{l} g(Y_i, Y_j)\Big| = o_P(1). \tag{30}$$

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

$$\mathrm{E}\left(\Big|n^{-3/2} \sum_{i=q+1}^{p} \sum_{j=h+1}^{l} g(Y_i, Y_j)\Big|^2\right) \leq \frac{C}{n^3}(p-q)(l-h). \tag{31}$$

*Proof of (31)* Lemma 1 of Dehling *et al.* (2015) showed if $f$ is a 1-continuous bounded degenerate kernel function and $\phi_f(\epsilon)$ satisfies

$$\sum_{k=1}^{\infty} k(\beta(k) + \sqrt{a_k} + \phi_f(a_k)) < \infty, \tag{32}$$

then

$$\mathrm{E}\Big(\sum_{i=1}^{k} \sum_{j=k+1}^{n} f(Y_i, Y_j)\Big)^2 \leq Ck(n-k), \qquad 1 \leq k \leq n, \tag{33}$$

where the constant $C$ depends on the left-hand side of (32). The proof of Lemma 1 in Dehling *et al.* (2015) shows that (33) can be extended to (31). 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 (29), $g(x, y) = h(x, y) + F(x) - F(y) - 1/2$. Note that $\mathrm{E}\, F(Y_1) = 1/2$, thus $\mathrm{E}\, g(x, Y_1) = \mathrm{E}\, g(Y_1, y) = 0$, that is, $g(x, y)$ is a degenerate kernel. Furthermore, $g(x, y)$ is bounded, since $h(x, y) = 1_{\{x\leq y\}}$ and $F(x)$ are bounded. By Remark 5.1 iii) $g(x, y)$ is 1-continuous with $\phi(\epsilon) = C\epsilon$, the latter satisfies (32) because of condition (10). This completes the proof of (31).

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

*Proof of (30)* 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 $\zeta_i$, $i = 1, 2, \ldots$ 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

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

for $\epsilon > 0$, $0 \leq i \leq j \leq n$, then for all $\epsilon > 0$, $n \geq 2$,

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

where $K > 0$ depends only on $\alpha$ and $\beta$.

Denote

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

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 using the reverse triangle inequality, for $1 \leq h \leq l \leq n$,

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

Let us now define

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

and note that $\tilde{S}_k$ depends on $h$ and $l$. Furthermore, note that for $1 \leq q \leq p \leq n$,

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

By Markov inequality and (31),

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

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 $\epsilon > 0$,

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

and moreover

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

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 $n \to \infty$,

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

which proves the lemma. □

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

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

and

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

appearing in the proof of Lemma 5.1.

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

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

with kernel function $h_n(x,y) = 1_{\{y<x\leq y+\Delta_n\}}$.

Applying Hoeffding's decomposition of $U$-statistics to $\tilde{U}_{a,b}(k)$, decomposes the kernel function $h_n$ into the sum

$$h_n(x,y) = \Theta_{\Delta_n} + h_{1,n}(x) + h_{2,n}(y) + g_n(x,y), \tag{34}$$

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

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

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

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

**Lemma 5.5.** Suppose that the assumptions of Lemma 5.1 hold. Then,

$$n^{-3/2}\left|\tilde{U}_{1,\hat{k}}(k^*) - k^*(\hat{k}-k^*)\Theta_{\Delta_n}\right| = o_P(1) \tag{35}$$

and

$$n^{-3/2}\left|\tilde{U}_{\hat{k}+1,n}(k^*) - (k^*-\hat{k})(n-k^*)\Theta_{\Delta_n}\right| = o_P(1), \tag{36}$$

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

*Proof.* Let us start with the proof of (35). The Hoeffding decomposition (34) yields

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

Therefore,

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

Note that the indicator function $h_n(x,y) = 1_{\{y<x\le y+\Delta_n\}}$ is bounded.
The distribution function $F$ of $Y_1$ has bounded second derivative. Hence, as $n\to\infty$,

$$\begin{aligned}\Theta_{\Delta_n} &= \mathrm{E}\,1_{\{Y_2'<Y_1'\le Y_2'+\Delta_n\}} = \mathrm{P}\left(Y_2' < Y_1' \le Y_2' + \Delta_n\right)\\ &= \int_{\mathbb{R}}\left(F\left(y+\Delta_n\right) - F(y)\right)\mathrm{d}F(y) = \Delta_n\left(\int_{\mathbb{R}} f^2(y)\,\mathrm{d}y + o(1)\right) \sim C\Delta_n.\end{aligned} \tag{37}$$

Thus,

$$\begin{aligned}|h_{1,n}(x)| &\le |F(x) - F(x-\Delta_n) - \Theta_{\Delta_n}| \le C\Delta_n + \Theta_{\Delta_n} \le C\Delta_n,\\ |h_{2,n}(x)| &\le |F(x+\Delta_n) - F(x) - \Theta_{\Delta_n}| \le C\Delta_n + \Theta_{\Delta_n} \le C\Delta_n,\end{aligned} \tag{38}$$

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, that is, $\mathrm{E}\,g_n(x,Y_1) = \mathrm{E}\,g_n(Y_1,y) = 0$. $h_n(x,y)$ satisfies (26) and (27) with $\phi_{h_n}(\epsilon) = C\epsilon$, see for example, Corollary 4.1 of Gerstenberger (2018), where constant $C$ does not depend on $n$. Then, with similar argument as in Remark 5.1, $h_{1,n}$ 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 (32). Hence, $g_n(x,y)$ satisfies the conditions on $g(x,y)$ in Lemma 5.4, which yields

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




DOI: 10.1111/jtsa.12554 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Thus, it remains to show $n^{-3/2}\big|(\hat{k}-k^*)\sum_{i=1}^{k^*} h_{1,n}\left(Y_i\right)+k^*\sum_{j=k^*+1}^{\hat{k}} h_{2,n}\left(Y_j\right)\big|=o_P(1)$.
By (38), we receive the following inequality

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

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 (35).
The proof of (36) follows using similar argument. □

*5.1.2. Proof of Lemma 5.1*
Before proceeding to Lemma 5.1, similarly to the notation $W_{m,n}(k)$ in (2), we define

$$U_{m,n}(k)=\sum_{i=m}^{k}\sum_{j=k+1}^{n}(1_{\{Y_i\leq Y_j\}}-1/2),\qquad m\leq k\leq n. \tag{39}$$

Note that $W_{m,n}(k)$ depends on $(X_m,\dots,X_n)$, where $U_{m,n}(k)$ depends on $(Y_m,\dots,Y_n)$.

**Lemma 5.1.** 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,

$$n^{-3/2}\max_{1\leq k\leq\hat{k}}\left|W_{1,\hat{k}}(k)\right|=n^{-3/2}\max_{1\leq k\leq\hat{k}}\left|U_{1,\hat{k}}(k)\right|+o_P(1) \tag{40}$$
$$n^{-3/2}\max_{\hat{k}<k\leq n}\left|W_{\hat{k}+1,n}(k)\right|=n^{-3/2}\max_{\hat{k}<k\leq n}\left|U_{\hat{k}+1,n}(k)\right|+o_P(1). \tag{41}$$

*Proof.* We have to distinguish between two cases, $\hat{k}\leq k^*$ and $\hat{k}>k^*$, where $k^*=[n\theta]$.
If $\hat{k}\leq k^*$, 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 $1_{\{Y_i+\Delta_n\leq Y_j+\Delta_n\}}=1_{\{Y_i\leq Y_j\}}$, $W_{\hat{k}+1,n}(k)$ can be decomposed into two terms,

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

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

$$W_{1,\hat{k}}(k)=\begin{cases}U_{1,\hat{k}}(k)+\sum_{i=1}^{k}\sum_{j=k^*+1}^{\hat{k}}1_{\{Y_j<Y_i\leq Y_j+\Delta_n\}}, & 1\leq k\leq k^*\\ U_{1,\hat{k}}(k)+\sum_{i=1}^{k^*}\sum_{j=k+1}^{\hat{k}}1_{\{Y_j<Y_i\leq Y_j+\Delta_n\}}, & k^*<k\leq\hat{k}.\end{cases} \tag{42}$$

*Proof of (40).* For $\hat{k}\leq k^*$, equation (40) holds trivially, since $W_{1,\hat{k}}(k)=U_{1,\hat{k}}(k)$, $k=1,\dots,\hat{k}$.

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

For $\hat{k} > k^*$, equation (42) yields,

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

for all $1 \leq k \leq \hat{k}$. Hence, using the reverse triangle inequality,

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

Thus, property (40) holds if $n^{-3/2} I_{1,\hat{k}}(k^*) = o_P(1)$.

By Lemma 5.5, $n^{-3/2} I_{1,\hat{k}}(k^*) = n^{-3/2} k^* (\hat{k} - k^*) \Theta_{\Delta_n} + o_P(1)$, where $\Theta_{\Delta_n} = \mathrm{E}\left(1_{\{Y'_2 < Y'_1 \leq Y'_2 + \Delta_n\}}\right)$ 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 $n \to \infty$, by (37),

$$\Theta_{\Delta_n} = \Delta_n \left( \int_{\mathbb{R}} f^2(y)\, \mathrm{d}y + o(1) \right).$$

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 (40). The proof of (41) follows using similar argument. □

*5.1.3. Proof of Lemma 5.2*

We will now state the proof of Lemma 5.2.

*Proof.* To prove Lemma 5.2 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 (39) can be written as a second order $U$-statistic

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

with kernel function $h(x, y) = 1_{\{x\leq y\}}$ and constant $\Theta = \mathrm{E}\, h(Y'_1, Y'_2) = 1/2$, where $Y'_1$ and $Y'_2$ are independent copies of $Y_1$. Furthermore, we can apply the Hoeffding's decomposition given in (29).

Therefore,

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

where

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

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 5.4 yields

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

Furthermore, by the triangle inequality,

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

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

$$\left|\frac{\hat{k}}{n} - \theta\right| = o_P(1). \tag{43}$$

It remains to show that

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

where $B^{(1)}$ and $B^{(2)}$ are independent Brownian bridges. By Slutsky's Lemma this implies (25). 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

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

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

Lemma 5.3 implies weak convergence on $D[0,1]$ of the partial sum process,

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

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

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

Set

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

and note that $\Gamma^{(1)}_{W,k}$ and $\Gamma^{(2)}_{W,k}$ are independent, since the increments of Brownian motions are independent. Thus,

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

By (43) and by the a.s. equicontinuity of the Brownian motion process $\{W_n\}$ and using the continuous mapping theorem, $\left|W_n\big(\hat{k}/n\big)-W_n(\theta)\right| = o_P(1)$. Hence,

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

and

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

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

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

since Brownian motions are scale invariant, that is, $\theta^{-1/2}W_n(t)\stackrel{d}{=}W_n(t/\theta)$, and

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

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

### 5.2. Proof of Theorem 2.2

Under the alternative we consider observations $X_1, \dots, X_n$ with $X_i = G(\xi_i) + \mu$, $i = 1, \dots, n$. Note that the indicator function $1_{\{x \le y\}}$ is invariant under strictly increasing functions, that is, $1_{\{G(\xi_i) \le G(\xi_j)\}} = 1_{\{\xi_i \le \xi_j\}}$, 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 $(\xi_j)$, that is, $T_{n,1}(\xi_1, \dots, \xi_{\hat{k}})$ and $T_{n,2}(\xi_{\hat{k}+1}, \dots, \xi_n)$, instead of $T_{n,1}(X_1, \dots, X_{\hat{k}})$ and $T_{n,2}(X_{\hat{k}+1}, \dots, X_n)$.

Before we prove that the test $M_n$ tends to infinity in probability under the alternative, we will consider the limit distribution of $T_{n,1}(\xi_1, \dots, \xi_{\hat{k}})$ and $T_{n,2}(\xi_{\hat{k}+1}, \dots, \xi_n)$ in Lemma 5.7, 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 5.6.** 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 5.6 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 $\mathrm{E}\, g^2(\xi) < \infty$, $\xi \sim \mathrm{N}(0,1)$, that is, $g \in L^2(\mathbb{R}, N)$, we could represent $g$ by its Hermite expansion

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

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, for example, in Taqqu (1978). If $m = 1$, $Z_1(t)$ is the standard Gaussian fractional Brownian motion.

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

*Proof of Lemma 5.6.* Dehling *et al.* (2013b) have shown in their Theorem 1 that

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

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

$$\left(\frac{1}{n^{d+3/2}c_d}\sum_{i=1}^{[ns]}\sum_{j=[nt]+1}^{n}(1_{\{X_i\le X_j\}}-1/2)\right)_{0\le s\le t\le 1}$$
$$\xrightarrow{d}\left(\frac{1}{m!}\big((1-t)Z_m(s)-s(Z_m(1)-Z_m(t)\big)\int_{\mathbb{R}}J_m(x)\mathrm{d}F(x)\right)_{0\le s\le t\le 1}. \tag{44}$$

Since $F$ is a continuous distribution function, $\int_{\mathbb{R}}F(x)\mathrm{d}F(x)=1/2$. 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,

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

Integration by parts yields,

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

Hence,

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

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

$$\left(\frac{[ns](n-[nt])}{n^{d+3/2}c_d}\int_{\mathbb{R}}(F_{[ns]}(x)-F(x))\mathrm{d}F_{[nt]+1,n}(x)\right)_{0\le s\le t\le 1}$$
$$\xrightarrow{d}\left(\frac{(1-t)}{m!}\int_{\mathbb{R}}J_m(x)Z_m(s)\mathrm{d}F(x)\right)_{0\le s\le t\le 1},$$




DOI: 10.1111/jtsa.12554 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

and

$$\left(\frac{[ns](n-[nt])}{n^{d+3/2}c_d}\int_{\mathbb{R}}(F_{[nt]+1,n}(x)-F(x))\mathrm{d}F(x)\right)_{0\le s\le t\le 1}\xrightarrow{d}\left(\frac{s}{m!}\int_{\mathbb{R}}J_m(x)(Z_m(1)-Z_m(t))\mathrm{d}F(x)\right)_{0\le s\le t\le 1}.$$

We do this by applying the Skorohod–Dudley–Wichura representation which yields almost sure convergence, that is

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

$$\frac{[ns](n-[nt])}{n^{d+3/2}c_d}\int_{\mathbb{R}}(F_{[nt]+1,n}(x)-F(x))\mathrm{d}F(x)-\frac{s}{m!}\int_{\mathbb{R}}J_m(x)(Z_m(1)-Z_m(t))\mathrm{d}F(x)\to 0, \tag{46}$$

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

Let us start with (45). We can write

$$\begin{aligned}
&\frac{[ns](n-[nt])}{n^{d+3/2}c_d}\int_{\mathbb{R}}(F_{[ns]}(x)-F(x))\mathrm{d}F_{[nt]+1,n}(x)-\frac{(1-t)}{m!}\int_{\mathbb{R}}J_m(x)Z_m(s)\mathrm{d}F(x)\\
&=\frac{(n-[nt])}{n}\int_{\mathbb{R}}\frac{[ns]}{n^{d+1/2}c_d}(F_{[ns]}(x)-F(x))\mathrm{d}F_{[nt]+1,n}(x)-(1-t)\int_{\mathbb{R}}J_m(x)\frac{Z_m(s)}{m!}\mathrm{d}F(x)\\
&=\frac{(n-[nt])}{n}\int_{\mathbb{R}}\Big(\frac{[ns]}{n^{d+1/2}c_d}(F_{[ns]}(x)-F(x))-J_m(x)\frac{Z_m(s)}{m!}\Big)\mathrm{d}F_{[nt]+1,n}(x)\\
&\quad+\frac{(n-[nt])}{n}\int_{\mathbb{R}}J_m(x)\frac{Z_m(s)}{m!}\mathrm{d}\big(F_{[nt]+1,n}-F\big)(x)\\
&\quad+\Big(\frac{(n-[nt])}{n}-(1-t)\Big)\int_{\mathbb{R}}J_m(x)\frac{Z_m(s)}{m!}\mathrm{d}F(x).
\end{aligned}\tag{47}$$

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, that is

$$\sup_{s,x}\big|d_n^{-1}[ns]\big(F_{[ns]}(x)-F(x)\big)-J(x)Z(x)\big|\to 0\qquad\text{a.s.} \tag{48}$$

Thus, the first term on the right-hand side of (47) converges to 0 almost surely, uniformly in $0<s\le t<1$.

Furthermore, we note that

$$\begin{aligned}
&\frac{(n-[nt])}{n}\int_{\mathbb{R}}J(x)Z(s)\mathrm{d}\big(F_{[nt]+1,n}-F\big)(x)\\
&=Z(s)\Big[\frac{(n-[nt])}{n}\int_{\mathbb{R}}J(x)\mathrm{d}F_{[nt]+1,n}(x)-\frac{(n-[nt])}{n}\int_{\mathbb{R}}J(x)\mathrm{d}F(x)\Big]
\end{aligned}$$

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

$$= Z(s)\Big[\frac{1}{n}\sum_{i=[nt]+1}^{n} J(X_i) - \frac{(n-[nt])}{n}\,\mathrm{E}(J(X_i))\Big]$$
$$= Z(s)\frac{1}{n}\sum_{i=1}^{n}\big(J(X_i)-\mathrm{E}(J(X_i))\big) - Z(s)\frac{1}{n}\sum_{i=1}^{[nt]}\big(J(X_i)-\mathrm{E}(J(X_i))\big).$$

Note that $(J(X_i))$ is ergodic since the process $(X_i)$ is ergodic and $J$ is a measurable function. By the ergodic theorem, $\frac{1}{n}\sum_{i=1}^{n}\big(J(X_i)-\mathrm{E}(J(X_i))\big) \to 0$ almost surely. This implies that $\sum_{i=1}^{n}\big(J(X_i)-\mathrm{E}(J(X_i))\big) = o(n)$ and hence

$$\max_{0\le k\le n}\Big|\sum_{i=1}^{k}\big(J(X_i)-\mathrm{E}(J(X_i))\big)\Big| = o(n)$$

almost surely as $n\to\infty$. Thus, $\frac{1}{n}\sum_{i=1}^{[nt]}\big(J(X_i)-\mathrm{E}(J(X_i))\big) \to 0$ almost surely for all $0\le t\le 1$. Therefore, the second term on the right-hand side of (47) converges to 0 almost surely, uniformly in $0<s\le t<1$.

Also the third term on the right-hand side of (47) converges to 0, since, as $n\to\infty$, $\big((n-[nt])/n-(1-t)\big)\to 0$, and $\int_{\mathbb{R}} J_m(x)\frac{Z_m(s)}{m!}\mathrm{d}F(x)$ is bounded. This finishes the proof of (45).

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 (46) follows using again (48). Thus, (44) is shown.

Note that this result holds for $X_i = G(\xi_i)$, but in our lemma we consider $X_i=\xi_i$, where $(\xi_j)$ 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 $\int_{\mathbb{R}} J_1(x)\mathrm{d}F(x) = -\frac{1}{2\sqrt{\pi}}$, 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 (44) 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 5.7.** Assume that the assumptions of Theorem 2.2 hold. Then,

$$\Big[\frac{1}{n^{d+3/2}c_d}\max_{1\le k\le\hat{k}}\Big|\sum_{i=1}^{k}\sum_{j=k+1}^{\hat{k}}(1_{\{\xi_i\le\xi_j\}}-1/2)\Big|, \frac{1}{n^{d+3/2}c_d}\max_{\hat{k}<k\le n}\Big|\sum_{i=\hat{k}+1}^{k}\sum_{j=k+1}^{n}(1_{\{\xi_i\le\xi_j\}}-1/2)\Big|\Big]$$
$$\xrightarrow{d}\Big[\frac{\zeta}{2\sqrt{\pi}}\sup_{0\le t\le\zeta}\Big|W_H(t)-\frac{t}{\zeta}W_H(\zeta)\Big|, \frac{1-\zeta}{2\sqrt{\pi}}\sup_{\zeta\le t\le 1}\Big|W_H(t)-W_H(\zeta)-\frac{t-\zeta}{1-\zeta}(W_H(1)-W_H(\zeta))\Big|\Big],$$

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

$$\zeta = \inf\Big\{t\ge 0 : \sup_{0\le s\le 1}|W_H(s)-sW_H(1)| = |W_H(t)-tW_H(1)|\Big\}. \tag{49}$$




DOI: 10.1111/jtsa.12554 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

*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 5.6, $(\tilde{U}_n(s,t))_{s,t} \xrightarrow{d} (\tilde{W}_H(s,t))_{s,t}$. 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 $\zeta$ is given in (49). Hence, by the continuous mapping theorem and Lemma 5.6

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

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 5.7,

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

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 $\frac{n^{d+3/2}c_d}{\hat{k}^{3/2}} \to_p \infty$ and $\frac{n^{d+3/2}c_d}{(n-\hat{k})^{3/2}} \to_p \infty$. The proof of Lemma 5.7 yields $\hat{k}/n \xrightarrow{d} \zeta$, where $\zeta$ is given in (49), and hence, $(\hat{k}/n)^{3/2}$ and

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554

$((n-\hat{k})/n)^{3/2}$ are asymptotically bounded away from zero. Since $d > 0$, $n^d \to \infty$ as $n \to \infty$. Thus, $T_{n,1} \to_p \infty$ and $T_{n,2} \to_p \infty$. This finishes the proof of Theorem 2.2. □

## ACKNOWLEDGEMENTS

The author thanks Herold Dehling, Liudas Giraitis and Isabel Garcia for valuable discussions and the anonymous referees for their constructive comments. The research was supported by the Collaborative Research Centre 823 *Statistical modelling of nonlinear dynamic processes* and the Konrad–Adenauer–Stiftung. The author thanks Svenja Fischer for providing the hydrologic data set.

Open access funding enabled and organized by ProjektDEAL.

## DATA AVAILABILITY STATEMENT

The data used in Section 4 are property of the German state Saxony and are therefore not openly available but can be requested from the Saxon State Office for Environment, Agriculture and Geology.

## REFERENCES

Baek C, Pipiras V. 2012. Statistical tests for a single change in mean against long-range dependence. *Journal of Time Series Analysis* **33**: 131–151.

Berkes I, Horváth L, Kokoszka P, Shao Q. 2006. On discriminating between long-range dependence and changes in mean. *Annals of Statistics* **34**: 1140–1165.

Bhattacharya R, Gupta V, Waymire E. 1983. The Hurst effect under trends. *Journal of Applied Probability* **20**: 649–662.

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

Borovkova S, Burton R, Dehling H. 2001. Limit theorems for functionals of mixing processes with applications to *U*-statistics and dimension estimation. *Transactions of the American Mathematical Society* **353**: 4261–4318.

Bradley RC. 2007. *Introduction to Strong Mixing Conditions*. Heber City: Kendrick Press.

Carlstein E. 1986. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics* **14**: 1171–1179.

Chang SY, Perron P. 2016. Inference on a structural break in trend with fractionally integrated errors. *Journal of Time Series Analysis* **37**: 555–574.

Dehling H, Fried R, Garcia Arboleda I, Wendler M. 2015. Change-point detection under dependence based on two-sample *U*-statistics. In *Asymptotic laws and methods in stochastics. Fields Institute Communication*, 195–220.

Dehling H, Fried R, Sharipov O, Vogel D, Wornowizki M. 2013a. Estimation of the variance of partial sums of dependent processes. *Statistics & Probability Letters* **83**: 141–147.

Dehling H, Rooch A, Taqqu MS. 2013b. Non-parametric change-point tests for long-range dependent data. *Scandinavian Journal of Statistics* **40**: 153–173.

Dehling H, Taqqu MS. 1989. The empirical process of some long-range dependent sequences with an application to *U*-statistics. *Annals of Statistics* **17**: 1767–1783.

Diebold FX, Inoue A. 2001. Long memory and regime switching. *Journal of Econometrics* **105**: 131–159.

Dürre A, Fried R, Liboschik T. 2015. Robust estimation of (partial) autocorrelation. *WIREs Computational Statistics* **7**: 205–222.

Gerstenberger C. 2018. Robust Wilcoxon-type estimation of change-point location under short-range dependence. *Journal of Time Series Analysis* **1**: 90–104.

Giraitis L, Kokoszka P, Leipus R. 2001. Testing for long memory in the presence of a general trend. *Journal of Applied Probability* **38**: 1033–1054.

Granger C, Hyung N. 2004. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance* **11**: 399–421.

Hansen BE. 1991. GARCH(1,1) processes are near epoch dependent. *Economics Letters* **36**: 181–186.

Hoeffding W. 1948. A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* **19**: 293–325.

Hurst H. 1951. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* **116**: 770–808.

Iacone F, Leybourne SJ, Taylor AMR. 2013a. Testing for a break in trend when the order of integration is unknown. *Journal of Econometrics* **176**: 30–45.

  
DOI: 10.1111/jtsa.12554 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

Iacone F, Leybourne SJ, Taylor AMR. 2013b. On the behavior of fixed-b trend break tests under fractional integration. *Econometric Theory* **29**: 393–418.
Iacone F, Leybourne SJ, Taylor AMR. 2017. Testing for a change in mean under fractional integration. *Journal of Time Series Econometrics* **9**(1).
Iacone F, Leybourne SJ, Taylor AMR. 2019. Testing the order of fractional integration of a time series in the possible presence of a trend break at an unknown point. *Econometric Theory* **35**: 1201–1233.
Künsch H. 1986. Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability* **4**: 1025–1030.
Lavielle M, Moulines E. 2000. Least squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis* **21**: 33–59.
Ma Y, Genton M. 2000. Highly robust estimation of the autocovariance function. *Journal of Time Series Analysis* **21**: 663–684.
Mandelbrot B, Van Ness J. 1968. Fractional Brownian motions, fractional noises and applications. *Society for Industrial and Applied Mathematics* **10**: 422–437.
Mandelbrot B, Wallis J. 1968. Noah, Joseph, and operational hydrology. *Water Resources Research* **4**: 909–918.
McCloskey A, Perron P. 2013. Memory parameter estimation in the presence of level shifts and deterministic trend. *Econometric Theory* **29**: 1196–1237.
Preuß P, Sen K, Dette H. 2017. Detecting long-range dependence in non-stationary time series. *Electronic Journal of Statistics* **1**: 1600–1659.
Qu Z. 2011. A test against spurious long memory. *Journal of Business and Economic Statistics* **29**: 426–438.
Rooch A. 2012. *Change-Point Tests For Long-Range Dependent Data*, PhD thesis, Ruhr-Universität Bochum, Germany.
Rousseeuw PJ, Croux C. 1993. Alternatives to the median absolute deviation. *Journal of the American Statistical Association* **88**: 1273–1283.
Shorack GR, Wellner JA. 2009. *Empirical Processes with Applications to Statistics*. New York: Wiley.
Taqqu M. 1978. A representation for self-similar processes. *Stochastic Processes and Their Applications* **7**: 55–64.
Wooldridge JM, White H. 1988. Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* **4**: 210–230.

  
*Journal of Time Series Analysis* published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12554