Electronic Journal of Statistics
Vol. 16 (2022) 862–891
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1964

Change-point detection based on weighted two-sample U-statistics*

Herold Dehling and Kata Vuk

Fakultät für Mathematik,
Ruhr-Universität Bochum,
44780 Bochum, Germany
e-mail: herold.dehling@ruhr-uni-bochum.de; kata.vuk@ruhr-uni-bochum.de

Martin Wendler

Institut für Mathematische Stochastik,
Otto-von-Guericke-Universität Magdeburg,
39106 Magdeburg, Germany
e-mail: martin.wendler@ovgu.de

Abstract: We investigate the large-sample behavior of change-point tests based on weighted two-sample U-statistics, in the case of short-range dependent data. Under some mild mixing conditions, we establish convergence of the test statistic to an extreme value distribution. A simulation study shows that the weighted tests are superior to the non-weighted versions when the change-point occurs near the boundary of the time interval, while they loose power in the center.

Keywords and phrases: Change-point tests, U-statistics, Wilcoxon test, short-range dependence, extreme value distribution.

Received February 2021.

Contents

1 Introduction . . . 863
2 Main theoretical results . . . 865
3 Simulations . . . 869
4 Data example . . . 873
5 Proofs . . . 875
5.1 Proof of Theorem 1 . . . 875
5.2 Proof of Theorem 2 . . . 878
5.3 Proof of Theorem 3 . . . 880
5.4 Proof of Theorem 4 . . . 884
A Auxiliary results . . . 887
References . . . 890

*H. Dehling and K. Vuk were supported by the Collaborative Research Grant SFB 823 *Statistical modelling of nonlinear dynamic processes.* M. Wendler was supported by the German Research Foundation (DFG), project WE 5988/3 *Analyse funktionaler Daten ohne Dimensionsreduktion.*

1. Introduction

In this paper, we study nonparametric tests for change-points in time series that are based on weighted two-sample U-statistics. By a suitable choice of the weights, we obtain tests that are able to detect changes that occur very early or very late during the observation period. Our results cover both the CUSUM test and the Wilcoxon change-point test, as well as many other robust and non-robust tests. We investigate the large-sample behavior of our tests in the case of short-range dependent data under mild conditions that cover, e.g. ARMA and ARCH processes. By means of a simulation study, we analyze the small sample behavior of our tests, e.g. regarding robustness and the ability to detect early and late changes. As an application, we analyze a data set of daily stock returns of Wirecard during the weeks prior to the detection of accounting fraud in June 2020.

There is a vast body of literature on change-point tests using U-statistics. Gombay [14], and Kirch and Stoehr [16] investigate sequential change-point tests based on U-statistics. Wang, Volgushev and Shao [27] apply certain U-statistics for change-point detection in high-dimensional time series. Zhang, Wang, and Shao [29] propose an adaptive change-point test based on U-statistics. Račkauskas and Wendler [20] apply U-statistics for the detection of epidemic changes.

We assume that the data are generated by a stochastic process $(X_i)_{i\geq 1}$ which follows the model

$$X_i = \mu_i + \xi_i,\ i \geq 1,$$

where $(\mu_i)_{i\geq 1}$ is an unknown signal, and where $(\xi_i)_{i\geq 1}$ is a short-range dependent stationary stochastic process. Given the observations $X_1, \ldots, X_n$, we want to test the hypothesis that the signal is constant, i.e.

$$H_0\ :\ \mu_1 = \ldots = \mu_n$$

against the alternative of a change in the mean at an unknown point k^* in time, i.e.

$$H_1\ :\ \mu_1 = \ldots = \mu_{k^*} \neq \mu_{k^*+1} = \ldots = \mu_n,\ \text{for some } k^* \in \{1, \ldots, n-1\}.$$

If the change-point k^* was known, we would have a two-sample problem with the samples $X_1, \ldots, X_{k^*}$ and $X_{k^*+1}, \ldots, X_n$, respectively, and we could apply standard tests such as the two-sample Student t-test or the Wilcoxon two-sample test for a change in mean. Up to normalization, both are special cases of two-sample U-statistics $\sum_{i=1}^{k^*} \sum_{j=k^*+1}^{n} h(X_i, X_j)$, with a suitably chosen kernel function $h : \mathbb{R}^2 \to \mathbb{R}$.

In the change-point setting, where a change occurs at an unknown point in time, we have a family of two-sample problems, indexed by the potential change-point k, and thus we are naturally led to the two-sample U-statistic process

$$\sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j),\ 0 \leq k \leq n.$$

A variety of change-point tests can be derived from this process by taking suitable functionals such as weighted maxima

$$\max_{1\le k\le n-1} \frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^{\gamma}} \frac{1}{n^{3/2}} \Big| \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j) \Big|,$$

where $0 \le \gamma \le \frac{1}{2}$ is some parameter to be chosen.

For $\gamma = 0$, we obtain non-weighted tests, which have been widely studied in the literature, starting with Darkhovskh [5] and Pettitt [19], who studied the special case of a Wilcoxon-type test statistic. Csörgő and Horvath [2] investigated U-statistics with general kernels, in the case of independent data. They could show that the asymptotic distribution under the null hypothesis is the Kolmogorov-Smirnov distribution, which is the distribution of the supremum of a Brownian bridge. Dehling, Fried, Garcia and Wendler [7] extended these results to weakly dependent data. Under long-range dependence, the limiting distribution is given by the supremum of a linear combination of Hermite processes; see Dehling, Rooch and Taqqu [10] for the Wilcoxon test, and Dehling, Rooch and Wendler [11] for arbitrary kernels. For $0 < \gamma < \frac{1}{2}$, the limit distribution under independence is the supremum of the appropriately weighted Brownian bridge, see e.g. the seminal monograph by Csörgő and Horváth [3]. The moment conditions for such a limit theorem have been relaxed by Csörgő, Szyszkowicz, and Wang [4].

In the present paper, we focus on the extreme case $\gamma = \frac{1}{2}$, where we obtain the test statistic

$$T_n := \max_{1\le k\le n-1} \frac{1}{\sqrt{k(n-k)n}} \Big| \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j) \Big| \tag{1}$$

Under the null hypothesis, after some suitable normalization and centering, T_n converges in distribution to the Gumbel extreme value distribution. This has been derived by Csörgő and Horváth [2] under independence. We will show that the same holds in the case of short-range dependent data. Antoch, Hušková und Prášková [1] have studied the large-sample behavior of weighted versions of the CUSUM test for dependent observations, in particular for linear processes. We also show that the test is consistent against a wide class of alternatives. For independent data, the behavior under the alternative has been studied by Ferger [13] and Gombay [14]. We have conducted an extensive simulation study comparing this test with the corresponding non-weighted test. Our simulations confirm the intuition that the weighted tests have more power against very early and very late changes, while the non-weighted tests are more powerful against changes in the middle of the observation period.

By the choice of the kernel function h, the weighted two-sample U-statistics lead to a flexible class of change-point tests. As special examples, we obtain the CUSUM test for $h(x, y) = y - x$, and the Wilcoxon test for $h(x, y) = 1_{\{x\le y\}} - \frac{1}{2}$. More generally, one can take the kernels $h(x, y) = \psi(y - x)$ for some anti-symmetric function $\psi : \mathbb{R} \longrightarrow \mathbb{R}$. Depending on the choice of ψ, one obtains

tests with specific properties, such as robustness against outliers, and tests that are powerful for certain alternatives.

The rest of the paper is organized as follows. In the next section, we present the detailed technical assumptions, and we give the main theoretical results of our paper. In section 3, we present the outcomes of a major simulation study comparing weighted and non-weighted tests as well as the robust Wilcoxon test and the non-robust CUSUM test. Full details of the proofs are presented in the final section.

2. Main theoretical results

In this section, we analyze the large sample behavior of the suitably normalized and centered test statistic T_n under the hypothesis, and under a wide class of alternatives. A major ingredient in the proof is a new Darling-Erdős type limit theorem for the tied-down random walk of dependent random variables, which might also be of independent interest. Before we present our results, we give some definitions. Throughout this paper, the stochastic process $(X_i)_{i\geq 1}$ will be assumed to be α-mixing in the sense of Rosenblatt [23].

Definition 2.1. *The stochastic process $(X_i)_{i\geq 1}$ is said to be α-mixing if*

$$\alpha(k) := \sup_{n\geq 1} \sup \left\{ |P(A\cap B) - P(A)\,P(B)| : A \in \mathcal{F}_1^n, B \in \mathcal{F}_{n+k}^{\infty} \right\} \longrightarrow 0,$$

as $k \longrightarrow \infty$, where $\mathcal{F}_a^b$ denotes the σ-field generated by the random variables $X_a, \ldots, X_b$. We define the generalized inverse $\alpha^{-1} : [0,1] \longrightarrow \mathbb{N}$ by

$$\alpha^{-1}(u) := \min\left\{ k \in \mathbb{N} : \alpha(k) \leq u \right\} = \sum_{i=0}^{\infty} 1_{\{\alpha(i)>u\}}.$$

Our theoretical results require assumptions on the rate of decay of the mixing coefficients $(\alpha(k))_{k\geq 1}$. We formulate these assumptions using a concept introduced by Rio [21] that is based on the quantile function which we define below.

Definition 2.2. *For a random variable X, the upper quantile function $Q_X : [0,1] \longrightarrow \mathbb{R}$ is defined by*

$$Q_X(u) = \inf\left\{ t \in \mathbb{R} : P(X > t) \leq u \right\}$$

Finally, we will assume that the kernel function $h : \mathbb{R}^2 \to \mathbb{R}$ satisfies the variation condition, which is a continuity assumption introduced by Denker and Keller [12], and that the kernel has uniform $(2+\delta)$-moments.

Definition 2.3. *A kernel $h : \mathbb{R}^2 \to \mathbb{R}$ satisfies the variation condition if there exist constants $L > 0$ and $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$*

$$E\Big(\sup_{\{(x,y):\|(x,y)-(X,Y)\|\leq\varepsilon\}} |h(x,y) - h(X,Y)| \Big) \leq L\,\varepsilon, \tag{2}$$

where X, Y are independent random variables with the same distribution as X_1, and where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^2$.

Definition 2.4. *Let $(X_i)_{i\geq 1}$ be a stationary process. A kernel h has uniform $(2+\delta)$-moments if for all $k \in \mathbb{N}_0$*

$$E|h(X_1, X_k)|^{(2+\delta)} \leq M,$$
$$E|h(X, Y)|^{(2+\delta)} \leq M,$$

where X, Y are independent copies of X_1, and where M is a constant.

The following theorem is the main theoretical result of this paper. Throughout, $Q_{|X|}$ will denote the common quantile function of the X_k's.

Theorem 1. *Let $(X_i)_{i\geq 1}$ be an α-mixing strictly stationary process. Let $h(x, y)$ be a bounded anti-symmetric kernel with uniform $(2+\delta)$-moments and satisfying the variation condition (2). Moreover, assume that there exist constants $p > 2$ and $\varepsilon > 6/\delta$ such that*

$$\int_0^1 (\alpha^{-1}(u))^{4+\varepsilon} Q^p_{|X|}(u)du < \infty. \tag{3}$$

Then, under the null hypothesis H_0, as $n \to \infty$,

$$\frac{\sqrt{2\log\log n}}{\sigma_h} T_n - b_n \xrightarrow{\mathcal{D}} G_2,$$

where G_2 is the Gumbel extreme value distribution with distribution function

$$G_2(x) = \exp(-2\exp(-x)), \tag{4}$$

and where the centering constants b_n and the long-run variance σ_h^2 are defined by

$$b_n = 2\log\log n + \frac{1}{2}\log\log\log n - \frac{1}{2}\log\pi \tag{5}$$

$$\sigma_h^2 = \mathrm{Var}(h_1(X_1) + 2\sum_{k=2}^{\infty} \mathrm{Cov}(h_1(X_1), h_1(X_k)). \tag{6}$$

Here, $h_1(x)$ denotes the first order term in the Hoeffding decomposition of h, as defined below.

The idea of the proof is to apply the Hoeffding decomposition, which was introduced by Hoeffding [15], and to show that the degenerate part is asymptotically negligible. Thus, it will remain to show that the linear part converges to the extreme value distribution G_2. For a two-sample U-statistic, the Hoeffding decomposition of the kernel h is given by

$$h(x, y) = \theta + h_1(x) + h_2(y) + \Psi(x, y), \tag{7}$$

where the terms on the right hand side are defined by

$$\theta = Eh(X, Y)$$

$$\begin{aligned} h_1(x) &= Eh(x,Y) - \theta \\ h_2(y) &= Eh(X,y) - \theta \\ \Psi(x,y) &= h(x,y) - h_1(x) - h_2(y) - \theta, \end{aligned}$$

and where X, Y are two independent random variables with the same distribution as X_1. Note that in our case, since h is assumed to be anti-symmetric, $\theta = 0$ and $h_2(x) = -h_1(x)$. Applying the Hoeffding decomposition of the kernel h to the test statistic T_n, we obtain

$$\begin{aligned} &\frac{1}{\sqrt{k(n-k)n}}\Big|\sum_{i=1}^{k}\sum_{j=k+1}^{n} h(X_i,X_j)\Big| \qquad (8)\\ &= \frac{1}{\sqrt{k(n-k)n}}\Big|\sum_{i=1}^{k}\sum_{j=k+1}^{n} \big(h_1(X_i) - h_1(X_j) + \Psi(X_i,X_j)\big)\Big| \\ &= \frac{1}{\sqrt{k(n-k)n}}\Big|(n-k)\sum_{i=1}^{k} h_1(X_i) - k\sum_{j=k+1}^{n} h_1(X_j) + \sum_{i=1}^{k}\sum_{j=k+1}^{n}\Psi(X_i,X_j)\Big| \\ &= \Big|\sqrt{\frac{n}{k(n-k)}}\Big(\sum_{i=1}^{k} h_1(X_i) - \frac{k}{n}\sum_{j=1}^{n} h_1(X_j)\Big) + \frac{1}{\sqrt{k(n-k)n}}\sum_{i=1}^{k}\sum_{j=k+1}^{n}\Psi(X_i,X_j)\Big| \end{aligned}$$

In order to show that $\frac{\sqrt{\log\log n}}{\sigma_h}T_n - b_n \xrightarrow{\mathcal{D}} G_2$, it thus suffices to show that

$$\sqrt{\log\log n}\max_{1\le k\le n-1}\frac{1}{\sqrt{k(n-k)n}}\Big|\sum_{i=1}^{k}\sum_{j=k+1}^{n}\Psi(X_i,X_j)\Big| \xrightarrow{P} 0, \qquad (9)$$

and that

$$\frac{\sqrt{2\log\log n}}{\sigma_h}\max_{1\le k\le n-1}\sqrt{\frac{n}{k(n-k)}}\Big(\sum_{i=1}^{k} h_1(X_i) - \frac{k}{n}\sum_{j=1}^{n} h_1(X_j)\Big) \xrightarrow{\mathcal{D}} G_2. \qquad (10)$$

The asymptotic negligibility of the degenerate part, i.e. (9), will be established in Proposition 5.1, while (10) will be a consequence of a suitable Darling-Erdős theorem; see Theorem 3 below.

In the next theorem, we investigate the large sample behavior of T_n under the alternative H_1. We define

$$\Delta = \Delta_n := Eh(X_1, X_n'),$$

where X_n' is independent of X_1 and has the same distribution as X_n. Note that Δ_n measures the size of the change in the distribution of X_i at the change point.

Theorem 2. *Assume that the degenerate kernel h has uniform $(2+\delta)$-moments and that it satisfies the variation condition (2). Let*

$$\sum_{k=1}^{\infty} k\left(\int_0^{\alpha(k)} Q_{|X|}(u)du\right)^{\frac{\delta}{3+2\delta}} < \infty. \qquad (11)$$

Moreover, assume that the alternative H_1 holds, and that

$$\sqrt{\frac{k_n^*(n-k_n^*)}{n\log\log n}}|\Delta_n| \longrightarrow \infty, \tag{12}$$

where $k^* = k_n^*$ denotes the location of the change. Then

$$\frac{1}{\sqrt{\log\log n}}T_n \xrightarrow{P} \infty. \tag{13}$$

Corollary 2.5. *The test that rejects the null hypothesis H_0 of no change when*

$$\frac{\sqrt{2\log\log n}}{\sigma_h}T_n - b_n \geq g_{2,\alpha},$$

where $g_{2,\alpha}$ denotes the upper α-th quantile of the Gumbel distribution G_2, has asymptotic level α. Moreover, the test is consistent against any alternatives H_1 that satisfy (12).

Proof. Under the null hypothesis, the test statistic converges by Theorem 1 to the Gumbel extreme value distribution G_2, and thus the test that rejects the null hypothesis when the statistic exceeds $g_{2,\alpha}$ has asymptotically level α. Regarding the behavior under the alternative, we will show that under the assumptions of the corollary

$$\frac{\sqrt{2\log\log n}}{\sigma_h}T_n - b_n \xrightarrow{P} \infty. \tag{14}$$

Let $K > 0$ be a given constant, then

$$\begin{aligned}
&P\Big(\frac{\sqrt{2\log\log n}}{\sigma_h}T_n - b_n \geq K\Big)\\
&= P\Big(T_n \geq \frac{(b_n+K)\sigma_h}{\sqrt{2\log\log n}}\Big)\\
&= P\Big(\frac{1}{\sqrt{\log\log n}}T_n \geq \frac{(b_n+K)\sigma_h/\sqrt{2}}{\log\log n}\Big)\\
&\geq P\Big(\frac{1}{\sqrt{\log\log n}}T_n \geq 3\sigma_h/\sqrt{2}\Big),
\end{aligned}$$

for all n large enough, since $(b_n+K)/\log\log n \to 2$ as $n\to\infty$. Now, the right hand side converges to 1 by Theorem 2. □

Remark 1. *The condition (12) puts restrictions on the time k_n^* as well as the magnitude Δ_n of the change, in order for our test to be consistent. For an early change, i.e. when $k_n^* = o(n)$, the condition (12) is equivalent to*

$$\frac{k_n^*\Delta_n^2}{\log\log n} \to \infty,$$

stating that $k_n^*\cdot\Delta_n^2$ has to grow faster than $\log\log n$ for the weighted test to be consistent. Keeping the magnitude of the change $\Delta_n \equiv \Delta$ constant, this means that k_n^* must grow faster than $\log\log n$.

In the next theorem, we derive the asymptotic distribution of the weighted tied-down random walk. This result will be used in the analysis of the asymptotic distribution of the linear part of the test statistic T_n.

Theorem 3. *Let $(X_i)_{i\geq 1}$ be an α-mixing stationary process, let $S_n := \sum_{i=1}^n X_i$ denote the partial sum process, and let*

$$\sigma^2 = \mathrm{Var}(X_1) + 2\sum_{k=2}^{\infty} \mathrm{Cov}(X_1, X_k) \tag{15}$$

denote the long-run variance. If there exists a $2 < p \leq 3$ such that

$$\int_0^1 (\alpha^{-1}(u))^{p-1} Q^p_{|X|}(u)\, du < \infty, \tag{16}$$

then

$$\frac{\sqrt{2\log\log n}}{\sigma} \max_{1\leq k\leq n-1} \sqrt{\frac{n}{k(n-k)}} \Big| S_k - \frac{k}{n} S_n \Big| - b_n \xrightarrow{\mathcal{D}} G_2,$$

where $G_2(x) = \exp(-2\exp(-x))$, and where b_n is defined as in (5).

The proof of the above theorem follows the ideas of Yao and Davis [28] who showed that for i.i.d. standard normally distributed data the likelihood ratio converges in distribution to a Gumbel extreme value distribution. An important tool in the proof is the celebrated Darling-Erdős theorem on the asymptotic distribution of $\max_{1\leq k\leq n} \frac{1}{\sqrt{k}} |\sum_{i=1}^k X_i|$. Theorem 4 below establishes such a result for dependent data. Such theorems have been proved before, see e.g. Shorack [25], but not under the conditions required in the present paper.

Theorem 4. *Let $(X_i)_{i\geq 1}$ be an α-mixing strictly stationary process satisfying (16) for some $2 < p \leq 3$, and let $S_n := \sum_{i=1}^n X_i$ denote the partial sum process. Then*

$$\frac{\sqrt{2\log\log n}}{\sigma} \max_{1\leq k\leq n} \frac{|S_k|}{\sqrt{k}} - b_n \xrightarrow{\mathcal{D}} G, \tag{17}$$

where $G(x) = \exp(-\exp(-x))$ denotes the Gumbel extreme value distribution function, and where σ and b_n is defined in (15) and (5), respectively.

The proof of Theorem 4, presented in Section 5.4 below, follows the ideas of Shorack [25] who proved that an almost sure invariance principle together with a suitable maximal inequality implies the Darling-Erdős theorem.

3. Simulations

In this section we present some simulation results for the weighted and the unweighted test statistic. We compare the power, the empirical size and the critical values, and consider the CUSUM and the Wilcoxon kernel, namely $h(x, y) = y - x$ and $h(x, y) = 1_{\{x<y\}} - \frac{1}{2}$.

Remark 2. *It is easy to see that the Wilcoxon kernel satisfies the variation condition. Let* $h(x,y) = 1_{\{x \le y\}} - \frac{1}{2}$. *Then, for* $\varepsilon > 0$,

$$\begin{aligned}
&\sup_{\|(x,y)-(X,Y)\| \le \varepsilon} |h(x,y) - h(x',y')| \\
&= \sup_{\|(x,y)-(X,Y)\| \le \varepsilon} |1_{\{x \le y\}} - 1_{\{x' \le y'\}}| \\
&= \begin{cases} 1 & \text{for } X - Y \in (-\sqrt{2}\varepsilon, \sqrt{2}\varepsilon], \\ 0 & \text{else.} \end{cases}
\end{aligned}$$

If the density of $X - Y$ *is bounded, then*

$$\begin{aligned}
&E\Big(\sup_{\|(x,y)-(X,Y)\| \le \varepsilon} |1_{\{x \le y\}} - 1_{\{x' \le y'\}}|\Big) \\
&\le P(X - Y \in (-\sqrt{2}\varepsilon, \sqrt{2}\varepsilon]) \le L\varepsilon.
\end{aligned}$$

Let T_n^C denote the CUSUM and T_n^{WC} the weighted CUSUM test statistic and let T_n^W and T_n^{WW} denote the Wilcoxon and the weighted Wilcoxon test statistic, all properly centered and normalized, i.e.

$$\begin{aligned}
T_n^C &= \frac{1}{n^{3/2}\sigma_C} \max_{1 \le k \le n-1} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} (X_j - X_i) \right|, \\
T_n^{WC} &= \frac{\sqrt{2 \log\log n}}{\sigma_C} \max_{1 \le k \le n-1} \frac{1}{\sqrt{k(n-k)n}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} (X_j - X_i) \right| - b_n, \\
T_n^W &= \frac{1}{n^{3/2}\sigma_W} \max_{1 \le k \le n-1} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(1_{\{X_i < X_j\}} - \frac{1}{2} \right) \right|, \\
T_n^{WW} &= \frac{\sqrt{2 \log\log n}}{\sigma_W} \max_{1 \le k \le n-1} \frac{1}{\sqrt{k(n-k)n}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(1_{\{X_i < X_j\}} - \frac{1}{2} \right) \right| - b_n.
\end{aligned}$$

For $0 < \alpha < 1$ we define $c_i(\alpha),\ i \in \{1,2,3,4\}$, such that

$$\begin{aligned}
P\left(T_n^C > c_1(\alpha)\right) = \alpha, \quad P\left(T_n^{WC} > c_2(\alpha)\right) = \alpha, \\
P\left(T_n^W > c_3(\alpha)\right) = \alpha, \quad P\left(T_n^{WW} > c_4(\alpha)\right) = \alpha.
\end{aligned}$$

Most of the simulation study is based on i.i.d. standard normally distributed observations. In this case

$$\sigma_C^2 = 1 \text{ and } \sigma_W^2 = \frac{1}{12}.$$

In Figure 3 we we will consider dependent observations. In this case the long run variance has to be estimated. We use a subsampling estimator introduced

by Dehling, Fried, Sharipov, Vogel and Wornowizki [8]. To achieve consistency under the alternative, we split the data into three disjoint sub-sequences of similar length and use the median of the resulting three separate estimations, see Dehling, Fried and Wendler [9].

First, we have simulated the critical values $c_i(\alpha)$ and compared them to the asymptotic ones. The results are summarized in Table 1. For the unweighted test statistics the simulated critical values are almost in agreement with the asymptotic ones, whereas for the weighted test statistics the difference is larger. An overview is also given in Figure 1, which shows the different empirical distribution functions compared to the asymptotic ones. On the left hand side the empirical distribution function for the CUSUM and Wilcoxon test statistic is compared to the distribution function of the Kolmogorov-Smirnov distribution, and on the right hand side the distribution functions for the weighted CUSUM and weighted Wilcoxon test statistics are compared to the distribution function of the Gumbel distribution with location parameter log(2) and scale parameter 1.

Table 1

Empirical and asymptotic critical values for the CUSUM (C)*, weighted CUSUM* (WC)*, Wilcoxon* (W) *and weighted Wilcoxon* (WW) *test statistics. The empirical critical values are based on i.i.d. standard normally distributed observations and 20000 simulation runs.*

		$\alpha = 0.05$					$\alpha = 0.10$				
	n	100	200	400	800	∞	100	200	400	800	∞
C	$c_1(\alpha)$	1.30	1.33	1.33	1.33	1.36	1.17	1.18	1.20	1.20	1.22
WC	$c_2(\alpha)$	2.64	2.72	2.76	2.82	3.66	2.20	2.26	2.28	2.32	2.94
W	$c_3(\alpha)$	1.30	1.32	1.33	1.33	1.36	1.17	1.19	1.20	1.21	1.22
WW	$c_4(\alpha)$	2.47	2.56	2.60	2.65	3.66	2.05	2.13	2.18	2.18	2.94

Next, we evaluate the performance of the test statistics by computing the empirical sizes and the power. Table 2 presents the empirical sizes for the unweighted and weighted CUSUM and Wilcoxon test statistics. The empirical sizes are lower than the nominal size in all cases considered. For the unweighted test statistics the size distortion shrinks to zero as the sample size increases, whereas for the weighted test statistics the difference between the empirical size and the nominal size is much larger for all considered sample sizes.

Figure 2 shows the size-corrected power of the CUSUM, weighted CUSUM and the Wilcoxon and weighted Wilcoxon test statistic for all possible change-point times with shift height $\Delta = 0.3$. The simulations are based on i.i.d. standard normally distributed observations. It is clear that the power of all test statistics improves as the change-point is more central. In the case of early and late changes the weighted test statistics have better power than the unweighted ones, whereas they have less power for changes in the middle of the time period. For standard normally distributed data the power of the CUSUM and the Wilcoxon test statistics differs only marginally. Considering data with heavier tails, such as $t(5)$ or $t(3)$ distributed observations, one can see that the CUSUM test statistics lose more power, especially the weighted CUSUM. In both cases,

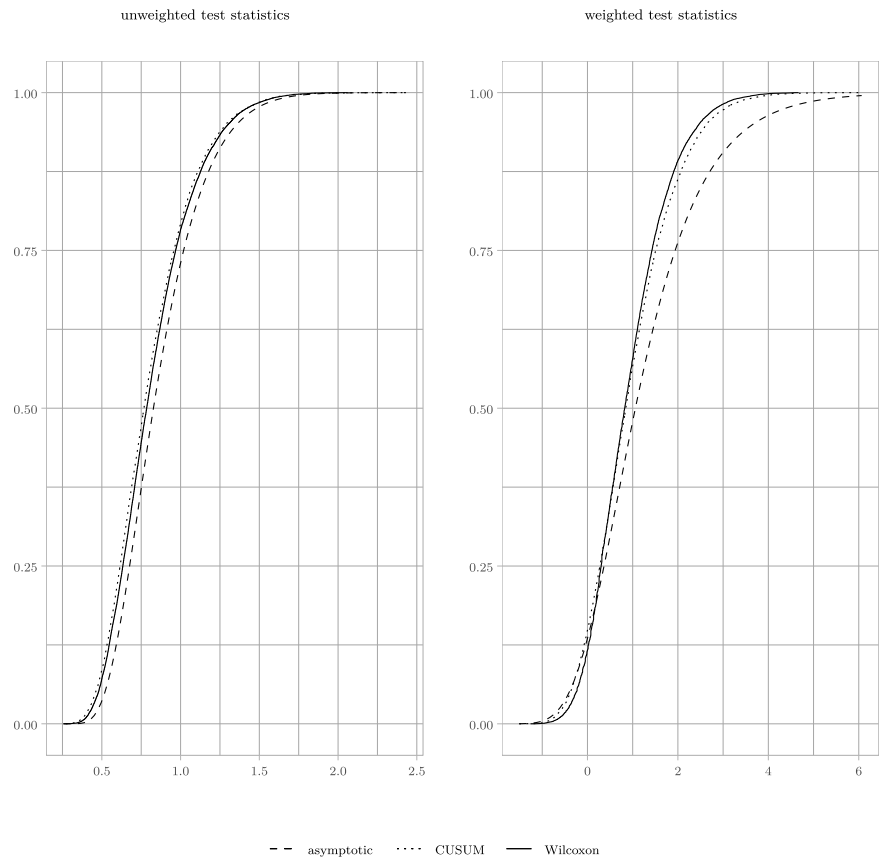

FIG 1. *Empirical distribution functions under the hypothesis for the unweighted and weighted CUSUM and Wilcoxon test statistics, compared to the distribution functions of the asymptotic distributions. The simulations are based on 100 i.i.d. standard normally distributed observations and 20000 runs.*

the Wilcoxon test statistics have better power.

TABLE 2
Empirical size based on n i.i.d. standard normally distributed observations and 5000 simulation runs.

n	α	T_n^C	T_n^{WC}	T_n^W	T_n^{WW}
200	0.05	0.039	0.010	0.042	0.006
400	0.05	0.046	0.010	0.046	0.010
800	0.05	0.049	0.014	0.047	0.012
1600	0.05	0.045	0.015	0.045	0.015
200	0.10	0.085	0.035	0.088	0.023
400	0.10	0.085	0.039	0.085	0.039
800	0.10	0.095	0.043	0.091	0.034
1600	0.10	0.091	0.043	0.091	0.043

Figure 3 shows analogous power plots, based on AR(1)-processes with $t(3)$-distributed innovations and correlation coefficients $\rho = 0.3, 0.5, 0.7$. It is easy to see that a higher correlation results in a lower power. Again, in the case of early and late changes the weighted test statistics have better power than the unweighted ones. Compared to the CUSUM test statistics, the Wilcoxon test statistics always have greater power.

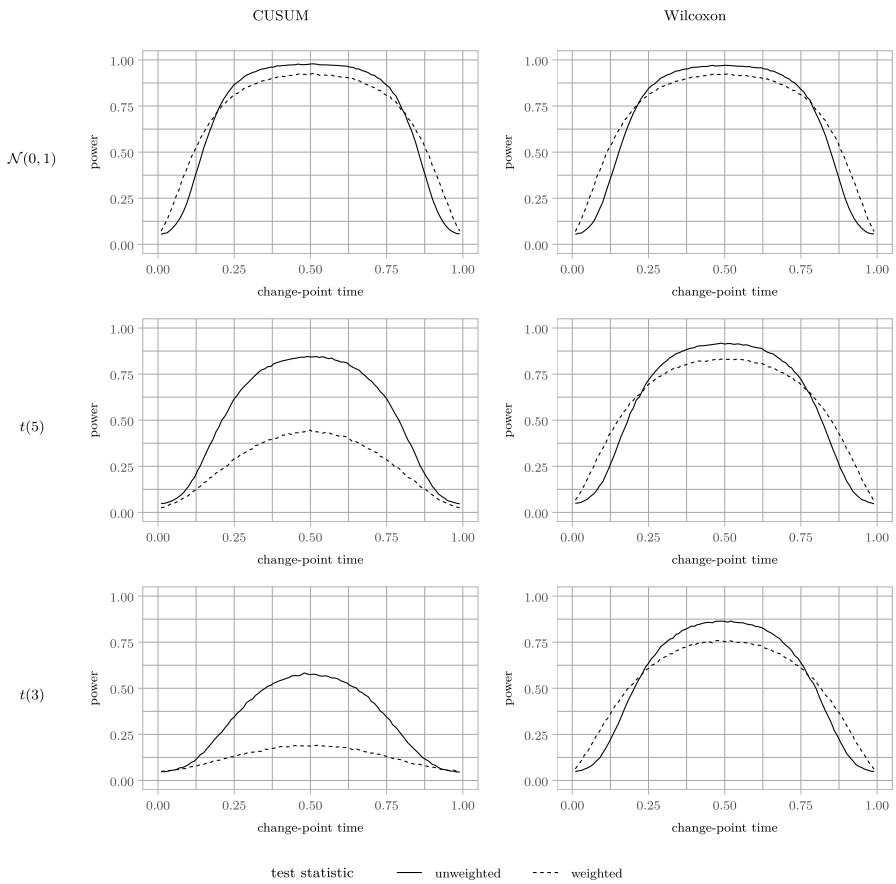

FIG 2. *Size-corrected power depending on the change-point time* $\tau = [\frac{k^*}{n}]$. *The simulations are based on 800 i.i.d. standard normally distributed observations with shift height* $\Delta = 0.3$ *and 20000 simulation runs.*

4. Data example

As an application we analyze the daily absolute log returns of the closing Wirecard stock prices (currency in EUR, downloaded from https://de.finance.yahoo.com/quote/WDI.DU/history?p=WDI.DU on June 14, 2021). We consider

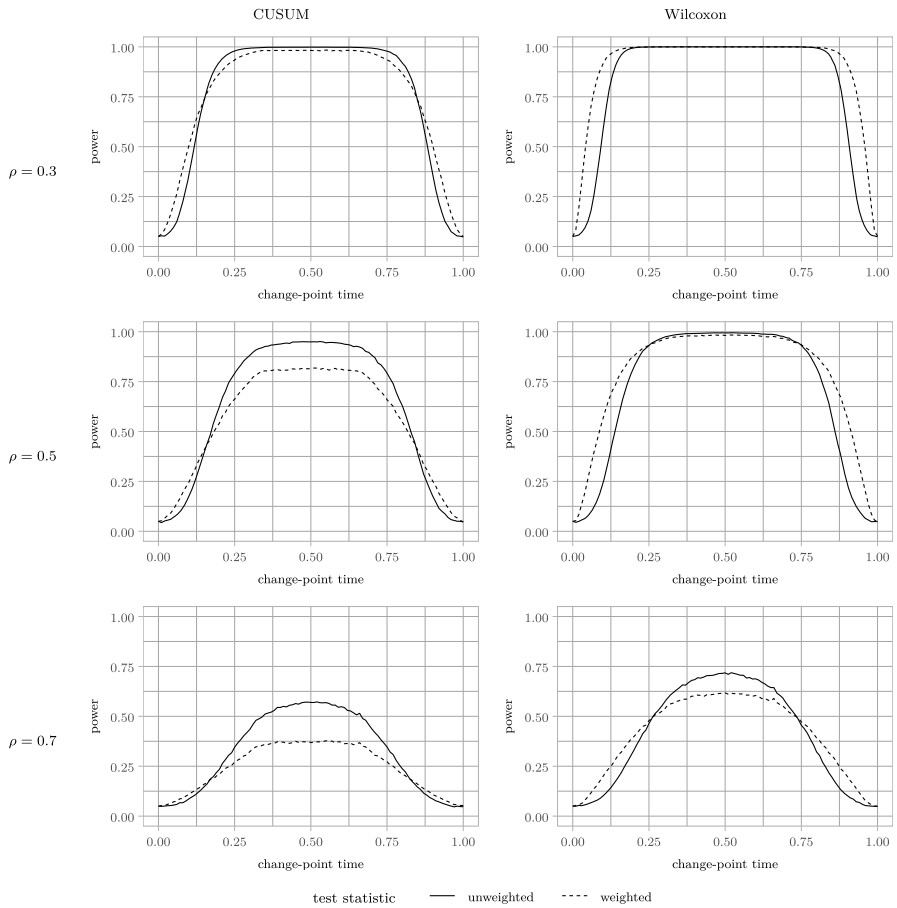

FIG 3. *Size-corrected power depending on the change-point time* $\tau = [\frac{k^*}{n}]$. *The data is generated from an AR(1) process with* $t(3)$*-distributed innovations and correlation coefficients* $\rho = 0.3, 0.5, 0.7$. *The simulations are based on 20000 runs with sample size* $n = 800$ *and shift height* $\Delta = 1$.

the time period February 10, 2020, to June 26, 2020 , which is 19 weeks and 95 observations (trading time from monday to friday). The absolute log returns in this time period are visualized in Figure 4.

As the focus of this paper is on robust tests, we apply the weighted and non weighted Wilcoxon test to the data. The long run variance is estimated by the same subsampling estimator used in the simulation study. Considering a significance level of 5%, the weighted Wilcoxon test statistic rejects the null hypothesis of a constant mean in the absolute log returns. In contrast, the unweighted Wilcoxon test doesn't detect any significant change (c.f. Table 3). The weighted Wilcoxon test detects a change at observation 88, which correspond to June 18, 2020. On that day Wirecard reported that, following an audit by

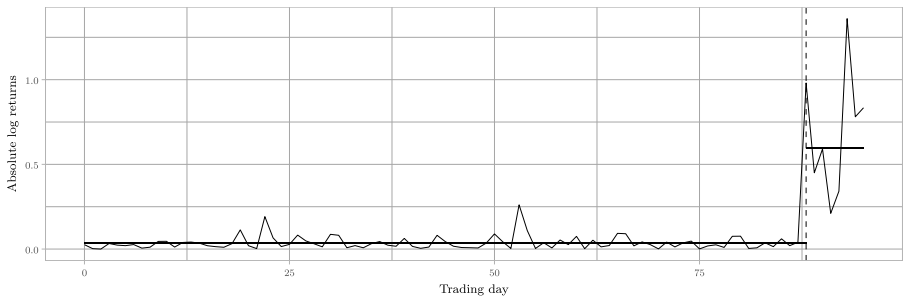

FIG 4. *Absolute log returns of the Wirecard stock price (Febraury 10, 2020 – June 26, 2020). Change-point detected by the weighted Wilcoxon test (vertical dashed line), and average between the detected change-point (horizontal black lines).*

Ernst & Young, about 1.9 billion euro was missing in certain trust accounts.

TABLE 3
Unweigted and weighted Wilcoxon test statistics together with the corresponding asymptotic critical values for the Wirecard stock price data, considering a significance level of 5%.

	Wilcoxon	weighted Wilcoxon
Test statistic	1.28	4.96
Critical value	1.36	3.66

5. Proofs

5.1. Proof of Theorem 1

First, we show in the following proposition that the degenerate part is asymptotically negligible.

Proposition 5.1. *Let Ψ be the kernel given by Hoeffding's decomposition of h in (7), satisfying the variation condition. Moreover, we assume that (11) holds. Then, under the null hypothesis H_0, as $n \to \infty$*

$$\max_{1\le k\le n-1} \sqrt{\frac{\log\log n}{k(n-k)n}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} \Psi(X_i, X_j) \right| \xrightarrow{\mathcal{P}} 0.$$

Proof. We can split the maximum into the stretch up to $\sqrt{n}$, the stretch between $\sqrt{n}$ and $n-\sqrt{n}$ and the stretch after $n-\sqrt{n}$, such that

$$\max_{1\le k\le n-1} \sqrt{\frac{\log\log n}{k(n-k)n}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} \Psi(X_i, X_j) \right|$$

$$\leq 2 \max_{\sqrt{n}\leq k<n-\sqrt{n}} \sqrt{\frac{\log\log n}{n^{5/2}}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} \Psi(X_i, X_j) \right| \tag{18}$$

$$+ \max_{k<\sqrt{n}} \sqrt{\frac{\log\log n}{n^2}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} \Psi(X_i, X_j) \right| \tag{19}$$

$$+ \max_{k\geq n-\sqrt{n}} \sqrt{\frac{\log\log n}{n^2}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} \Psi(X_i, X_j) \right|. \tag{20}$$

Now we can deal with every single maximum. To show that these maxima converge in probability to 0, we use Theorem A of Serfling [24]. To apply that theorem, we need a functional $g(F_{a,n})$ depending on the joint distribution of a vector $(Y_{a+1}, \ldots, Y_{a+n})$ of n random variables, and satisfying

$$g(F_{a,k}) + g(F_{a+k,l}) \leq g(F_{a,k+l}),\ 1 \leq k < k+l,$$

such that

$$E|S_{a,l}|^2 \leq g(F_{a,l}),\ \text{for all } l \geq 1,$$

where

$$S_{a,l} = \sum_{i=a+1}^{a+l} Y_i.$$

Then it follows

$$E\left(\max_{1\leq i\leq l} |S_{a,i}|\right)^2 \leq \left(\frac{\log(2l)}{\log(2)}\right)^2 (g(F_{a,l}))^2.$$

We set $V_k = \sum_{i=1}^{k} \sum_{j=k+1}^{n} \Psi(X_i, X_j)$. By Lemma A.3

$$E(V_m - V_k)^2 \leq C_1|m-k|n,$$

where C_1 is a constant. Define $Y_i := V_i - V_{i-1},\ i = 1, \ldots, n$ and set $V_0 = 0$. Then one has $V_m = \sum_{i=1}^{m} Y_i$ and $V_m - V_k = \sum_{i=k+1}^{m} Y_i$. As $k \leq m$ we can set $m := k+l,\ l \in \mathbb{N}$. Thus

$$E\left(\sum_{i=k+1}^{k+l} Y_i\right)^2 \leq C_1 ln.$$

With

$$g(F_{a,l}) = C_1 ln,$$

where C_1 is a constant, the required conditions are satisfied. We have

$$g(F_{a,k}) + g(F_{a+k,l}) = C_1 kn + C_1 ln = C_1 n(k+l) = g(F_{a,k+l})$$

and

$$E|S_{a,l}|^2 = E\left(\sum_{i=a+1}^{a+l} Y_i\right) \leq C_1 ln = g(F_{a,l}).$$

Thus

$$E\left(\max_{1\leq j\leq l}\left|\sum_{i=k+1}^{k+j} Y_i\right|\right)^2 \leq \left(\frac{\log(2l)}{\log(2)}\right)^2 C_1 ln.$$

For $k=0$ and $l=\sqrt{n}$ we get

$$E\left(\max_{1\leq j\leq\sqrt{n}}|V_j|\right)^2 = E\left(\max_{1\leq j\leq\sqrt{n}}\left|\sum_{i=1}^{j} Y_i\right|\right)^2 \leq \left(\frac{\log\left(2\sqrt{n}\right)}{\log(2)}\right)^2 C_1 n^{3/2}.$$

Thus we have

$$E\left(\max_{1\leq k\leq\sqrt{n}}\sqrt{\frac{\log\log n}{n^2}}|V_k|\right)^2 \leq \frac{\log\log n}{\sqrt{n}}\left(\frac{\log\left(2\sqrt{n}\right)}{\log(2)}\right)^2 C_1.$$

This goes to 0 for $n\to\infty$. Applying Chebyshev's inequality, we obtain that (19) converges to 0 in probability as $n\to\infty$. By stationarity, this also holds for (20). An analogous procedure leads to

$$E\left(\max_{1\leq k\leq n-\sqrt{n}}\sqrt{\frac{\log\log n}{n^{5/2}}}|V_k|\right)^2 \leq \frac{(n-\sqrt{n})\log\log n}{n^{3/2}}\left(\frac{\log\left(2\sqrt{n}\right)}{\log(2)}\right)^2 C_1.$$

As this goes to zero for $n\to\infty$ and as

$$\max_{\sqrt{n}\leq k\leq n-\sqrt{n}}\sqrt{\frac{\log\log n}{n^{5/2}}}|V_k| \leq \max_{1\leq k\leq n-\sqrt{n}}\sqrt{\frac{\log\log n}{n^{5/2}}}|V_k|,$$

we obtain that (18) converges in probability to 0. □

Proof of Theorem 1. In the same way as in (8), we apply Hoeffding's decomposition to the kernel $h(x,y)$ and get

$$\begin{aligned}&\frac{1}{\sqrt{k(n-k)n}}\sum_{i=1}^{k}\sum_{j=k+1}^{n} h(X_i,X_j)\\ =&\sqrt{\frac{n}{k(n-k)}}\left(\sum_{i=1}^{k} h_1(X_i) - \frac{k}{n}\sum_{i=1}^{n} h_1(X_i)\right) + \frac{1}{\sqrt{k(n-k)n}}\sum_{i=1}^{k}\sum_{j=k+1}^{n}\Psi(X_i,X_j).\end{aligned}$$

As the variation condition holds for the kernel h, it also holds for Ψ. In order to be able to apply Proposition 5.1, we need to verify assumption (11). From (3), we obtain

$$\int_0^1 (\alpha^{-1}(u))^{4+\varepsilon}\, Q_{|X|}(u)du < \infty,$$

which is equivalent to

$$\sum_{k=1}^{\infty} k^{3+\varepsilon} \int_0^{\alpha(k)} Q_{|X|}(u)du < \infty;$$

see Remark 2.2 in Merlevède and Rio [17] or Annex C in Rio [22]. Now, we define

$$q_k := \int_0^{\alpha(k)} Q_{|X|}(u)du.$$

As q_k is nonincreasing and by assumption, we obtain

$$n^{4+\varepsilon} q_n \le c_1 \sum_{k=1}^{n} k^{3+\varepsilon} q_n \le c_1 \sum_{k=1}^{n} k^{3+\varepsilon} q_k \le c_1 \sum_{k=1}^{\infty} k^{3+\varepsilon} q_k < \infty,$$

where c_1 is a constant. Hence $n^{\frac{(4+\varepsilon)\delta}{3+2\delta}} q_n^{\frac{\delta}{3+2\delta}}$ is bounded by some constant c_2, and thus

$$n q_n^{\frac{\delta}{3+2\delta}} \le c_2 \frac{1}{n^{1+\frac{\varepsilon\delta-6}{3+2\delta}}},$$

which implies

$$\sum_{k=1}^{\infty} k q_k^{\frac{\delta}{3+2\delta}} < \infty.$$

Thus, all conditions of Proposition 5.1 are satisfied and we can conclude that

$$\max_{1\le k\le n-1} \sqrt{\frac{\log\log n}{k(n-k)n}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} \Psi(X_i, X_j) \right| \xrightarrow{\mathcal{P}} 0.$$

Hence, by Slutsky's theorem, it remains to show that

$$\max_{1\le k\le n-1} \sqrt{\frac{n\log\log n}{k(n-k)}} \left| \sum_{i=1}^{k} h_1(X_i) - \frac{k}{n} \sum_{i=1}^{n} h_1(X_i) \right| - b_n$$

converges in distribution to the desired extreme value distribution. This follows from Theorem 3 with $S_k = \sum_{i=1}^{k} h_1(X_i)$. □

5.2. Proof of Theorem 2

Without loss of generality, we assume that $\Delta > 0$. Since

$$\max_{1\le k\le n-1} \frac{1}{\sqrt{k(n-k)n}} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} h(X_i, X_j) \right| \ge \frac{1}{\sqrt{k^*(n-k^*)n}} \left| \sum_{i=1}^{k^*} \sum_{j=k^*+1}^{n} h(X_i, X_j) \right|,$$

it suffices to show that

$$\frac{1}{\sqrt{(\log\log n)k^*(n-k^*)n}}\left|\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{n} h(X_i,X_j)\right| \xrightarrow{\mathcal{P}} \infty.$$

We obtain the following Hoeffding decomposition

$$h(X_i,X_j)=\Delta+h_1(X_i)+h_2(X_j)+\Psi(X_i,X_j), \quad 1\le i\le k^*<j\le n,$$

where Δ is given as in Theorem 2, and where

$$\begin{aligned} h_1(x) &:= Eh(x,X_n)-\Delta,\\ h_2(y) &:= Eh(X_1,y)-\Delta. \end{aligned}$$

It holds

$$\begin{aligned}\frac{T_n}{\sqrt{\log\log n}} \ge& \frac{1}{\sqrt{(\log\log n)k^*(n-k^*)n}}\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{n}(\Delta+h_1(X_i)+h_2(X_j)+\Psi(X_i,X_j))\\ =&\Delta\sqrt{\frac{k^*(n-k^*)}{(\log\log n)n}}+\sqrt{\frac{n-k^*}{(\log\log n)k^*n}}\sum_{i=1}^{k^*}h_1(X_i)\\ &+\sqrt{\frac{k^*}{(\log\log n)(n-k^*)n}}\sum_{j=k^*+1}^{n}h_2(X_j)\\ &+\frac{1}{\sqrt{(\log\log n)k^*(n-k^*)n}}\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{n}\Psi(X_i,X_j). \end{aligned}$$

By the law of iterated logarithm

$$\begin{aligned}\sqrt{\frac{n-k^*}{k^*n\log\log n}}\left|\sum_{i=1}^{k^*}h_1(X_i)\right| \le& \frac{1}{\sqrt{k^*\log\log n}}\left|\sum_{i=1}^{k^*}h_1(X_i)\right| = O(1)\ a.s.,\\ \sqrt{\frac{k^*}{(n-k^*)n\log\log n}}\left|\sum_{j=k^*+1}^{n}h_2(X_j)\right| \le& \frac{1}{\sqrt{(n-k^*\log\log n)}}\left|\sum_{j=k^*+1}^{n}h_2(X_j)\right|\\ \overset{\mathcal{D}}{=}& \frac{1}{\sqrt{(n-k^*)\log\log n}}\left|\sum_{j=1}^{n-k^*}h_2(X_{j+k^*})\right|\\ =& O_p(1). \end{aligned}$$

It remains to show that

$$\frac{1}{\sqrt{(\log\log n)k^*(n-k^*)n}}\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{n}\Psi(X_i,X_j) \xrightarrow{\mathcal{P}} 0,$$

as $n \to \infty$. By Lemma A.3 we have

$$E\left(\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{n}\Psi(X_i,X_j)\right)^2 \le C(n-k^*)k^*,$$

for some constant C. We get

$$E\left(\frac{1}{\sqrt{(\log\log n)k^*(n-k^*)n}}\sum_{i=1}^{k^*}\sum_{j=k^*+1}^{n}\Psi(X_i,X_j)\right)^2 \le \frac{C}{(\log\log n)n} \to 0,$$

as $n \to \infty$, which completes the proof.

5.3. Proof of Theorem 3

For the proof of Theorem 3, we need the following lemmas, which all hold under the assumptions of Theorem 3.

Lemma 5.2.

$$\max_{1\le k\le\left[\frac{n}{\log n}\right]}\sqrt{\frac{n}{k(n-k)}}\left|S_k-\frac{k}{n}S_n\right| - \max_{1\le k\le\left[\frac{n}{\log n}\right]}\left|\frac{S_k}{\sqrt{k}}\right| = o_p\left(\frac{1}{\sqrt{2\log\log n}}\right)$$

Proof. For n large, we have

$$\begin{aligned}
&\left|\max_{1\le k\le\left[\frac{n}{\log n}\right]}\sqrt{\frac{n}{k(n-k)}}\left|S_k-\frac{k}{n}S_n\right| - \max_{1\le k\le\left[\frac{n}{\log n}\right]}\left|\frac{S_k}{\sqrt{k}}\right|\right| \\
&\le \max_{1\le k\le\left[\frac{n}{\log n}\right]}\left|\sqrt{\frac{n}{k(n-k)}}\left|S_k-\frac{k}{n}S_n\right| - \left|\frac{S_k}{\sqrt{k}}\right|\right| \\
&\le \max_{1\le k\le\left[\frac{n}{\log n}\right]}\left|\sqrt{\frac{n}{k(n-k)}}\left(S_k-\frac{k}{n}S_n\right) - \frac{S_k}{\sqrt{k}}\right| \\
&= \max_{1\le k\le\left[\frac{n}{\log n}\right]}\left|\left(\sqrt{\frac{n}{(n-k)}}-1\right)\frac{S_k}{\sqrt{k}} - \sqrt{\frac{k}{n-k}}\frac{S_n}{\sqrt{n}}\right| \\
&\le \max_{1\le k\le\left[\frac{n}{\log n}\right]}\left(\left|\left(\sqrt{\frac{n}{(n-k)}}-1\right)\frac{S_k}{\sqrt{k}}\right| + \left|\sqrt{\frac{k}{n-k}}\frac{S_n}{\sqrt{n}}\right|\right) \qquad (21)
\end{aligned}$$

For large n and $1 \le k \le \left[\frac{n}{\log n}\right]$ the inequalities

$$\sqrt{\frac{n}{n-k}}-1 \le \frac{k}{n} \text{ and } \sqrt{\frac{k}{n-k}} \le 2\sqrt{\frac{k}{n}}$$

are satisfied. So we have that (21) is less or equal to

$$\max_{1\le k\le\left[\frac{n}{\log n}\right]}\left(\left|\frac{k}{n}\frac{S_k}{\sqrt{k}}\right| + \left|2\sqrt{\frac{k}{n}}\frac{S_n}{\sqrt{n}}\right|\right)$$

$$
\begin{aligned}
&= \max_{1\le k\le\left[\frac{n}{\log n}\right]} \sqrt{\frac{k}{n}}\left(\left|\frac{S_k}{\sqrt{n}}\right| + 2\left|\frac{S_n}{\sqrt{n}}\right|\right)\\
&\le \frac{1}{\sqrt{\log n}} \max_{1\le k\le\left[\frac{n}{\log n}\right]} \left(\left|\frac{S_k}{\sqrt{n}}\right| + 2\left|\frac{S_n}{\sqrt{n}}\right|\right)\\
&= O_p\left(\frac{1}{\sqrt{\log n}}\right),
\end{aligned}
$$

where the last equality holds, as

$$
\max_{1\le k\le\left[\frac{n}{\log n}\right]} \left(\left|\frac{S_k}{\sqrt{n}}\right| + 2\left|\frac{S_n}{\sqrt{n}}\right|\right) = O_p(1),
$$

since $\max_{1\le k\le\left[\frac{n}{\log n}\right]}\left|\frac{S_k}{\sqrt{n}}\right|$ and $\left|\frac{S_n}{\sqrt{n}}\right|$ converge in distribution. Now the claim follows, as $(\frac{1}{\sqrt{\log n}})/(\frac{1}{\sqrt{2\log\log n}}) \to 0$, for $n\to\infty$. □

Lemma 5.3.

$$
\max_{\left[\frac{n}{\log n}\right]\le k\le\left[\frac{n}{2}\right]} \sqrt{\frac{n}{k(n-k)}}\left|S_k - \frac{k}{n}S_n\right| = O_p\left(\sqrt{\log\log\log n}\right)
$$

Proof. Since

$$
\begin{aligned}
&\max_{\left[\frac{n}{\log n}\right]\le k\le\left[\frac{n}{2}\right]} \sqrt{\frac{n}{k(n-k)}}\left|S_k - \frac{k}{n}S_n\right|\\
\le &\max_{\left[\frac{n}{\log n}\right]\le k\le\left[\frac{n}{2}\right]} 2\left|\frac{S_k}{\sqrt{k}}\right| + \max_{\left[\frac{n}{\log n}\right]\le k\le\left[\frac{n}{2}\right]}\left|\frac{S_n}{\sqrt{n}}\right|
\end{aligned}
$$

and the second summand is $O_p(1)$, it suffices to show that

$$
\max_{\left[\frac{n}{\log n}\right]\le k\le\left[\frac{n}{2}\right]}\left|\frac{S_k}{\sqrt{k}}\right| = O_p\left(\sqrt{\log\log\log n}\right).
$$

By the almost sure invariance principle of Merlevède and Rio (2012), one can find a Brownian motion $(W_t)_{t\ge 0}$ such that

$$
\frac{|S_t - W_t|}{\sqrt{t}} = O\left(t^{-\lambda}\right) \ \ a.s.,
$$

for some $\lambda > 0$. Hence

$$
\max_{\left[\frac{n}{\log n}\right]\le k\le\left[\frac{n}{2}\right]}\left|\frac{S_k - W_k}{\sqrt{k}}\right| = O\left(\left(\frac{\log n}{n}\right)^{\lambda}\right) \ \ a.s.,
$$

and thus

$$
\max_{\left[\frac{n}{\log n}\right]\le k\le\left[\frac{n}{2}\right]}\left|\frac{S_k}{\sqrt{k}}\right| \le \max_{\left[\frac{n}{\log n}\right]\le k\le\left[\frac{n}{2}\right]}\left|\frac{W_k}{\sqrt{k}}\right| + \max_{\left[\frac{n}{\log n}\right]\le k\le\left[\frac{n}{2}\right]}\left|\frac{S_k - W_k}{\sqrt{k}}\right|. \tag{22}
$$

Set $k = nr,\ r \in (0,1)$ and consider the first summand of the right-hand side of (22). We have

$$\max_{\left[\frac{n}{\log n}\right] \le nr \le \left[\frac{n}{2}\right]} \left| \frac{W_{nr}}{\sqrt{nr}} \right| \stackrel{\mathcal{D}}{=} \max_{\left[\frac{n}{\log n}\right] \le nr \le \left[\frac{n}{2}\right]} \left| \frac{W_r}{\sqrt{r}} \right| \le \max_{\frac{1}{\log n} \le r \le \frac{1}{2}} \left| \frac{W_r}{\sqrt{r}} \right|.$$

By the law of the iterated logarithm, it holds for $r \to 0$

$$|W_r| = O_p\left(\sqrt{2r \log\log r^{-1}}\right).$$

It follows that

$$\max_{\frac{1}{\log n} \le r \le \frac{1}{2}} \left| \frac{W_r}{\sqrt{r}} \right| = O_p\left(\sqrt{\log\log\log n}\right),$$

which completes the proof. □

For abbreviation, we define $a_n := \sqrt{2 \log\log n}$. Recall the definition of b_n (5).

Lemma 5.4.

$$\lim_{n\to\infty} P\left(\frac{a_n}{\sigma} \max_{1 \le k \le \left[\frac{n}{\log n}\right]} \left| \frac{S_k}{\sqrt{k}} \right| - b_n \le x \right) = \exp(-\exp(-x))$$

Proof. By Theorem 4, one has

$$\frac{a_{\left[\frac{n}{\log n}\right]}}{\sigma} \max_{1 \le k \le \left[\frac{n}{\log n}\right]} \frac{|S_k|}{\sqrt{k}} - b_{\left[\frac{n}{\log n}\right]} \xrightarrow{\mathcal{D}} G.$$

Since

$$\frac{a_n}{a_{\left[\frac{n}{\log n}\right]}} \to 1, \quad b_{\left[\frac{n}{\log n}\right]} \frac{a_n}{a_{\left[\frac{n}{\log n}\right]}} - b_n \to 0,$$

as $n \to \infty$, we obtain with Slutsky's Theorem

$$\begin{aligned} &\frac{a_n}{\sigma} \max_{1 \le k \le \left[\frac{n}{\log n}\right]} \frac{|S_k|}{\sqrt{k}} - b_n \\ &= \frac{a_n}{a_{\left[\frac{n}{\log n}\right]}} \left(\frac{a_{\left[\frac{n}{\log n}\right]}}{\sigma} \max_{1 \le k \le \left[\frac{n}{\log n}\right]} \frac{|S_k|}{\sqrt{k}} - b_{\left[\frac{n}{\log n}\right]} \right) + b_{\left[\frac{n}{\log n}\right]} \frac{a_n}{a_{\left[\frac{n}{\log n}\right]}} - b_n \xrightarrow{\mathcal{D}} G. \end{aligned}$$

□

Lemma 5.5.

$$\max_{1 \le k \le \left[\frac{n}{2}\right]} \sqrt{\frac{n}{k(n-k)}} \left| S_k - \frac{k}{n} S_n \right| - \max_{1 \le k \le \left[\frac{n}{\log n}\right]} \left| \frac{S_k}{\sqrt{k}} \right| = o_p\left(\frac{1}{\sqrt{2 \log\log n}} \right)$$

Proof. Since $\frac{1}{a_n^2} \to 0$, we can conclude from Lemma 5.4 and Slutsky's Theorem

$$\frac{1}{a_n^2}\left(a_n \max_{1\le k\le\left[\frac{n}{\log n}\right]} \frac{|S_k|}{\sqrt{k}} - b_n\right) \xrightarrow{\mathcal{P}} 0.$$

As $\frac{b_n}{a_n^2} \to 1$, we get

$$\frac{1}{a_n} \max_{1\le k\le\left[\frac{n}{\log n}\right]} \frac{|S_k|}{\sqrt{k}} \xrightarrow{\mathcal{P}} 1.$$

By Lemma 5.2, this implies

$$\frac{1}{a_n} \max_{1\le k\le\left[\frac{n}{\log n}\right]} \sqrt{\frac{n}{k(n-k)}}\left|S_k - \frac{k}{n}S_n\right| \xrightarrow{\mathcal{P}} 1.$$

By Lemma 5.3, we have

$$\frac{1}{a_n} \max_{\left[\frac{n}{\log n}\right]\le k\le\left[\frac{n}{2}\right]} \sqrt{\frac{n}{k(n-k)}}\left|S_k - \frac{k}{n}S_n\right| \xrightarrow{\mathcal{P}} 0.$$

Thus

$$P\left(\max_{\left[\frac{n}{\log n}\right]\le k\le\left[\frac{n}{2}\right]} \sqrt{\frac{n}{k(n-k)}}\left|S_k - \frac{k}{n}S_n\right| \ge \max_{1\le k\le\left[\frac{n}{\log n}\right]} \sqrt{\frac{n}{k(n-k)}}\left|S_k - \frac{k}{n}S_n\right|\right) \longrightarrow 0,$$

as $n \to \infty$, and hence the lemma follows from Lemma 5.2. □

Lemma 5.6.

$$\max_{1\le n-k\le\left[\frac{n}{2}\right]} \sqrt{\frac{n}{k(n-k)}}\left|S_k - \frac{k}{n}S_n\right| - \max_{1\le n-k\le\left[\frac{n}{\log n}\right]} \left|\frac{S_n - S_k}{\sqrt{n-k}}\right| = o_p\left(\frac{1}{\sqrt{2\log\log n}}\right)$$

Proof. Follows from Lemma 5.5 and the stationarity under the hypothesis. □

Applying the above lemmas, we can now prove Theorem 3.

Proof of Theorem 3. We have

$$\begin{aligned} &P\left(\frac{a_n}{\sigma} \max_{1\le k\le n-1} \sqrt{\frac{n}{k(n-k)}}\left|S_k - \frac{k}{n}S_n\right| - b_n \le x\right) \\ &= P\left(\max_{1\le k\le\left[\frac{n}{2}\right]} \sqrt{\frac{n}{k(n-k)}}\left|S_k - \frac{k}{n}S_n\right| \le \frac{(x+b_n)\sigma}{a_n},\right. \\ &\qquad \left.\max_{1\le n-k\le\left[\frac{n}{2}\right]} \sqrt{\frac{n}{k(n-k)}}\left|S_k - \frac{k}{n}S_n\right| \le \frac{(x+b_n)\sigma}{a_n}\right). \end{aligned}$$

From Lemma 5.5 and Lemma 5.6 we get

$$\frac{a_n}{\sigma} \max_{1\le k\le\left[\frac{n}{2}\right]} \sqrt{\frac{n}{k(n-k)}} \left|S_k - \frac{k}{n}S_n\right| - \frac{a_n}{\sigma} \max_{1\le k\le\left[\frac{n}{\log n}\right]} \left|\frac{S_k}{\sqrt{k}}\right| \xrightarrow{\mathcal{P}} 0$$

and

$$\frac{a_n}{\sigma} \max_{1\le n-k\le\left[\frac{n}{2}\right]} \sqrt{\frac{n}{k(n-k)}} \left|S_k - \frac{k}{n}S_n\right| - \frac{a_n}{\sigma} \max_{1\le n-k\le\left[\frac{n}{\log n}\right]} \left|\frac{S_n - S_k}{\sqrt{n-k}}\right| \xrightarrow{\mathcal{P}} 0.$$

Applying Lemma 5.4 we get

$$\lim_{n\to\infty} P\left(\max_{1\le k<\left[\frac{n}{\log n}\right]} \left|\frac{S_k}{\sqrt{k}}\right| \le \frac{(x+b_n)\sigma}{a_n}\right) = \exp(-\exp(-x)).$$

And again applying Lemma 5.4 to the reversed time series, we get

$$\lim_{n\to\infty} P\left(\max_{1\le n-k\le\left[\frac{n}{\log n}\right]} \left|\frac{S_n - S_k}{\sqrt{n-k}}\right| \le \frac{(x+b_n)\sigma}{a_n}\right) = \exp(-\exp(-x)).$$

Due to the underlying $\alpha-$mixing process it holds

$$\left|P\left(\max_{1\le k\le\left[\frac{n}{\log n}\right]} \left|\frac{S_k}{\sqrt{k}}\right| \le \frac{(x+b_n)\sigma}{a_n}, \max_{1\le n-k\le\left[\frac{n}{\log n}\right]} \left|\frac{S_n - S_k}{\sqrt{n-k}}\right| \le \frac{(x+b_n)\sigma}{a_n}\right)\right.$$
$$\left. -P\left(\max_{1\le k\le\left[\frac{n}{\log n}\right]} \left|\frac{S_k}{\sqrt{k}}\right| \le \frac{(x+b_n)\sigma}{a_n}\right) P\left(\max_{1\le n-k\le\left[\frac{n}{\log n}\right]} \left|\frac{S_n - S_k}{\sqrt{n-k}}\right| \le \frac{(x+b_n)\sigma}{a_n}\right)\right| \longrightarrow 0.$$

All in all we get the desired result

$$\lim_{n\to\infty} P\left(\frac{a_n}{\sigma} \max_{1\le k\le n-1} \sqrt{\frac{n}{k(n-k)}} \left|S_k - \frac{k}{n}S_n\right| - b_n \le x\right) = \exp(-2\exp(-x)).$$

□

5.4. Proof of Theorem 4

First, we state two lemmas, both valid under the assumptions of Theorem 4. We define the random variables Y_n and Z_n by

$$\begin{aligned} Y_n &= \max_{1\le k\le(\log\log n)^\gamma} \frac{|S_k|}{\sqrt{k}}, \\ Z_n &= \max_{(\log\log n)^\gamma\le k\le n} \frac{|S_k|}{\sqrt{k}}, \end{aligned}$$

for some $\gamma > \frac{1}{2}$.

Lemma 5.7.

$$\sqrt{2\log\log n}\,Y_n - 2\log\log n \xrightarrow{\mathcal{P}} -\infty$$

Proof. We apply the following maximal inequality for α-mixing processes, due to Rio [22] Theorem 3.1

$$E\left(\max_{1\le k\le n}\left|\sum_{j=1}^{k}\xi_j\right|\right)^2 \le 16\sum_{k=1}^{n}\int_0^1 \alpha^{-1}(u)\,Q_k^2(u)\,du, \tag{23}$$

where Q_k denotes the quantile function of ξ_k, to the random variables $\xi_k = \frac{X_k}{\sqrt{k}}$. Note that

$$Q_k(u) = \frac{Q_X(u)}{\sqrt{k}}.$$

Furthermore, we use the inequality in Lemma A.1, i.e.

$$\max_{1\le k\le n}\frac{|S_k|}{\sqrt{k}} \le 2\max_{1\le k\le n}\left|\sum_{j=1}^{k}\xi_j\right|.$$

Then, for $K>0$, we obtain for all $n\ge n_K$

$$\begin{aligned}
&P\left(\sqrt{2\log\log n}\,Y_n - 2\log\log n \ge -K\right)\\
&\le P\left(Y_n \ge \sqrt{\log\log n}\right)\\
&\le P\left(2\max_{1\le k\le(\log\log n)^\gamma}\left|\sum_{j=1}^{k}\xi_j\right| \ge \sqrt{\log\log n}\right)\\
&\le \frac{4}{\log\log n}E\left(\max_{1\le k\le(\log\log n)^\gamma}\left|\sum_{j=1}^{k}\xi_j\right|\right)^2\\
&\le \frac{4}{\log\log n}16\sum_{k=1}^{(\log\log n)^\gamma}\int_0^1 \alpha^{-1}(u)\,Q_k^2(u)\,du\\
&= \frac{4}{\log\log n}16\sum_{k=1}^{(\log\log n)^\gamma}\int_0^1 \alpha^{-1}(u)\,\frac{Q_X^2(u)}{k}\,du\\
&\le \frac{4}{\log\log n}16\sum_{k=1}^{(\log\log n)^\gamma}\int_0^1 \alpha^{-1}(u)\,\frac{Q_{|X|}^2(u)}{k}\,du\\
&\le C\frac{1}{\log\log n}\sum_{k=1}^{(\log\log n)^\gamma}\frac{1}{k}
\end{aligned}$$

$$\leq C \frac{\gamma \log\log\log n}{\log\log n} \longrightarrow 0. \qquad \square$$

Lemma 5.8. *Let U_n and V_n be real-valued random variables satisfying $U_n \xrightarrow{\mathcal{P}} -\infty$. Then $V_n \xrightarrow{\mathcal{D}} G$ if and only if $\max(U_n, V_n) \xrightarrow{\mathcal{D}} G$.*

Proof. We have the following chain of inequalities

$$P(V_n > x) \leq P(\max(U_n, V_n) > x) \leq P(U_n > x) + P(V_n > x),$$

and hence $|P(\max(U_n, V_n) > x) - P(V_n > x)| \leq P(U_n > x) \to 0$. $\square$

Proof of Theorem 4. Let $(W)_{t\geq 0}$ be a Brownian motion with $\mathrm{Var}(W_1) = \sigma^2$. Define analogues of the random variables Y_n and Z_n, replacing the partial sum process by Brownian motion

$$\widetilde{Y}_n = \max_{1\leq k\leq (\log\log n)^\gamma} \frac{|W_k|}{\sqrt{k}},$$
$$\widetilde{Z}_n = \max_{(\log\log n)^\gamma \leq k\leq n} \frac{|W_k|}{\sqrt{k}}.$$

Note that $\widetilde{M}_n := \max(\widetilde{Y}_n, \widetilde{Z}_n) = \max_{1\leq k\leq n} \frac{|W_k|}{\sqrt{k}}$. By the Darling-Erdős theorem for Brownian motion, see Darling-Erdős [6], we know that

$$\frac{\sqrt{2\log\log n}}{\sigma}\, \widetilde{M}_n - b_n \xrightarrow{\mathcal{D}} G.$$

Applying Lemma 5.7 to $\widetilde{Y}_n$, and Lemma 5.8, we obtain that

$$\frac{\sqrt{2\log\log n}}{\sigma}\, \widetilde{Z}_n - b_n \xrightarrow{\mathcal{D}} G.$$

In the final step, we employ an almost sure invariance principle which is stated as Theorem 1 in Merlevède and Rio [17] under a weaker strong mixing condition than we have. We can conclude that, under the assumptions of Theorem 4, there exists a Brownian motion $(W_t)_{t\geq 0}$ with $\mathrm{Var}(W_1) = \sigma^2$, such that

$$|S_k - W_k| = O\left(k^{\frac{1}{2}-\lambda}\right) \ a.s. \text{ for } \lambda > 0.$$

Hence, we obtain

$$\begin{aligned}\left| \max_{(\log\log n)^\gamma \leq k\leq n} \frac{|S_k|}{\sqrt{k}} - \max_{(\log\log n)^\gamma \leq k\leq n} \frac{|W_k|}{\sqrt{k}} \right| &\leq \max_{(\log\log n)^\gamma \leq k\leq n} \frac{|S_k - W_k|}{\sqrt{k}} \\ &\leq C \max_{(\log\log n)^\gamma \leq k\leq n} \frac{k^{\frac{1}{2}-\lambda}}{\sqrt{k}} \\ &\leq C(\log\log n)^{-\lambda\gamma}.\end{aligned}$$

Since $\lambda\gamma > \frac{1}{2}$, this implies that $\frac{\sqrt{\log\log n}}{\sigma}(\widetilde{Z}_n - Z_n) \xrightarrow{\mathcal{P}} 0$. Hence, we obtain using Slutsky's lemma that

$$\frac{\sqrt{2\log\log n}}{\sigma} Z_n - b_n \xrightarrow{\mathcal{D}} G.$$

Now, the statement of Theorem 4 follows from Lemma 5.7 and Lemma 5.8. □

Appendix A: Auxiliary results

Lemma A.1. *For all $a_1, \ldots, a_k \in \mathbb{R}$, the following inequality holds*

$$\max_{1\le k\le n} \frac{1}{\sqrt{k}} \left| \sum_{j=1}^{k} a_j \sqrt{j} \right| \le 2 \max_{1\le k\le n} \left| \sum_{j=1}^{k} a_j \right|$$

Proof. Define the partial sums $A_j := \sum_{i=1}^{j} a_i$ and set $A_0 = 0$. Then $a_j = A_j - A_{j-1}$, and thus

$$\begin{aligned}
\left| \sum_{j=1}^{k} a_j \sqrt{j} \right| &= \left| \sum_{j=1}^{k} (A_j - A_{j-1}) \sqrt{j} \right| \\
&= \left| \sum_{j=1}^{k-1} A_j \left(\sqrt{j} - \sqrt{j+1} \right) + A_k \sqrt{k} \right| \\
&\le \max_{1\le j\le k} |A_j| \sum_{j=1}^{k-1} \left(\sqrt{j+1} - \sqrt{j} \right) + |A_k| \sqrt{k} \\
&= \max_{1\le j\le k} |A_j| \sqrt{k} + |A_k| \sqrt{k} \\
&\le 2\sqrt{k} \max_{1\le j\le k} |A_j|.
\end{aligned}$$

□

Remark 3. *This is a special case of an inequality stated as Lemma 1 in Shorack and Smythe [26].*

Lemma A.2. *Assume that the kernel g is degenerated and that it satisfies the variation condition. Let $m = \max\{i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)}\}$, where $\{i_1, i_2, i_3, i_4\} = \{i_{(1)}, i_{(2)}, i_{(3)}, i_{(4)}\}$ and $i_{(1)} \le i_{(2)} \le i_{(3)} \le i_{(4)}$. If g is a kernel with uniform $(2+\delta)$-moments for a $\delta > 0$, then there exists a constant C such that*

$$|E\left(g(X_{i_1}, X_{i_2}) g(X_{i_3}, X_{i_4})\right)| \le C \left(\int_0^{\alpha(m)} Q_{|X|}(u) du \right)^{\frac{\delta}{3+2\delta}}.$$

Proof. Let $\varepsilon > 0,\ K > 0$ and define

$$g_K(x,y) = \begin{cases} g(x,y) & \text{if } |g(x,y)| \le \sqrt{K} \\ \sqrt{K} & \text{if } g(x,y) > \sqrt{K} \\ -\sqrt{K} & \text{if } g(x,y) < -\sqrt{K}. \end{cases}$$

Let g satisfy the variation condition with constant L. Then, g_K satisfies also the variation condition with the same constant L. Assume, without loss of generality, that $m = i_2 - i_1$. Moreover, we can assume that there exists a uniform on $[0,1]$ random variable that is independent of $(X_i)_{i\ge 1}$. With Theorem 1 of Peligrad [18], choose a random variable X'_{i_1} independent of $X_{i_2}, X_{i_3}, X_{i_4}$ with the same distribution as X_{i_1} and

$$P\left(|X_{i_1} - X'_{i_1}| \ge \varepsilon\right) \le \frac{4\int_0^{\alpha(m)} Q_{|X|}(u)du}{\varepsilon}.$$

As g is a degenerate kernel, we have

$$E\left(g(X'_{i_1}, X_{i_2})g(X_{i_3}, X_{i_4})\right) = 0.$$

We get

$$\begin{aligned}
&|E\left(g(X_{i_1}, X_{i_2})g(X_{i_3}, X_{i_4})\right)| \\
=&\left|E\left(g(X_{i_1}, X_{i_2})g(X_{i_3}, X_{i_4})\right) - E\left(g(X'_{i_1}, X_{i_2})g(X_{i_3}, X_{i_4})\right)\right| \\
=&E\left(\left|(g_K(X_{i_1}, X_{i_2}) - g_K(X'_{i_1}, X_{i_2}))g_K(X_{i_3,X_{i_4}})\right|1_{\{|X_{i_1}-X'_{i_1}|<\varepsilon\}}\right) && (24)\\
&+E\left(\left|(g_K(X_{i_1}, X_{i_2}) - g_K(X'_{i_1}, X_{i_2}))g_K(X_{i_3,X_{i_4}})\right|1_{\{|X_{i_1}-X'_{i_1}|\ge\varepsilon\}}\right) && (25)\\
&+E\left(\left|g_K(X_{i_1}, X_{i_2})g_K(X_{i_3}, X_{i_4}) - g(X_{i_1}, X_{i_2})g(X_{i_3}, X_{i_4})\right|\right) && (26)\\
&+E\left(\left|g_K(X'_{i_1}, X_{i_2})g_K(X_{i_3}, X_{i_4}) - g(X'_{i_1}, X_{i_2})g(X_{i_3}, X_{i_4})\right|\right). && (27)
\end{aligned}$$

Due to the variation condition and $|g_K(X_{i_3}, X_{i_4})| \le \sqrt{K}$, which holds by definition, we have that (24) is smaller than $L\varepsilon\sqrt{K}$. The second summand (25) is bounded by

$$2KP(|X_{i_1} - X'_{i_1}| \ge \varepsilon) \le \frac{8\int_0^{\alpha(m)} Q_{|X|}(u)du}{\varepsilon}K.$$

Let M be the bound of the $(2+\delta)$-moments of g. For the third term of the sum (26) we get the following chain of inequalities.

$$\begin{aligned}
&E\left(\left|g_K(X_{i_1}, X_{i_2})g_K(X_{i_3}, X_{i_4}) - g(X_{i_1}, X_{i_2})g(X_{i_3}, X_{i_4})\right|\right)\\
\le&E\left(\left|g(X_{i_1}, X_{i_2})\sqrt{K} - g(X_{i_1}, X_{i_2})g(X_{i_3}, X_{i_4})\right|1_{\{|g(X_{i_1},X_{i_2})|\le\sqrt{K}, g(X_{i_3},X_{i_4})>\sqrt{K}\}}\right)\\
&+E\left(\left|g(X_{i_1},X_{i_2})(-\sqrt{K}) - g(X_{i_1},X_{i_2})g(X_{i_3},X_{i_4})\right|1_{\{|g(X_{i_1},X_{i_2})|\le\sqrt{K}, g(X_{i_3},X_{i_4})<-\sqrt{K}\}}\right)\\
&+E\left(\left|\sqrt{K}g(X_{i_3}, X_{i_4}) - g(X_{i_1}, X_{i_2})g(X_{i_3}, X_{i_4})\right|1_{\{g(X_{i_1},X_{i_2})>\sqrt{K}, |g(X_{i_3},X_{i_4})|\le\sqrt{K}\}}\right)
\end{aligned}$$

$$
\begin{aligned}
&+E\big(\big|-\sqrt{K}g(X_{i_3},X_{i_4})-g(X_{i_1},X_{i_2})g(X_{i_3},X_{i_4})\big|1_{\{g(X_{i_1},X_{i_2})<-\sqrt{K},|g(X_{i_3},X_{i_4})|\le\sqrt{K}\}}\big)\\
&+E\big(\big|\sqrt{K}\sqrt{K}-g(X_{i_1},X_{i_2})g(X_{i_3},X_{i_4})\big|1_{\{g(X_{i_1},X_{i_2})>\sqrt{K},g(X_{i_3},X_{i_4})>\sqrt{K}\}}\big)\\
&+E\big(\big|(-\sqrt{K})(-\sqrt{K})-g(X_{i_1},X_{i_2})g(X_{i_3},X_{i_4})\big|1_{\{g(X_{i_1},X_{i_2})<-\sqrt{K},g(X_{i_3},X_{i_4})<-\sqrt{K}\}}\big)\\
&+E\big(\big|\sqrt{K}(-\sqrt{K})-g(X_{i_1},X_{i_2})g(X_{i_3},X_{i_4})\big|1_{\{g(X_{i_1},X_{i_2})>\sqrt{K},g(X_{i_3},X_{i_4})<-\sqrt{K}\}}\big)\\
&+E\big(\big|(-\sqrt{K})\sqrt{K}-g(X_{i_1},X_{i_2})g(X_{i_3},X_{i_4})\big|1_{\{g(X_{i_1},X_{i_2})<-\sqrt{K},g(X_{i_3},X_{i_4})>\sqrt{K}\}}\big).
\end{aligned}
$$

Considering the first term of the sum, we have

$$
\begin{aligned}
&E\big(\big|g(X_{i_1},X_{i_2})\sqrt{K}-g(X_{i_1},X_{i_2})g(X_{i_3},X_{i_4})\big|1_{\{|g(X_{i_1},X_{i_2})|\le\sqrt{K},g(X_{i_3},X_{i_4})>\sqrt{K}\}}\big)\\
&=E\big(\big|(\sqrt{(}K)-g(X_{i_3},X_{i_4}))g(X_{i_1},X_{i_2}\big|1_{\{g(X_{i_3},X_{i_4})>\sqrt{K}\}}\big)\\
&\le\sqrt{K}E\big(\big|\sqrt{(}K)-g(X_{i_3},X_{i_4})\big|1_{\{g(X_{i_3},X_{i_4})>\sqrt{K}\}}\big)\\
&\le\sqrt{K}E\big((|g(X_{i_3},X_{i_4})|-\sqrt{K})1_{\{|g(X_{i_3},X_{i_4})|>\sqrt{K}\}}\big)\\
&\le\sqrt{K}\frac{E|g(X_{i_3},X_{i_4})|^{2+\delta}}{(\sqrt{K})^{1+\delta}}=\frac{E|g(X_{i_3},X_{i_4})|^{2+\delta}}{K^{\delta/2}}\le\frac{M}{K^{\delta/2}}.
\end{aligned}
$$

The next three terms can be treated analogously. For the fifth summand we get the following inequality

$$
\begin{aligned}
&E\big(\big|\sqrt{K}\sqrt{K}-g(X_{i_1},X_{i_2})g(X_{i_3},X_{i_4})\big|1_{\{g(X_{i_1},X_{i_2})>\sqrt{K},g(X_{i_3},X_{i_4})>\sqrt{K}\}}\big)\\
&\le E\big(\big(|g(X_{i_1},X_{i_2})||g(X_{i_3},X_{i_4})|-K\big)1_{\{|g(X_{i_1},X_{i_2})|>\sqrt{K},|g(X_{i_3},X_{i_4})|>\sqrt{K}\}}\big)\\
&\le\big(E\big(|g(X_{i_1},X_{i_2})|^2 1_{\{|g(X_{i_1},X_{i_2})|>\sqrt{K}\}}\big)\big)^{1/2}\big(E\big(|g(X_{i_3},X_{i_4})|^2 1_{\{|g(X_{i_3},X_{i_4})|>\sqrt{K}\}}\big)\big)^{1/2}\\
&\le\left(\frac{E|g(X_{i_1},X_{i_2})|^{2+\delta}}{(\sqrt{K})^{\delta}}\right)^{1/2}\left(\frac{E|g(X_{i_3},X_{i_4})|^{2+\delta}}{(\sqrt{K})^{\delta}}\right)^{1/2}\le\frac{M}{K^{\delta/2}}.
\end{aligned}
$$

Analogously, we get for the last three terms that they are also bounded by $\frac{M}{K^{\delta/2}}$. Altogether, we have

$$
E\big(\big|g_K(X_{i_1},X_{i_2})g_K(X_{i_3},X_{i_4})-g(X_{i_1},X_{i_2})g(X_{i_3},X_{i_4})\big|\big)\le 8\frac{M}{K^{\delta/2}}.
$$

Similarly, we obtain for (27)

$$
E\big(\big|g_K(X'_{i_1},X_{i_2})g_K(X_{i_3},X_{i_4})-g(X'_{i_1},X_{i_2})g(X_{i_3},X_{i_4})\big|\big)\le 8\frac{M}{K^{\delta/2}}.
$$

Altogether, we get

$$
\big|E\big(g(X_{i_1},X_{i_2})g(X_{i_3},X_{i_4})\big)\big|\le L\varepsilon\sqrt{K}+\frac{8\int_0^{\alpha(m)}Q_{|X|}(u)du}{\varepsilon}K+16\frac{M}{K^{\delta/2}}.
$$

Choosing $\varepsilon=\sqrt{\int_0^{\alpha(m)}Q_{|X|}(u)du}K^{1/4}$ and $K=(\int_0^{\alpha(m)}Q_{|X|}(u)du)^{-\frac{1}{3/2+\delta}}$, we obtain

$$
\big|E\big(g(X_{i_1},X_{i_2})g(X_{i_3},X_{i_4})\big)\big|\le(L+8+16M)\left(\int_0^{\alpha(m)}Q_{|X|}(u)du\right)^{\frac{\delta}{3+2\delta}}. \qquad \square
$$

Lemma A.3. *In addition to the conditions of Lemma A.2, let (11) be satisfied. Then there exists a constant C such that*

$$E\left(\sum_{i=1}^{k}\sum_{j=k+1}^{n} g(X_i, X_j)\right)^2 \leq Ck(n-k).$$

Proof. This was proved for functionals of absolutely regular processes by Dehling et al. [7] in Lemma 1. They make use of an upper bound for the expectations $|E(g(X_{i_1}, X_{i_2})g(X_{i_3}, X_{i_4}))|$. Such a bound for an α-mixing process is stated in Lemma A.2. The rest of the proof is analogous. □

References

[1] Antoch, J., Hušková, M. and Prášková, Z. (1997). Effect of dependence on statistics for determination of change. *J. Statist. Plann. Inference* **60** 291–310. MR1456633

[2] Csörgő, M. and Horváth, L. (1988). Invariance Principles for Change-point Problems. *J. Multivariate Anal.* **27** 151–168. MR971179

[3] Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis. Wiley Series in Probability and Statistics.* John Wiley & Sons, Ltd., Chichester. MR2743035

[4] Csörgő, M., Szyszkowicz, B. and Wang, Q. (2008). Asymptotics of Studentized U-type processes for changepoint problems. *Acta Math. Hungar.* **121** 333–357. MR2461439

[5] Darkhovskh, B. S. (1976). A Nonparametric Method for the a Posteriori Detection of the "Disorder" Time of a Sequence of Independent Random Variables. *Theory Probab. Appl.* **21** 178–183.

[6] Darling, D. and Erdös, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Mathematical Journal* **23** 143–155. MR0074712

[7] Dehling, H., Fried, R., Garcia, I. and Wendler, M. (2015). Change-Point Detection Under Dependence Based on Two-Sample U-Statistics. In *Asymptotic Laws and Methods in Stochastics*, (D. Dawson, R. Kulik, M. Ould Haye, B. Szyszkowicz and Y. Zhao, eds.) **76** 195–220. Springer, New York, NY. MR3409833

[8] Dehling, H., Fried, R., Sharipov, O., Vogel, D. and Wornowizki, M. (2013). Estimation of the variance of partial sums of dependent processes. *Statist. Probab. Lett.* **83** 141–147. MR2998735

[9] Dehling, H., Fried, R. and Wendler, M. (2020). A robust method for shift detection in time series. *Biometrika* **107** 647–660. MR4138981

[10] Dehling, H., Rooch, A. and Taqqu, M. S. (2013). Non-parametric change-point tests for long-range dependent data. *Scandinavian Journal of Statistics* **40** 153–173. MR3024037

[11] Dehling, H., Rooch, A. and Wendler, M. (2017). Two-Sample U-Statistic Processes for Long-Range Dependent Data. *Statistics* **51** 84–104. MR3600463

[12] Denker, M. and Keller, G. (1986). Rigorous Statistical Procedures for Data from Dynamical Systems. *J. Stat. Phys.* **44** 67–93. MR0854400

[13] Ferger, D. (1994). On the power of nonparametric changepoint-tests. *Metrika* **41** 277–292. MR1292400

[14] Gombay, E. (2000). U-statistics for sequential change detection. *Metrika* **52** 133–145. MR1811267

[15] Hoeffding, W. (1948). A Class of Statistics with Asymptotically Normal Distribution. *Ann. Math. Statist.* **19** 293–325. MR0026294

[16] Kirch, C. and Stoehr, C. (2020). Asymptotic delay times of sequential tests based on U-statistics for early and late change points. *arXiv preprint arXiv:2003.08847*.

[17] Merlevède, F. and Rio, E. (2012). Strong approximation of partial sums under dependence conditions with application to dynamical systems. *Stochastic Process. Appl.* **122** 386–417. MR2860454

[18] Peligrad, M. (2002). Some remarks on coupling of dependent random variables. *Statist. Probab. Lett.* **60** 201–209. MR1945442

[19] Pettitt, A. N. (1979). A Non-Parametric Approach to the Change-Point Problem. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **28** 126–135. MR0539082

[20] Račkauskas, A. and Wendler, M. (2020). Convergence of U-processes in Hölder spaces with application to robust detection of a changed segment. *Statistical Papers* **61** 1409–1435. MR4127480

[21] Rio, E. (1993). Covariance inequalities for strongly mixing processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **29** 587–597. MR1251142

[22] Rio, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes* **80**. Springer, Berlin. MR3642873

[23] Rosenblatt, M. (1956). A Central Limit Theorem and a Strong Mixing Condition. *Proc. Natl. Acad. Sci. USA* **42** 43–47. MR0074711

[24] Serfling, R. J. (1970). Moment Inequalities for the Maximum Cumulative Sum. *Ann. Math. Statist.* **41** 1227–1234. MR0268938

[25] Shorack, G. R. (1979). Extension of the Darling and Erdös Theorem on the Maximum of Normalized Sums. *Ann. Probab.* **7** 1092–1096. MR0548907

[26] Shorack, G. R. and Smythe, R. T. (1976). Inequalities for $\max|S_k|/b_k$ where $k \in N^r$. *Proc. Amer. Math. Soc.* **54** 331–336. MR0400386

[27] Wang, R., Zhu, C. and Volgushev, S. and Shao, X. (2022). Inference for change points in high dimensional data via self-normalization. *Ann. Math. Statist.*

[28] Yao, Y.-C. and Davis, R. (1986). The Asymptotic Behavior of the Likelihood Ratio Statistic for Testing a Shift in Mean in a Sequence of Independent Normal Variates. *Sankhyā Ser. A* **48** 339–353. MR0905446

[29] Zhang, Y., Wang, R. and Shao, X. (2021). Adaptive Inference for Change Points in High-Dimensional Data. *J. Amer. Statist. Assoc.* 1–12.