Electronic Journal of Statistics
Vol. 18 (2024) 2208–2240
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2248

Power of weighted test statistics for structural change in time series

Herold Dehling[1], **Kata Vuk**[2] **and Marin Wendler**[3]

[1] *Faculty of Mathematics, Ruhr University Bochum, Germany,*
e-mail: herold.dehling@rub.de

[2] *Faculty of Informatics and Data Science, University of Regensburg, Germany,*
e-mail: kata.vuk@ur.de

[3] *Institute for Mathematical Stochastics, Otto-von-Guericke University Magdeburg, Germany,*
e-mail: martin.wendler@ovgu.de

Abstract: We investigate the power of some common change-point tests as a function of the location of the change-point. The test statistics are maxima of weighted U-statistics, with the CUSUM test and the Wilcoxon change-point test as special examples. We study the power under local alternatives, where we vary both the change-point's location and the magnitude of the change. We quantify in which way weighted versions of the tests exhibit greater power when the change occurs near the beginning or the end of the time interval, while losing power against changes located in the center.

Keywords and phrases: Weighted change-point tests, power of tests, U-statistics, local alternatives.

Received March 2023.

1. Introduction

In this paper, we compare the power of some standard change-point tests with varying weight functions. We consider alternatives where a jump occurs in the center of the observation period, as well as alternatives where a jump occurs very early or very late. According to the change-point folklore, very early and very late changes are better detected by tests with weights that increase near the boundary of the observation period. In this paper, we aim to shed some light on this problem, both by precise mathematical results and by simulations. We will do so by considering local alternatives that express the phenomenon of change-points near the border of the observation period. Our results indicate that optimal weights depend on the rate at which the change-point converges to the border of the observation time.

We investigate the model of at most one change, assuming that the data are generated by the signal plus noise model

$$X_i = \mu_i + \xi_i,\ i \geq 1, \tag{1}$$

where $(\mu_i)_{i\geq 1}$ is an unknown signal, and where $(\xi_i)_{i\geq 1}$ is a mean zero i.i.d.

process. Based on the observations $X_1, \dots, X_n$, we want to test the hypothesis

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

that there is no change in the location during the observation period $\{1, \dots, n\}$ against the alternative that there is a change at some unknown point in time k^*, i.e.,

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

We will specifically consider alternatives where the location k^* as well as the height $\Delta = \mu_{k^*+1} - \mu_{k^*}$ of the change is allowed to vary with the sample size. Thus, strictly speaking, our model is a triangular array where the signal is given by $(\mu_{n,i})_{1\le i\le n, n\ge 1}$. We study U-process based test statistics defined as

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

where $0 \le \gamma < \frac{1}{2}$ is a tuning parameter and where $h : \mathbb{R}^2 \to \mathbb{R}$ is a kernel function. We consider kernels of the type

$$h(x, y) = g(y - x),$$

where g is an odd function. That type of kernel function covers many test statistics including CUSUM and Wilcoxon. The parameter γ defines the strength of the weight near the borders of the observation period. The greater γ is, the higher are the weights at the border. The limit $\gamma = \frac{1}{2}$ is exceptional because in this case the test statistics asymptotic distribution is an extreme value distribution. The choice $\gamma = 0$ yields the non-weighted version of the test.

The asymptotic distribution of the U-process based test statistic with kernel $h(x, y)$ has been studied by various authors, both for i.i.d. data and for data with serial correlations. For i.i.d. data, the theory is summarized in the seminal monograph by Csörgő and Horváth [4], where also dependent data are treated in connection with the CUSUM test. For short-range dependent data and general U-statistic based tests, Dehling et al. investigated the case $\gamma = 0$ in [6]. The special weighted case $\gamma = \frac{1}{2}$ was treated in our former paper [10].

Most research on change-point tests has been devoted to the distribution of test statistics under the null hypothesis of no change. In the case of i.i.d. data, the non-weighted U-statistics process for detecting a change in the distribution under contiguous alternatives, where the location of the change-point is in the center of the observation period, was studied by Szyszkowicz [18]. Also for i.i.d. data, the power of non-weighted tests based on anti-symmetric U-statistics under local alternatives, allowing both the location and the height of the jump to vary with the sample size, including possible jumps near the boundary of the observation period, was studied by Ferger [11]. For long-range dependent processes the asymptotic distribution of the Wilcoxon test statistic under the hypothesis as well as under the alternative was studied Dehling et al. in [8] and

[9]. In Horváth's, Rice's, and Zhao's paper from 2021 [15], the norms of weighted functional CUSUM processes were studied, and the asymptotic distribution under the null hypothesis of no change, as well as under local alternatives in the presence of a change in the covariance, was derived. Additionally, Robbins et al. [17] consider weighted versions of CUSUM tests for detecting early changes and compares various tests through simulation studies. Horváth et al. [14] investigated the power of non-weighted CUSUM tests under local alternatives for short-range dependent data. In that paper, they also investigated the power of so-called Rényi change-point tests, which are CUSUM tests weighted even more heavily towards the end of the observation period.

In the context of weighted and non-weighted U-statistics based processes, Račkauskas and Wendler [16] focus on epidemic changes, while Berkes et al. [1] address changes in the mean of the covariance structure of a linear process. Additionally, Gombay [12] compares the power of U-statistic based change-point tests for both online and offline scenarios, describing their large sample behavior under local alternatives. However, these analyses do not encompass contiguous alternatives with $O(n^{-1/2})$ size changes. In a Monte Carlo simulation study, Xie et al. [19] investigate how the location of the change-point influences the ability of the Wilcoxon-based Pettitt test.

In this paper, we investigate changes where the change-point occurs on the scale of n^κ, considering various values of κ within the range $(0, 1]$. For $\kappa = 1$, changes occur in the center of the observation period, while $0 < \kappa < 1$ corresponds to very early changes. We study both fixed size jumps and jumps whose size decreases as the sample size increases. Additionally, we compare the power functions of different tests and study their relationship with the envelope power. The envelope power is defined as the maximal power achievable by testing the hypothesis of no change against a fixed alternative (k, Δ), where k denotes the location and Δ represents the height of the change.

To the best of our knowledge, there are no results in the literature where the power of weighted change-point test statistics under local alternatives has been studied analytically. New are specifically our results on the power of weighted tests against local alternatives where the change-point occurs near the very beginning of the time series, on a scale of n^κ, where $0 < \kappa < 1$. We identify exactly the combinations of weights γ and scales κ that yield consistent tests, and we analyze the limit distribution on the border.

2. Preliminary remarks and definitions

Before we present our main results in the next sections, we introduce some notations. We consider the signal plus noise model (1) and test the hypothesis

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

against the alternative

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

The next sections attend to two different types of alternative:

The section *Small change after fixed proportion of time* deals with local alternatives in which the time of change is proportional to the sample size and the jump height decreases as the sample size increases. We call this alternative A_1 and define more precisely

$$A_1: \ \mu_1 = \ldots = \mu_{k_n^*} \neq \mu_{k_n^*+1} = \ldots = \mu_n, \text{ with}$$
$$k_n^* = [\tau^* n] \text{ and } \Delta_n = \mu_{k_n^*+1} - \mu_{k_n^*} = \frac{c}{\sqrt{n}},$$

where $\tau^* \in (0,1)$ and c is a constant.

In the section *Early change with fixed height*, we consider another type of alternative in which the jump height is kept constant, while the time of change moves closer to the border of the observation range. We model this alternative as follows

$$A_2: \ \mu_1 = \ldots = \mu_{k_n^*} \neq \mu_{k_n^*+1} = \ldots = \mu_n, \text{ with}$$
$$k_n^* \approx cn^\kappa, \text{ meaning that } \frac{k_n^*}{cn^\kappa} \to 1, \text{ and } \Delta_n = \mu_{k_n^*+1} - \mu_{k_n^*} \equiv \Delta,$$

where c is a constant and where the parameter κ is defined as

$$\kappa = \frac{1-2\gamma}{2(1-\gamma)}, \ \gamma \in [0, \tfrac{1}{2}).$$

Note that by definition $\kappa \in (0, \frac{1}{2}]$.

In short, we can write the corresponding model as

$$X_i = \begin{cases} \mu + \xi_i & \text{for } i \leq k_n^* \\ \mu + \Delta_n + \xi_i & \text{for } i \geq k_n^* + 1, \end{cases} \tag{2}$$

where $(\xi_i)_{\geq 1}$ is a mean zero i.i.d. process and where k_n^* and Δ_n are chosen as in A_1 or A_2.

Remark 1. The specific choice of κ in A_2 leads to a non trivial limit distribution under the alternative.

In order to test H vs. $A_i, \ i \in \{1,2\}$, we use the test statistic

$$G_n^\gamma(k) := \frac{1}{n^{3/2}\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma} \sum_{i=1}^{k} \sum_{j=k+1}^{n} g(X_j - X_i), \tag{3}$$

where $\gamma \in [0, 1/2)$ and where g is an odd function, i.e., $g(-x) = -g(x)$. Note that the case $\gamma = 0$ refers to the non-weighted test statistic. We determine the limiting distribution of the test statistic under the alternatives A_1 and A_2. In that proceeding, slightly different terms appear depending on whether $k \leq k_n^*$ or $k \geq k_n^*$. For the sake of simplicity we combine these terms into one function. We define $\phi_n : \{1, \ldots, n\} \to \mathbb{N}$ by

$$\phi_n(k) := \begin{cases} k(n - k_n^*) & \text{for } k \leq k_n^* \\ k_n^*(n-k) & \text{for } k \geq k_n^*, \end{cases}$$

and analogously the continuous version $\phi_{\tau^*} : [0,1] \to \mathbb{R}$ by

$$\phi_{\tau^*}(\lambda) = \begin{cases} \lambda(1-\tau^*) & \text{for } \lambda \le \tau^* \\ \tau^*(1-\lambda) & \text{for } \lambda \ge \tau^*. \end{cases}$$

For later use, we denote a Wiener process by $\{W(\lambda), 0 \le \lambda \le 1\}$ and a Brownian bridge process by $\{W^{(0)}(\lambda), 0 \le \lambda \le 1\}$.

3. Small changes after fixed proportion of a sample

In this section, we establish the asymptotic distribution of $\max_{1 \le k < n} G_n^\gamma(k)$ under the alternative A_1, i.e., where $k_n^* = [\tau^* n]$ and $\Delta_n = \frac{c}{\sqrt{n}}$. For the special CUSUM and Wilcoxon kernel functions, the results are stated in Corollary 3.1, and 3.2.

Theorem 3.1. *We consider model* (2) *under* A_1. *Assume that* $g(\xi_2 - \xi_1)$ *has finite second moments. Moreover, assume that* $\mathrm{Var}(h_1(\xi_1)) \to 0$, *and that* $c_g = \lim_{n\to\infty} \sqrt{n} u(\Delta_n)$ *exists. Then, for* $0 \le \gamma < \frac{1}{2}$, *and as* $n \to \infty$,

$$\max_{1 \le k < n} G_n^\gamma(k) \xrightarrow{\mathcal{D}} \sup_{0 \le \lambda \le 1} \frac{1}{(\lambda(1-\lambda))^\gamma} [\sigma W^{(0)}(\lambda) + c_g \phi_{\tau^*}(\lambda)],$$

where $\sigma^2 = \mathbb{E}(g_1^2(\xi_1)) > 0$, *and*

$$\begin{aligned} g_1(x) &= \mathbb{E}[g(\xi - x)] - \mathbb{E}[g(\xi - \eta)], \\ u(\Delta_n) &= \mathbb{E}[g(\xi - \eta + \Delta_n) - g(\xi - \eta)], \\ h_1(x) &= \mathbb{E}[g(\xi - x + \Delta_n) - g(\xi - x)] - u(\Delta_n), \end{aligned}$$

where ξ *and* η *are independent and have the same distribution as* ξ_1.

Remark 2. (i) h_1 and u are obtained from Hoeffding's decomposition, applied to the kernel $h(x,y) = g(y - x + \Delta_n) - g(y - x)$. More details are given in the proof of Theorem 3.1 in Section 7.
(ii) For $c = 0$ we obtain the limit under the null hypothesis of stationarity. In order to calculate the asymptotic critical values, we need to determine the quantiles of the distribution of

$$\sup_{0 \le \lambda \le 1} \frac{1}{(\lambda(1-\lambda))^\gamma} W^{(0)}(\lambda).$$

The upper α-quantiles, $\alpha \in (0.01, 0.05, 0.1)$, for various choices of γ, are tabulated in Table 1.
(iii) We conjecture that the results of Theorem 1 and Theorems 2 and 3 below would also hold for short-range dependent observations under suitable conditions, similar to those in [6] or [10]. Under dependence, the variance parameter $\sigma^2 = \mathrm{Var}(g_1(\xi_1))$ would have to be replaced by the long-run variance $\sigma_\infty^2 = \mathrm{Var}(g_1(\xi_1)) + 2\sum_{k=1}^\infty \mathrm{Cov}(g_1(\xi_1), g_1(\xi_{1+k}))$.

(iv) For $\gamma = 1/2$, Theorem 3.1 does not hold, as the limit is infinite almost surely. Under the null hypothesis, and with $\gamma = 1/2$, one can apply a different normalization to obtain the convergence to an extreme value distribution, see Csörgő and Horváth's paper from 1988 [3].

TABLE 1
Upper α-Quantiles of $\sup_{0\le\lambda\le 1} \frac{1}{(\lambda(1-\lambda))^\gamma} W^{(0)}(\lambda)$ for different values of the parameter γ, based on 10,000 repetitions.

γ \ α	0.1	0.05	0.01
0	1.05	1.20	1.51
0.1	1.24	1.41	1.72
0.2	1.45	1.63	2.05
0.3	1.75	1.96	2.40
0.4	2.10	2.31	2.83

Theorem 3.1 covers both the CUSUM and Wilcoxon test statistic. Choosing $g(x) = x$ leads to the CUSUM test statistic and satisfies the assumptions. We have $u(\Delta_n) = E[\Delta_n] = \Delta_n$ and $c_g = \lim_{n\to\infty} \sqrt{n}\Delta_n = c$, as $\Delta_n = \frac{c}{\sqrt{n}}$. Moreover,

$$h_1(x) = \mathbb{E}[g(\xi - x + \Delta_n) - g(\xi - x)] - u(\Delta_n) = \mathbb{E}[\xi - x + \Delta_n - (\xi - x)] - \Delta_n = 0.$$

Thus, we can deduce the following corollary for the weighted CUSUM test statistic.

Corollary 3.1. *Under the assumptions of Theorem 3.1, it holds*

$$\max_{1\le k<n} \frac{1}{n^{3/2}\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma} \sum_{i=1}^{k} \sum_{j=k+1}^{n} (X_j - X_i) \xrightarrow{\mathcal{D}} \sup_{0\le\lambda\le 1} \frac{1}{(\lambda(1-\lambda))^\gamma} \left[\sigma W^{(0)}(\lambda) + c\phi_{\tau^*}(\lambda)\right],$$

where $\sigma^2 = \mathrm{Var}(\xi_1) < \infty$.

To obtain the Wilcoxon test statistic, choose $g(x) = 1_{\{0\le x\}} - \frac{1}{2}$. Then

$$\begin{aligned} u(\Delta_n) &= \mathbb{E}[1_{\{\eta-\Delta_n\le\xi\}} - 1_{\{\eta\le\xi\}}] = \mathbb{E}[1_{\{\eta-\Delta_n<\xi\le\eta\}}] = \mathbb{P}(\eta - \Delta_n < \xi \le \eta) \\ &= \int_{\mathbb{R}} (F(y) - F(y-\Delta_n))dF(y) \approx -\Delta_n \int_{\mathbb{R}} f^2(y)dy, \end{aligned}$$

where F is the distribution function and f the density function of ξ. This yields $c_g = c\int_{\mathbb{R}} f^2(y)dy$. Furthermore,

$$\begin{aligned} |h_1(x)| &= |\mathbb{E}[1_{\{0\le\xi-x+\Delta_n\}} - 1_{\{0\le\xi-x\}}] - u(\Delta_n)| \\ &= |\mathbb{P}(x - \Delta_n < \xi \le x) - u(\Delta_n)| = |F(x) - F(x-\Delta_n) - u(\Delta_n)| \end{aligned}$$

$$= \left|\Delta_n \frac{F(x) - F(x-\Delta_n)}{\Delta_n} - u(\Delta_n)\right| \approx |\Delta_n f(x) - u(\Delta_n)|$$
$$= \left|\Delta_n\Big(\int_{\mathbb{R}} f^2(y)dy - f(x)\Big)\right| \leq |c\Delta_n|,$$

where c is a finite constant if the density is bounded. Thus, $\mathrm{Var}(h_1(\xi_1)) \to 0$. As all required assumptions are satisfied, we derive the following corollary.

Corollary 3.2. *Assume that ξ_1 has bounded density. Under the assumptions of Theorem 3.1 it holds*

$$\max_{1\leq k<n} \frac{1}{n^{3/2}\left(\frac{k}{n}(1-\frac{k}{n})\right)^{\gamma}} \sum_{i=1}^{k}\sum_{j=k+1}^{n}\Big(1_{\{X_i\leq X_j\}} - \frac{1}{2}\Big)$$
$$\xrightarrow{\mathcal{D}} \sup_{0\leq\lambda\leq 1} \frac{1}{(\lambda(1-\lambda))^{\gamma}}\left[\frac{1}{\sqrt{12}}W^{(0)}(\lambda) + c\phi_{\tau^*}(\lambda)\int_{\mathbb{R}} f^2(y)dy\right].$$

4. Early change with fixed height

Next, we consider Alternative A_2, where the jump height remains constant while the time of change moves closer to the border of the observation range. We show that the choice of γ influences the scales at which change-points can be detected.

We distinguish between the cases $\gamma = 0$ and $\gamma \in (0, 1/2)$. First, we consider the case $\gamma = 0$, i.e., the case where the norming sequence is $\frac{1}{n^{3/2}}$ and does not depend on k. For $\gamma = 0$, we get $\kappa = 1/2$, which yields the alternative where the change-point occurs at time $k_n^* \approx c\sqrt{n}$.

Theorem 4.1. *We consider model (2) under A_2. Assume that $g(\xi_2 - \xi_1)$ has finite second moments. Moreover, assume that $\mathrm{Var}(h_1(\xi_1)) < \infty$. Then, for $\gamma = 0$ and as $n \to \infty$,*

$$\max_{1\leq k<n} |G_n^0(k)| \xrightarrow{\mathcal{D}} \sup_{0\leq\lambda\leq 1}\left|\sigma W^{(0)}(\lambda) + c(1-\lambda)u(\Delta)\right|,$$

where σ, $u(\Delta_n)$, and $h_1(\xi_1)$ are defined as in Theorem 3.1, albeit with $\Delta_n \equiv \Delta$.

For the CUSUM kernel, we have $u(\Delta) = \Delta$ and $\mathrm{Var}(h_1(\xi_1)) = 0$. For the Wilcoxon kernel, we get $u(\Delta) = \mathbb{P}(0 \leq \xi_2 - \xi_1 \leq \Delta)$ and $\mathrm{Var}(h_1(\xi_1)) \to 0$. Thus, we can deduce the following corollaries.

Corollary 4.1. *Under the assumptions of Theorem 4.1, it holds*

$$\max_{1\leq k<n} \frac{1}{n^{3/2}}\left|\sum_{i=1}^{k}\sum_{j=k+1}^{n}(X_j - X_i)\right| \xrightarrow{\mathcal{D}} \sup_{0\leq\lambda\leq 1} |\sigma W^{(0)}(\lambda) + c(1-\lambda)\Delta|,$$

where $\sigma^2 = \mathrm{Var}(\xi_1) < \infty$.

Corollary 4.2. *Under the assumptions of Theorem 4.1, it holds*

$$\max_{1\le k<n} \frac{1}{n^{3/2}}\Big|\sum_{i=1}^{k}\sum_{j=k+1}^{n}\Big(1_{\{X_i\le X_j\}}-\frac{1}{2}\Big)\Big| \xrightarrow{\mathcal{D}} \sup_{0\le\lambda\le 1}\Big|\frac{1}{12}W^{(0)}(\lambda)+c(1-\lambda)\mathbb{P}(0\le \xi_2-\xi_1\le\Delta)\Big|.$$

Remark 3. (i) When $c=0$, i.e., when $k_n^*/\sqrt{n}\to 0$, the distribution of the test statistic under the alternative is asymptotically the same as under the null hypothesis, and thus the test has no power to detect such alternatives. The test has asymptotically only trivial power α, the same as the size.

(ii) The test is consistent if and only if $\lim_{n\to\infty}\frac{k_n^*}{\sqrt{n}}=\infty$.

(iii) In this sense, $k_n\approx c\sqrt{n}$ is the critical time for a change-point, when one wants to obtain a consistent test. Depending on the value $c>0$, the power might asymptotically approach any value between α (the size) and 1. This holds, as the distribution of $\sup|\sigma W^{(0)}(\lambda)|$ is continuous and for $c\to c'$ we have

$$\sup_{0\le\lambda\le 1}\big|\sigma W^{(0)}(\lambda)+c(1-\lambda)u(\Delta)\big| \to \sup_{0\le\lambda\le 1}\big|\sigma W^{(0)}(\lambda)+c'(1-\lambda)u(\Delta)\big|$$

in $\mathcal{D}[0,1]$. Thus, for $c\to c'$,

$$\mathbb{P}_{(c)}\big(\max_{1\le k<n}|G_n^0(k)|>q_\alpha\big)\to \mathbb{P}_{(c')}\big(\max_{1\le k<n}|G_n^0(k)|>q_\alpha\big),$$

where q_α is the critical value depending on the asymptotical size α. For $c=0$ we have $\mathbb{P}_{(0)}\big(\max_{1\le k<n}|G_n^0(k)|>q_\alpha\big)=\alpha$ and for c large enough, we have $\mathbb{P}_{(c)}\big(\max_{1\le k<n}|G_n^0(k)|>q_\alpha\big)=1$. As the mapping

$$c\mapsto \mathbb{P}_{(c)}\big(\max_{1\le k<n}|G_n^0(k)|>q_\alpha\big)$$

is continuous, it takes any value between α and 1.

Now, we consider the case $\gamma\in(0,1/2)$, i.e., where the norming sequence depends on k. Under the alternative A_2, we determine the asymptotic distribution of the test statistic $\max_{1\le k\le n} G_n^\gamma(k)$.

Theorem 4.2. *We consider model (2) under A_2. Assume that $g(\xi_2-\xi_1)$ has finite second moments. Then, for $0<\gamma<\frac{1}{2}$ and as $n\to\infty$,*

$$\max_{1\le k<n}|G_n^\gamma(k)| \xrightarrow{\mathcal{D}} \max\left\{c^{1-\gamma}u(\Delta),\ \sup_{0\le\lambda\le 1}\frac{\sigma}{(\lambda(1-\lambda))^\gamma}\big|W^{(0)}(\lambda)\big|\right\},$$

where σ and $u(\Delta)$ are defined as in Theorem 3.1, albeit with $\Delta_n\equiv\Delta$.

For the special case of the CUSUM and Wilcoxon kernel we obtain the following corollaries.

Corollary 4.3. *Under the assumptions of Theorem 4.2, we obtain*

$$\max_{1\le k<n} \frac{1}{n^{3/2}\left(\frac{k}{n}(1-\frac{k}{n})\right)^{\gamma}}\Big|\sum_{i=1}^{k}\sum_{j=k+1}^{n}(X_j-X_i)\Big| \xrightarrow{\mathcal{D}} \max\left(c^{1-\gamma}\Delta, \sup_{0\le\lambda\le 1}\frac{\sigma}{(\lambda(1-\lambda))^{\gamma}}|W^{(0)}(\lambda)|\right),$$

where $\sigma^2 = \mathrm{Var}(\xi_1) < \infty$.

Corollary 4.4. *Under the assumptions of Theorem 4.2, we obtain*

$$\max_{1\le k<n} \frac{1}{n^{3/2}\left(\frac{k}{n}(1-\frac{k}{n})\right)^{\gamma}}\Big|\sum_{i=1}^{k}\sum_{j=k+1}^{n}\left(1_{\{X_i\le X_j\}}-\frac{1}{2}\right) \xrightarrow{\mathcal{D}} \max\left(c^{1-\gamma}\mathbb{P}(0\le \xi_2-\xi_1\le\Delta), \sup_{0\le\lambda\le 1}\frac{1}{12}\frac{1}{(\lambda(1-\lambda))^{\gamma}}|W^{(0)}(\lambda)|\right).$$

In the next theorem, we identify conditions on the limit behavior of k_n^*/n^κ that guarantee consistency of the test statistic $\max_{1\le k<n}|G_n^\gamma(k)|$. We will see that the special form of the limit distribution under the local alternative results in a peculiar behavior of the asymptotic power.

Theorem 4.3. *The change-point test with test statistic*

$$\max_{1\le k<n}|G_n^\gamma(k)|$$

is consistent if $\liminf_{n\to\infty} k_n^*/n^\kappa > (q_\alpha/u(\Delta))^{1/(1-\gamma)}$, *where* q_α *is the critical value depending on the asymptotical size* α. *In contrast, the test has asymptotically only trivial power* α *if* $\limsup_{n\to\infty} k_n^*/n^\kappa < (q_\alpha/u(\Delta))^{1/(1-\gamma)}$.

Proof. First note that in order to achieve asymptotic size α, we have to choose q_α such that

$$\mathbb{P}\left(\sup_{0\le\lambda\le 1}\frac{1}{(\lambda(1-\lambda))^{\gamma}}|W^{(0)}(\lambda)| > q_\alpha\right) = \alpha.$$

We will show that for any subseries, where exists a subsubseries $(n_j)_{j\in\mathbb{N}}$, such that the probabilities for $\max_{1\le k<n_j}|G_{n_j}^\gamma(k)| > q_\alpha$ converge to 1 respectively to α. Because the limit is the same for any subsubseries, we will then conclude that the probabilities for $\max_{1\le k<n}|G_n^\gamma(k)| > q_\alpha$ converge to 1 respectively to α. If $\liminf_{n\to\infty} k_n^*/n^\kappa > (q_\alpha/u(\Delta))^{1/(1-\gamma)}$, we can choose the subsubseries such that $k_{n_j}^* \approx c\cdot {n_j}^\kappa$ with $c > (q_\alpha/u(\Delta))^{1/(1-\gamma)},\ c<\infty$, so $c^{1-\gamma}u(\Delta) > q_\alpha$. So from Theorem 4.2, we know that the limit distribution of our test statistic is given by the distribution of

$$\max\left(c^{1-\gamma}u(\Delta), \sup_{0\le\lambda\le 1}\frac{1}{(\lambda(1-\lambda))^{\gamma}}|W^{(0)}(\lambda)|\right),$$

which exceeds the critical value q_α with probability 1.

To prove the other case, note that if $\limsup_{n\to\infty} k_n^*/n^\kappa < (q_\alpha/u(\Delta))^{1/(1-\gamma)}$, there exists a subsubseries with $k_{n_j}^* \approx c \cdot {n_j}^\kappa$ for a $c < (q_\alpha/u(\Delta))^{1/(1-\gamma)}$, so $c^{1-\gamma}u(\Delta) < q_\alpha$ and for the limit distribution, it holds that

$$\mathbb{P}\left(\max\left(c^{1-\gamma}u(\Delta), \sup_{0\le\lambda\le 1}\frac{1}{(\lambda(1-\lambda))^\gamma}|W^{(0)}(\lambda)|\right) > q_\alpha\right) = \alpha.$$

□

Remark 4. (i) It is interesting to note that for $\gamma \in (0, 1/2)$, the asymptotic power is either α or 1, unlike in the case $\gamma = 0$, where the asymptotic power can take any value in the interval $(\alpha, 1)$. This implies that a change of height Δ occurring a little too early, say at $k_n^* \approx 0.99n^\kappa(q_\alpha/\Delta)^{1/(1-\gamma)}$, will only be detected with a probability close to α in a large sample using this weighting. Meanwhile, a change with the same height occurring a little later, for example, at $k_n^* \approx 1.01n^\kappa(q_\alpha/\Delta)^{1/(1-\gamma)}$, will be detected with high probability. In contrast, for $\gamma = 0$, a change at $k_n^* \approx 0.99n^{1/2}(q_\alpha/\Delta)$ would still be detected with asymptotically nontrivial power.
(ii) For $\gamma = 1/2$, in the model A_2, even fewer observations before the change are needed to consistently detect a change of a fixed height Δ. In our paper from 2022, see [10], we have shown that $k_n^*/\log\log n \longrightarrow 0$ suffices. On the other hand, a change of height Δ_n in the middle of the data will be detected with probability going to 1 for $\gamma = 1/2$ when $\Delta_n\sqrt{n}/\log\log n \longrightarrow \infty$, while for $\gamma < 1/2$ the weaker condition $\Delta_n\sqrt{n} \longrightarrow \infty$ suffices.

4.1. Envelope power function

In this section, we calculate the envelope power function for the change-point problem with normal data. We determine the test that maximizes the power in any point (k, Δ), $1 \le k \le n-1$, $\Delta \in \mathbb{R}$ in the alternative. For simplicity, we focus on the case when $\Delta > 0$, and we assume that the variance is known. By the Neyman-Pearson fundamental lemma, the most powerful level α test for the hypothesis of no change against the alternative of a change of size Δ at time k rejects the hypothesis for large values of

$$T_k := \frac{1}{\sqrt{\sigma^2\left(\frac{1}{k}+\frac{1}{n-k}\right)}}\left(\frac{1}{n-k}\sum_{i=k+1}^{n} X_i - \frac{1}{k}\sum_{i=1}^{k} X_i\right),$$

specifically when $T_k \ge z_{1-\alpha}$, where $z_{1-\alpha}$ is the upper α-quantile of the standard normal distribution. Under the alternative (k, Δ), the test statistic T_k has a normal distribution with mean $\Delta\sqrt{\frac{k(n-k)}{n\sigma^2}}$ and variance 1. Hence, the power is given by

$$\mathbb{P}_{(k,\Delta)}(T_k \ge z_{1-\alpha}) = 1 - \Phi\left(z_{1-\alpha} - \Delta\sqrt{\frac{k(n-k)}{n\sigma^2}}\right),$$

where Φ denotes the standard normal density function. This function defines the envelope power function, i.e., the maximal power that can be attained by any level α test for the hypothesis of stationarity.

5. Simulation study

5.1. Simulations under alternative A_1

Initially, to compare the finite-sample case to the asymptotic results, Figures 1 and 2 display kernel density estimates for the CUSUM and Wilcoxon test statistics at various values of γ for a sample size of $n = 200$, provided for data following $N(0,1)$, $t(5)$, and $t(3)$ distributions. These plots also include density plots for the corresponding asymptotic distributions of the test statistics. As the Wilcoxon test is robust, the kernel density estimates are closer to the asymptotic distribution function compared to the CUSUM test statistics. For the same reason, the Wilcoxon test statistics don't show any differences in the comparison of the kernel density plots with the asymptotic density plots for the different distributed observations. However, for the CUSUM test statistic, the deviations are more significant for heavy-tailed data, i.e., the deviations are largest for $t(3)$-distributed data. No differences were observed across different values of γ.

Next, we conduct simulations to compare the power of the CUSUM and Wilcoxon test statistics, as presented in Corollaries 3.1 and 3.2, across various values of $\gamma \in [0, 1/2)$. We examine the power under the first alternative A_1 and generate $n = 1000$ independent, standard normally distributed observations with one change-point occurring after some fraction $\tau^* \in (0,1)$ of time. We consider three different jump heights, namely $\Delta = \frac{5}{\sqrt{n}}, \frac{7}{\sqrt{n}}, \frac{9}{\sqrt{n}}$. In Figure 3, the size-corrected power functions, along with the envelope power function, are plotted. Additionally, we include power curves for the test statistics with $\gamma = 1/2$. Note that, in this case, under the hypothesis of no change, we obtain another limiting distribution, namely a Gumbel extreme value distribution, as shown in, e.g., [4] or in [10] for dependent observations. Clearly, for change-points occurring at the beginning or end, the power increases with higher values of γ. However, if the change-point is around the middle of the time period, we observe higher power with smaller values of γ. Regarding jump heights, it is clear that the power improves for larger jumps at each time point. The disparity in power for different γ diminishes with higher jumps and for changes occurring in the middle of the time period. Specifically, for $\Delta = \frac{9}{\sqrt{n}}$ and a change in the middle, the power is nearly equal for $\gamma \in 0, 0.1, 0.2, 0.3, 0.4$. This is in contrast to change-points near the boundary of the time interval, where the difference in power for various γ values becomes slightly greater with higher jumps.

In Figure 4, the plots depict the difference between the power of the most powerful level $\alpha = 0.05$ test and the power of the CUSUM and Wilcoxon test statistics. Basically, we see the same behavior as in Figure 3. For larger jumps, the difference increases at the boundary and decreases in the middle of the time period. Specifically, with $\Delta = \frac{9}{\sqrt{n}}$ and changes in the middle, the differences in

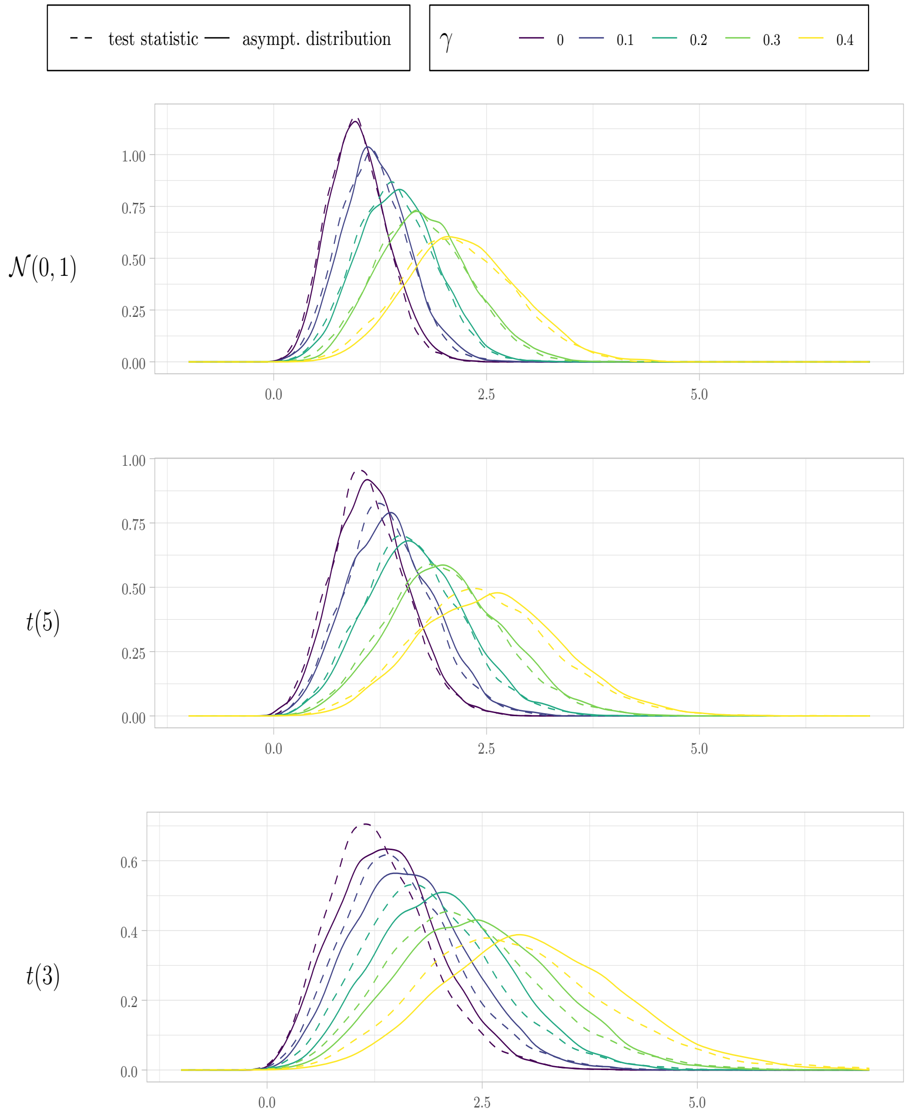

FIG. 1. *Kernel estimate density plots for the CUSUM test statistic for different values of γ compared to the density plots of the asymptotic distributions. The simulations are based on 5000 runs, with sample size $n = 200$, change-point time $\tau^* = 0.1$, and jump height $\Delta_n = \frac{7}{\sqrt{200}}$.*

power are close to zero. In other words, the CUSUM and Wilcoxon tests nearly achieve the optimal power for all γ. We observe that for a change in the middle,

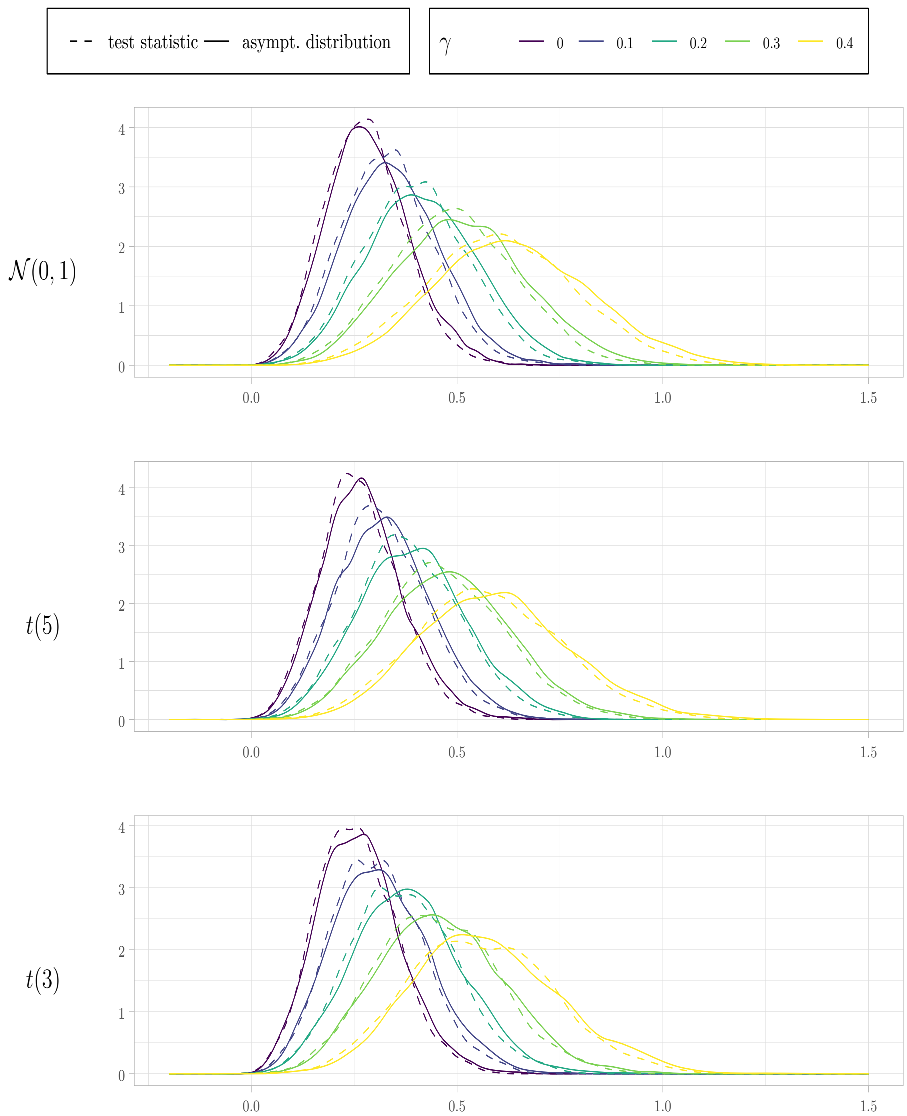

FIG. 2. *Kernel estimate density plots for the Wilcoxon test statistic for different values of* γ *compared to the density plots of the asymptotic distributions. The simulations are based on 5000 runs, with sample size* $n = 200$*, change-point time* $\tau^* = 0.1$*, and jump height* $\Delta = \frac{7}{\sqrt{200}}$.

the most significant power loss occurs with $\gamma = 0.5$. Conversely, for a change at the border, the most substantial power loss is observed with $\gamma = 0$.

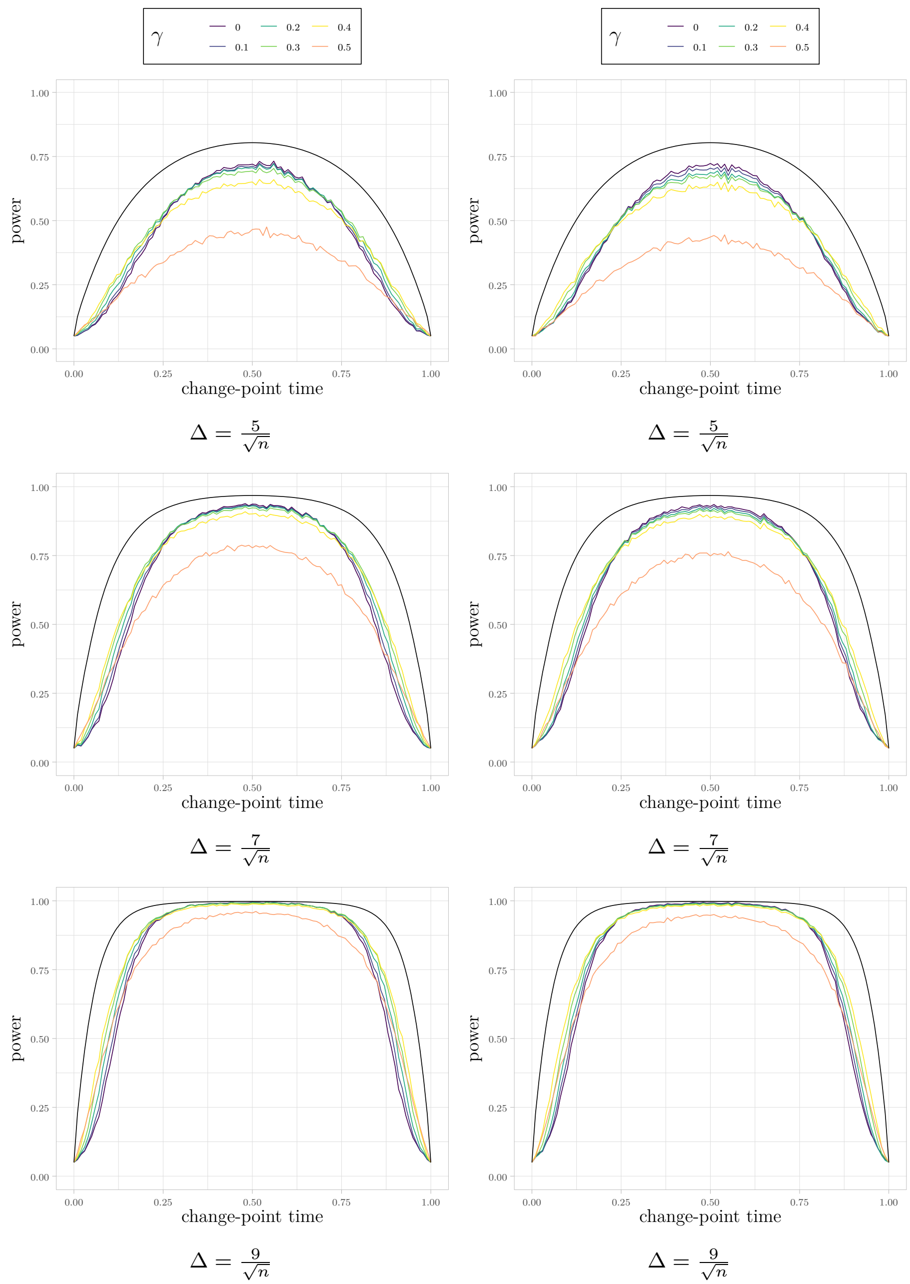

FIG. 3. *Size-corrected power for the CUSUM (left) and Wilcoxon (right) test statistics for different values of* γ. *The black line represents the envelope power function. The simulations are based on* $n = 1000$ *standard normally distributed observations and 5000 runs.*

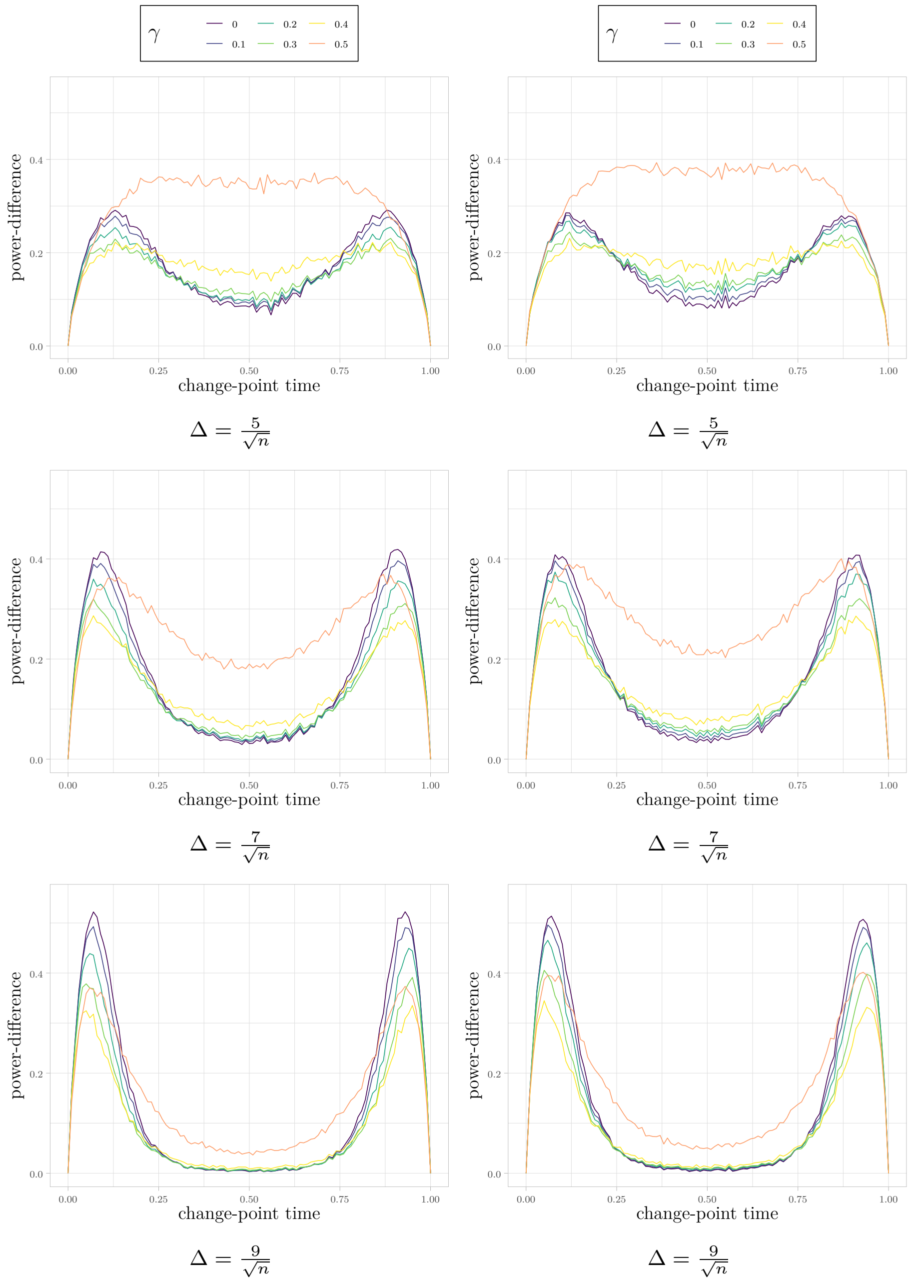

FIG. 4. *Difference between the envelope power and the size-corrected power for the CUSUM (left) and Wilcoxon (right) test statistics for different values of γ. The simulations are based on $n = 1000$ standard normally distributed observations and 5000 runs.*

A comparison of the overall-power is summarized in Table 2 for the CUSUM test statistics, and in Table 3 for the Wilcoxon test statistics. We evaluated the power obtained with the CUSUM and Wilcoxon tests in comparison to the most powerful level $\alpha = 0.05$ test. In order to compare the overall-power, i.e., the power for all $\tau^* \in (0,1)$, we considered the area under the curves in Figure 3, with the assumption that the area under the black curve (representing the envelope power function) corresponds to 100% power. As before, we considered different jump heights $\Delta = \frac{5}{\sqrt{n}}, \frac{7}{\sqrt{n}}, \frac{9}{\sqrt{n}}$. As an example, let us look at the overall-power in Table 2 for $\Delta = \frac{5}{\sqrt{n}}$. The most powerful test yields 100% power, whereas the CUSUM test with $\gamma = 0.3$ yields 74.75% power, slightly more compared to all other γ values. For $\Delta = \frac{9}{\sqrt{n}}$, the CUSUM test with $\gamma = 0.4$ yields the highest overall-power.

TABLE 2

Overall-power for the CUSUM test statistics compared to the envelope power for different values of the parameter γ and different shift heights Δ_n. The simulations are based on $n = 1000$ independent, standard normally distributed observations and 5000 runs.

Δ_n \ γ	0	0.1	0.2	0.3	0.4	0.5
$\frac{5}{\sqrt{n}}$	72.30%	72.71%	74.13%	74.75%	71.53%	50.34%
$\frac{7}{\sqrt{n}}$	78.97%	79.86%	81.45%	82.65%	81.85%	68.47%
$\frac{9}{\sqrt{n}}$	83.98%	85.14%	86.87%	88.52%	89.22%	82.97%

TABLE 3

Overall-power for the Wilcoxon test statistics compared to the envelope power for different values of the parameter γ and different shift heights Δ_n. The simulations are based on $n = 1000$ independent, standard normally distributed observations and 5000 runs.

Δ_n \ γ	0	0.1	0.2	0.3	0.4	0.5
$\frac{5}{\sqrt{n}}$	72.80%	72.20%	71.54%	72.36%	70.55%	46.76%
$\frac{7}{\sqrt{n}}$	78.98%	79.28%	79.65%	81.04%	80.92%	65.14%
$\frac{9}{\sqrt{n}}$	83.95%	84.74%	85.67%	87.38%	88.48%	80.69%

5.2. *Simulations under alternative A_2*

In Figure 5, we simulate the situation under alternative A_2, where a change-point occurs at some point in time $k_n^* \approx cn^\kappa$ with a constant change-point height $\Delta_n \equiv \Delta$. The simulations are based on $n = 5000$ (plots in Figure 5 at the top) and $n = 20000$ (plots in Figure 5 at the bottom) standard normally distributed observations with a fixed shift height $\Delta = 1$ at time $k_n^* = [cn^{2/7}]$, $0 < c < 4.3866$. In other words, we consider jumps that occur very early, namely after

$k_n^* = 1, 2, \ldots, 50$ observations in the smaller sample with $n = 5000$ and after $k_n^* = 1, 2, \ldots, 74$ in the larger sample with $n = 20000$, which corresponds to the first 1% and 0.37% of the observations, respectively. We compare the power functions for the weighted CUSUM and weighted Wilcoxon tests for different values of γ. In our simulations, we chose $\kappa = 2/7$, which corresponds to $\gamma = 0.3$. However, we also applied test statistics with smaller γ values ($\gamma = 0.1, 0.2$) and larger γ values ($\gamma = 0.4$). What we observe in our plots is that for smaller γ values, the power converges to the level $\alpha = 0.5$, and for larger γ values, the power converges to 1.

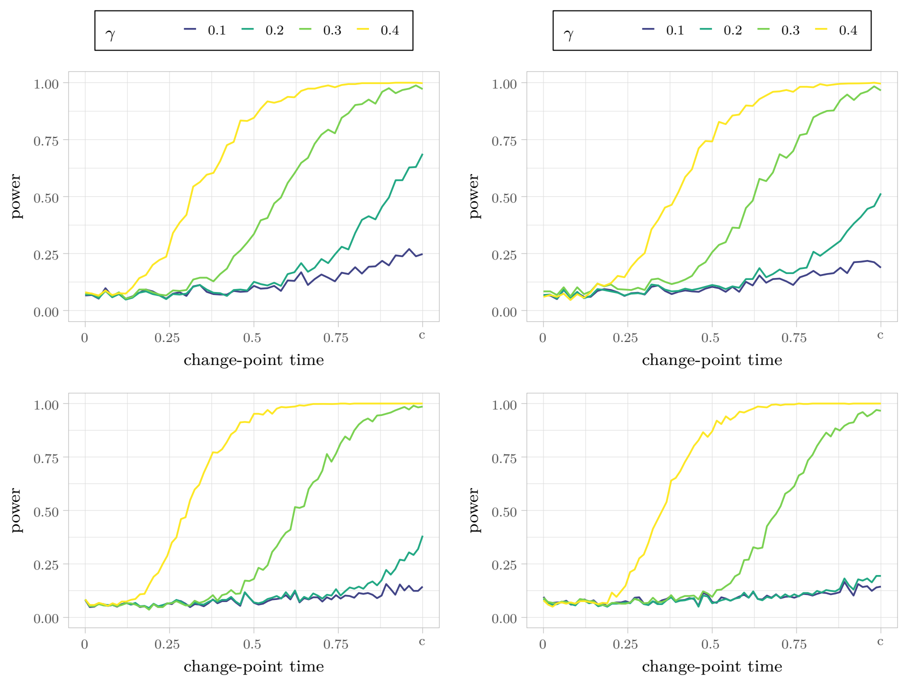

FIG. 5. *Size corrected power for the weighted CUSUM (left) and weighted Wilcoxon (right) tests for* $n = 5000$ *(top) and* $n = 20000$ *(bottom) standard normally distributed observations with a change of size* $\Delta = 1$ *at time* $k_n^* = [cn^{2/7}]$, *where* $c = \tau\tilde{c} = \tau\frac{50}{5000^{2/7}}$, $0 \le \tau \le 1$. *The simulations are based on 500 runs.*

6. Data example

We analyze the daily absolute log returns of 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, as an application to real-life data. We consider the time period from February 10, 2020, to June 26, 2020, which spans 19 weeks and comprises 95 observations (trading occurring from Monday to Friday). This dataset was analyzed in our previous paper [10], where both the non-weighted ($\gamma = 0$) and weighted ($\gamma = 1/2$) Wilcoxon test statistics were applied. In [10], the weighted Wilcoxon test statistic detected a change-point on June 18, 2020, the day when Wirecard reported that approximately 1.9 billion euros were missing from certain trust accounts. In contrast, the non-weighted Wilcoxon test ($\gamma = 0$) did not detect any change.

We apply the Wilcoxon test statistic with $\gamma \in (0.1, 0.2, 0.3, 0.4)$ to the same dataset. As mentioned in Remark 2 (iii), for dependent observations, the variance parameter must be replaced by the long-run variance, which must be estimated. We use the same estimator as in [10], a subsampling estimator introduced by Dehling et al. [7]. Similar to the procedure in [10], we split the data into three disjoint subsequences of similar length and use the median of the resulting three separate estimations to achieve consistency under the alternative. The values of the Wilcoxon test statistics and the corresponding asymptotic critical values are summarized in Table 4. The test statistics with $\gamma = 0.1$ and $\gamma = 0.2$ do not exceed the corresponding critical values, whereas the test statistics with $\gamma = 0.3$ and $\gamma = 0.4$ do. For both values of γ, the test statistics take their maximum at observation 88 (out of a total of 95 observations), consistent with the findings in [10].

TABLE 4
Value of Wilcoxon test statistics with different values of γ, applied to the Wirecard stock price data with 95 observations, together with the corresponding asymptotic critical values (considering a significance level of 5%).

	Wilcoxon test statistic	critical value
$\gamma = 0.1$	1.30	1.41
$\gamma = 0.2$	1.62	1.63
$\gamma = 0.3$	2.09	1.96
$\gamma = 0.4$	2.71	2.31

Furthermore, we investigate the minimum number of observations required for the test statistics to detect a change-point at the border. We systematically add or remove observations at the end of the time interval. For the Wilcoxon test statistic with $\gamma = 0.4$, 91 observations are sufficient for the test statistic to exceed the critical value at observation 88. With $\gamma = 0.3$, 93 observations are required, while for $\gamma = 0.2$, 96 observations are needed for the test statistic to detect a change-point, although the test statistic takes its maximum at observation 87. For $\gamma = 0.1$ and $\gamma = 0$, a minimum of 98 observations is necessary. However, in this case, the test statistics reach their maximum slightly earlier, at observation 85. Figure 6 visually represents these findings, plotting the absolute log-returns of the Wirecard stock prices. The dashed line indicates the time when the test statistic reaches its maximum, while the solid line represents the number of observations needed for the test statistic to exceed the critical value. The different colors correspond to the various values of gamma for the weighted Wilcoxon test statistic.

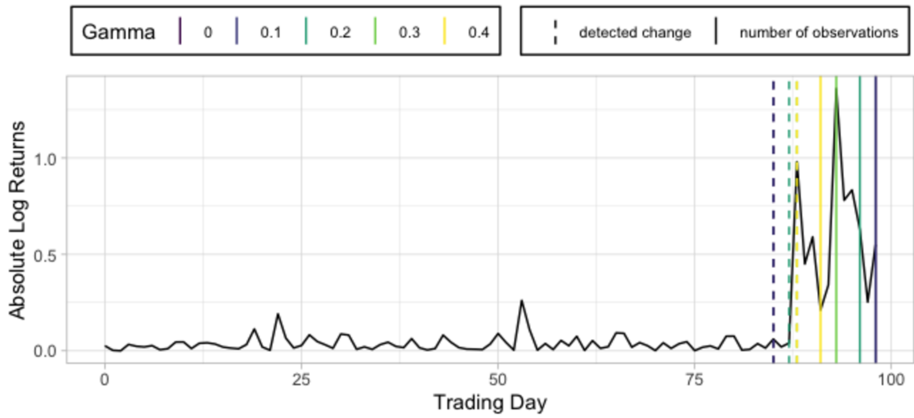

Fig. 6. *Absolute log returns of the Wirecard stock price (Febraury 10, 2020 – July 01, 2020). The dashed lines show the time of the detected change-point and the corresponding solid line the number of observations needed.*

7. Proofs

7.1. Proof of Theorem 3.1

We recall some definitions and assumptions. We assume that $(\xi_i)_{i\geq 1}$ is an i.i.d. process, and that the observations are given by

$$X_i = \begin{cases} \mu + \xi_i & \text{for } i \leq k_n^* \\ \mu + \Delta_n + \xi_i & \text{for } i \geq k_n^* + 1, \end{cases} \tag{4}$$

where μ is an unknown constant, and where $k_n^* = [n\tau^*]$, for some $\tau^* \in [0,1]$, and $\Delta_n = \frac{c}{\sqrt{n}}$. We consider a kernel of the type $g(y-x)$, where g is an odd function, i.e., $g(-x) = -g(x)$. We consider the process

$$G_n^\gamma(k) = \frac{1}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^\gamma} \frac{1}{n^{3/2}} \sum_{i=1}^{k} \sum_{j=k+1}^{n} g(X_j - X_i).$$

By (4), we obtain the following decomposition

$$\begin{aligned} G_n^\gamma(k) &= \frac{1}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^\gamma} \frac{1}{n^{3/2}} \sum_{i=1}^{k} \sum_{j=k+1}^{n} g(\xi_j - \xi_i) \\ &\quad + \frac{1}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^\gamma} \frac{1}{n^{3/2}} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(g(X_j - X_i) - g(\xi_j - \xi_i)\right) \\ &= \frac{1}{\left(\frac{k}{n}\left(1-\frac{k}{n}\right)\right)^\gamma} \left(I_n(k) + J_n(k)\right), \end{aligned}$$

where the processes $I_n(k)$ and $J_n(k)$ are defined as

$$I_n(k) = \frac{1}{n^{3/2}} \sum_{i=1}^{k} \sum_{j=k+1}^{n} g(\xi_j - \xi_i)$$
$$J_n(k) = \frac{1}{n^{3/2}} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \big(g(X_j - X_i) - g(\xi_j - \xi_i)\big).$$

We now analyze these two processes separately. Regarding $I_n(k)$, we obtain from the weighted functional central limit theorem for two-sample U-statistics that

$$\Big(\frac{1}{(\lambda(1-\lambda))^\gamma} I_n([n\lambda])\Big)_{0\le\lambda\le1} \xrightarrow{\mathcal{D}} \Big(\frac{\sigma}{(\lambda(1-\lambda))^\gamma} W^{(0)}(\lambda)\Big)_{0\le\lambda\le1,} \tag{5}$$

see Theorem 2.11 in [4]. We will analyze the limit behavior of $J_n(k)$ in two steps which we formulate as separate lemmas.

Lemma 7.1. *Under the conditions of Theorem 3.1*

$$\max_{1\le k<n} \frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^\gamma} \big|J_n(k) - \mathbb{E}(J_n(k))\big| \xrightarrow{P} 0. \tag{6}$$

Lemma 7.2. *Under the conditions of Theorem 3.1*

$$\max_{1\le k<n} \frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^\gamma} \big|\mathbb{E}(J_n(k)) - c_g \phi_{\tau^*}(\frac{k}{n})\big| \longrightarrow 0, \tag{7}$$

where $c_g = \lim_{n\to\infty} \sqrt{n}\, u(\Delta_n)$.

Proof of Lemma 7.1. Observe that by definition of the process $(X_i)_{i\ge1}$, we get

$$g(X_j - X_i) = \begin{cases} g(\xi_j - \xi_i) & \text{for } 1 \le i,\ j \le k_n^* \text{ or } k_n^*+1 \le i,\ j \le n \\ g(\xi_j - \xi_i + \Delta_n) & \text{for } 1 \le i \le k_n^*,\ k_n^*+1 \le j \le n. \end{cases} \tag{8}$$

Thus we obtain

$$J_n(k) = \begin{cases} \frac{1}{n^{3/2}} \sum_{i=1}^{k} \sum_{j=k_n^*+1}^{n} [g(\xi_j - \xi_i + \Delta_n) - g(\xi_j - \xi_i)] & \text{for } k \le k_n^* \\ \frac{1}{n^{3/2}} \sum_{i=1}^{k_n^*} \sum_{j=k+1}^{n} [g(\xi_j - \xi_i + \Delta_n) - g(\xi_j - \xi_i)] & \text{for } k \ge k_n^*+1. \end{cases} \tag{9}$$

By the Hoeffding decomposition, applied to the kernel

$$h(x,y) = g(y - x + \Delta_n) - g(y - x),$$

we obtain

$$u(\Delta_n) = \mathbb{E}(h(\xi,\eta)) = \mathbb{E}[g(\eta - \xi + \Delta_n) - g(\eta - \xi)]$$
$$h_1(x) = \mathbb{E}(h(x,\eta)) - u(\Delta_n) = \mathbb{E}[g(\eta - x + \Delta_n) - g(\eta - x)] - u(\Delta_n)$$

$$h_2(y) = \mathbb{E}(h(\xi, y)) - u(\Delta_n) = \mathbb{E}[g(y - \xi + \Delta_n) - g(y - \xi)] - u(\Delta_n)$$
$$\psi(x, y) = h(x, y) - u(\Delta_n) - h_1(x) - h_2(y),$$

where ξ and η are two independent random variables with the same distribution as ξ_1. Note that by definition,

$$h(\xi_i, \xi_j) = u(\Delta_n) + h_1(\xi_i) + h_2(\xi_j) + \psi(\xi_i, \xi_j)$$

and that all the terms on the r.h.s. are mutually uncorrelated. Then we get for $k \le k_n^*$

$$\begin{aligned}
&J_n(k) - \mathbb{E}(J_n(k)) \\
=&\frac{1}{n^{3/2}} \sum_{i=1}^{k} \sum_{j=k_n^*+1}^{n} [h_1(\xi_i) + h_2(\xi_j)] + \frac{1}{n^{3/2}} \sum_{i=1}^{k} \sum_{j=k_n^*+1}^{n} \psi(\xi_i, \xi_j) \\
=&\frac{n - k_n^*}{n^{3/2}} \sum_{i=1}^{k} h_1(\xi_i) + \frac{k}{n^{3/2}} \sum_{i=k_n^*}^{n} h_2(\xi_i) + \frac{1}{n^{3/2}} \sum_{i=1}^{k} \sum_{j=k_n^*+1}^{n} \psi(\xi_i, \xi_j) \qquad (10)
\end{aligned}$$

We will now analyze the three terms of (10) separately. Regarding the first term, using $k_n^* = [n\tau^*]$, we obtain

$$\max_{1 \le k \le k_n^*} \frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma} \frac{n-k_n^*}{n^{3/2}} \Big| \sum_{i=1}^{k} h_1(\xi_i) \Big| \le (1 - \tau^*)^{-\gamma} n^{\gamma - \frac{1}{2}} \max_{1 \le k \le k_n^*} \frac{1}{k^\gamma} \Big| \sum_{i=1}^{k} h_1(\xi_i) \Big|.$$

Note that $(k^\gamma)_{1 \le k \le k_n^*}$ is an increasing sequence, so that we may apply the Hájek-Rényi inequality, see [13], to obtain

$$\begin{aligned}
&\mathbb{P}\Big(\max_{1 \le k \le k_n^*} \frac{1}{\left(\frac{k}{n}(1 - \frac{k}{n})\right)^\gamma} \frac{n - k_n^*}{n^{3/2}} \Big| \sum_{i=1}^{k} h_1(\xi_i) \Big| \ge \epsilon \Big) \\
&\le \mathbb{P}\Big(\max_{1 \le k \le k_n^*} \frac{1}{k^\gamma} \Big| \sum_{i=1}^{k} h_1(\xi_i) \Big| \ge \epsilon (1 - \tau^*)^\gamma n^{1/2-\gamma} \Big) \\
&\le \frac{1}{\epsilon^2 (1 - \tau^*)^{2\gamma} n^{1-2\gamma}} \sum_{j=1}^{k_n^*} \frac{1}{j^{2\gamma}} \mathrm{Var}(h_1(\xi)) \\
&\le C \, \mathrm{Var}(h_1(\xi)).
\end{aligned}$$

This converges to zero for $n \to \infty$, as $\mathrm{Var}(h_1(\xi)) \to 0$.
Regarding the second term of (10), we obtain

$$\begin{aligned}
\max_{1 \le k \le k_n^*} \frac{1}{\left(\frac{k}{n}(1 - \frac{k}{n})\right)^\gamma} \frac{k}{n^{3/2}} \Big| \sum_{i=k_n^*+1}^{n} h_1(\xi_i) \Big| &\le \frac{k^{1-\gamma} n^\gamma}{(1 - \tau^*)^\gamma n^{3/2}} \Big| \sum_{i=k_n^*+1}^{n} h_1(\xi_i) \Big| \\
&\le \frac{(\tau^*)^{1-\gamma}}{(1 - \tau^*)^\gamma} \frac{1}{n^{1/2}} \Big| \sum_{i=k_n^*+1}^{n} h_1(\xi_i) \Big|.
\end{aligned}$$

Hence, using Chebychev's inequality, we obtain

$$\begin{aligned}
&\mathbb{P}\big(\max_{1\le k\le k_n^*} \frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^\gamma}\frac{k}{n^{3/2}}\big|\sum_{i=k_n^*+1}^{n} h_1(\xi_i)\big| \ge \epsilon\big)\\
&\le \mathbb{P}\big(\big|\sum_{i=k_n^*+1}^{n} h_1(\xi_i)\big| \ge C\epsilon n^{1/2}\big)\\
&\le \frac{1}{C^2\epsilon^2 n}\mathrm{Var}\big(\sum_{i=k_n^*+1}^{n} h_2(\xi_i)\big) \le C\,\mathrm{Var}(h_2(\xi_1)),
\end{aligned}$$

where $\mathrm{Var}(h_2(\xi_1)) \to 0$.
Regarding the third term of (10), the process

$$\Big(\sum_{i=1}^{k}\sum_{j=k_n^*+1}^{n}\psi(\xi_i,\xi_j)\Big)_{1\le k\le k_n^*}$$

is a martingale with respect to the filtration $\mathcal{F}_k = \sigma(\xi_1,\ldots,\ldots\xi_k,\xi_{k_n^*+1},\ldots,\xi_n)$. Clearly, $\sum_{i=1}^{k}\sum_{j=k_n^*+1}^{n}\psi(\xi_i,\xi_j)$ is adapted to $\mathcal{F}_k$. Moreover, for $m>k$,

$$\begin{aligned}
&\mathbb{E}\Big(\sum_{i=1}^{m}\sum_{j=k_n^*+1}^{n}\psi(\xi_i,\xi_j)\Big|\mathcal{F}_k\Big)\\
&=\mathbb{E}\Big(\sum_{i=1}^{k}\sum_{j=k_n^*+1}^{n}\psi(\xi_i,\xi_j)\Big|\mathcal{F}_k\Big)+\mathbb{E}\Big(\sum_{i=k+1}^{m}\sum_{j=k_n^*+1}^{n}\psi(\xi_i,\xi_j)\Big|\mathcal{F}_k\Big)\\
&=\sum_{i=1}^{k}\sum_{j=k_n^*+1}^{n}\psi(\xi_i,\xi_j),
\end{aligned}$$

as $\sum_{i=1}^{k}\sum_{j=k_n^*+1}^{n}\psi(\xi_i,\xi_j)$ is $\mathcal{F}_k$-measurable and

$$\begin{aligned}
\mathbb{E}\Big(\sum_{i=k+1}^{m}\sum_{j=k_n^*+1}^{n}\psi(\xi_i,\xi_j)\Big|\mathcal{F}_k\Big) &= \sum_{i=k+1}^{m}\sum_{j=k_n^*+1}^{n}\mathbb{E}(\psi(\xi_i,\xi_j)|\mathcal{F}_k)\\
&= \sum_{i=k+1}^{m}\sum_{j=k_n^*+1}^{n}\mathbb{E}(\psi(\xi_i,\xi_j)|\xi_j)=0,
\end{aligned}$$

where the last equality holds, as ψ is degenerate, i.e., $\mathbb{E}(\psi(\xi_1,x))=0$. Furthermore, we get $(\frac{k}{n}(1-\frac{k}{n}))^\gamma \ge n^{-\gamma}(1-\frac{k_n^*}{n}) \ge (1-\tau^*)n^{-\gamma}$, and hence

$$\frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^\gamma n^{3/2}} \le Cn^{\gamma-3/2}.$$

Thus, we finally obtain from Doob's maximal inequality

$$\mathbb{P}\Big(\max_{1\le k\le k_n^*}\frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^\gamma n^{3/2}}\big|\sum_{i=1}^{k}\sum_{j=k_n^*+1}^{n}\psi(\xi_i,\xi_j)\big|\ge\epsilon\Big)$$

$$\leq \mathbb{P}\Big(\max_{1\leq k\leq k_n^*} Cn^{\gamma-3/2}\big|\sum_{i=1}^{k}\sum_{j=k_n^*+1}^{n}\psi(\xi_i,\xi_j)\big| \geq \epsilon\Big)$$
$$\leq \mathbb{P}\Big(\max_{1\leq k\leq k_n^*}\big|\sum_{i=1}^{k}\sum_{j=k_n^*+1}^{n}\psi(\xi_i,\xi_j)\big| \geq C\epsilon n^{3/2-\gamma}\Big)$$
$$\leq \frac{1}{C\epsilon^2 n^{3-2\gamma}}\mathrm{Var}\Big(\sum_{i=1}^{k}\sum_{j=k_n^*+1}^{n}\psi(\xi_i,\xi_j)\Big)$$
$$\leq Cn^{2\gamma-3}k_n^*(n-k_n^*)\mathrm{Var}(\psi(\xi_1,\xi_2)) \leq Cn^{2\gamma-1}\mathrm{Var}(\psi(\xi_1,\xi_2))$$

Since $\gamma < 1/2$, the right hand side converges to zero as $n \to \infty$. Hence we have shown that $\max_{1\leq k\leq k_n^*} \frac{1}{(\frac{k}{n}(1-\frac{k}{n}))^\gamma}|J_n(k) - \mathbb{E}J_n(k)| \to 0$. In an analogous way, we can establish that $\max_{k_n^*\leq k<n} \frac{1}{(\frac{k}{n}(1-\frac{k}{n}))^\gamma}|J_n(k) - \mathbb{E}J_n(k)| \to 0$. □

Proof of Lemma 7.2. Observe that

$$\mathbb{E}(J_n(k)) = \begin{cases} k(n-k_n^*)\frac{1}{n^{3/2}}u(\Delta_n) & \text{for } k \leq k_n^* \\ k_n^*(n-k)\frac{1}{n^{3/2}}u(\Delta_n) & \text{for } k \geq k_n^*. \end{cases} \tag{11}$$

By the definition of $\phi_n(k)$ and $\phi_{\tau^*}(\lambda)$, we obtain

$$\mathbb{E}(J_n(k)) = \phi_n(k)\frac{1}{n^{3/2}}u(\Delta_n) = \phi_{\tau^*}(\frac{k}{n})\sqrt{n}\,u(\Delta_n).$$

Then we have

$$\max_{1\leq k<n} \frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^\gamma}\big|\mathbb{E}(J_n(k)) - c_g\,\phi_{\tau^*}(\frac{k}{n})\big|$$
$$= \max_{1\leq k<n} \frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^\gamma}\big|\sqrt{n}\ u(\Delta_n)\,\phi_{\tau^*}(\frac{k}{n}) - c_g\,\phi_{\tau^*}(\frac{k}{n})\big|$$
$$\leq \max_{1\leq k<n} \frac{1}{\big(\phi_{\tau^*}(\frac{k}{n})\big)^\gamma}\big|\sqrt{n}\ u(\Delta_n)\,\phi_{\tau^*}(\frac{k}{n}) - c_g\,\phi_{\tau^*}(\frac{k}{n})\big|$$
$$= \max_{1\leq k<n} \Big(\phi_{\tau^*}(\frac{k}{n})\Big)^{1-\gamma}\big|\sqrt{n}\ u(\Delta_n) - c_g\big|$$
$$= \big(\phi_{\tau^*}(\tau^*)\big)^{1-\gamma}\big|\sqrt{n}\ u(\Delta_n) - c_g\big|.$$

This converges to zero for $n \to \infty$, as $c_g = \lim_{n\to\infty}\sqrt{n}u(\Delta_n)$. □

Now from Lemma 7.1 and Lemma 7.2, we can deduce the limit behavior of $J_n(k)$. Together with (5), we can conclude the statement of the theorem. □

7.2. *Proof of Theorem 4.1*

Let $I_n(k)$, $J_n(k)$ and $u(\Delta)$ be defined as in the proof of Theorem 3.1. Set

$$Z_n := \max_{1\leq k<n} |I_n(k) + J_n(k)|,$$

$$Z_{n,m} := \max\left\{\max_{k\le n/m}|J_n(k)|, \max_{k>n/m}|I_n(k)+J_n(k)|\right\},$$

$$Z_{(m)} := \max\left\{|cu(\Delta)|, \sup_{\lambda>1/m}|\sigma W^{(0)}(\lambda)+c(1-\lambda)u(\Delta)|\right\},$$

$$Z := \sup_{0\le\lambda\le 1}|\sigma W^{(0)}(\lambda)+c(1-\lambda)u(\Delta)|.$$

The idea is to show

$$Z_{n,m} \xrightarrow{\mathcal{D}} Z_{(m)}, \text{ as } n\to\infty, \tag{12}$$

$$Z_{(m)} \xrightarrow{\mathcal{D}} Z, \text{ as } m\to\infty, \tag{13}$$

$$\lim_{m\to\infty}\limsup_{n\to\infty}\mathbb{P}(|Z_{n,m}-Z_n|\ge\varepsilon)=0, \tag{14}$$

and to deduce the convergence $Z_n \xrightarrow{\mathcal{D}} Z$, for $n\to\infty$, from Billingsley's triangle theorem (Theorem 3.2 in [2]).

First, we show (12). From Lemma 7.3, we know that $\max_{1\le k<n}|J_n(k) - \mathbb{E}(J_n(k))| \xrightarrow{\mathcal{P}} 0$. Thus, in order to show $\max_{k\le n/m}|J_n(k)|\to|cu(\Delta)|$, it suffices to show that $\max_{k\le n/m}|\mathbb{E}(J_n(k))| \to |cu(\Delta)|$. As indicated in the proof of Lemma 7.2, we have

$$\psi_n(k) := \mathbb{E}(J_n(k)) = \frac{\phi_n(k)}{n^{3/2}}u(\Delta) = \begin{cases} \frac{k(n-c\sqrt{n})}{n^{3/2}}u(\Delta) & \text{for } k\le k_n^* = c\sqrt{n} \\ \frac{c\sqrt{n}(n-k)}{n^{3/2}}u(\Delta) & \text{for } k\ge k_n^* = c\sqrt{n}. \end{cases}$$

As $\psi_n(k)$ is monotonically increasing for $k\le k_n^*$ and monotonically decreasing for $k\ge k_n^*$, it takes its maximum value at $k=k_n^*=c\sqrt{n}$. We obtain

$$\psi_n(k_n^*) = \frac{c\sqrt{n}(n-c\sqrt{n})}{n^{3/2}}u(\Delta) = c\big(1-\frac{1}{\sqrt{n}}\big)u(\Delta) \longrightarrow cu(\Delta), \text{ as } n\to\infty.$$

Thus, as $\frac{n}{m} > c\sqrt{n}$ for n large enough, we obtain

$$\max_{k\le n/m}|\mathbb{E}(J_n(k))| = |\psi_n(k_n^*)| \longrightarrow cu(\Delta), \text{ as } n\to\infty.$$

Moreover, we have

$$\begin{aligned}&\sup_{\lambda>1/m}\big|\mathbb{E}(J_n(\lambda n)) - c(1-\lambda)u(\Delta)\big| \\ &= \sup_{\lambda>1/m}\Big|\frac{c\sqrt{n}(n-\lambda n)}{n^{3/2}}u(\Delta) - c(1-\lambda)u(\Delta)\Big| = 0.\end{aligned}$$

Together with the functional central limit theorem for two-sample U-statistics, i.e.,

$$\Big(I_n([\lambda n])\Big)_{0\le\lambda\le 1} \xrightarrow{\mathcal{D}} \Big(\sigma W^{(0)}([\lambda n])\Big)_{0\le\lambda\le 1},$$

we can deduce weak convergence of the process $(I_n([\lambda n]) + J_n([\lambda n]))_{(\lambda\in[1/m,1])}$ to $(\sigma W^{(0)}(\lambda) + c(1-\lambda)u(\Delta))_{\lambda\in[1/m,1]}$. Thus we can conclude (12). From the continuity of $(\sigma W^{(0)}(\lambda) + c(1-\lambda)u(\Delta))_{0\le\lambda\le 1}$ and as

$$\begin{aligned}&\max\{|cu(\Delta)|, \sup_{0\le\lambda\le 1} |\sigma W^{(0)}(\lambda) + c(1-\lambda)u(\Delta)|\}\\ &= \sup_{0\le\lambda\le 1} |\sigma W^{(0)}(\lambda) + c(1-\lambda)u(\Delta)|,\end{aligned}$$

we can deduce (13). For (14), note that $|Z_{n,m} - Z_n| \le \max_{k\le n/m} |I_n(k)|$ and

$$\max_{k\le n/m} |I_n(k)| \xrightarrow{\mathcal{D}} \sup_{\lambda<1/m} |\sigma W^0(\lambda)|.$$

Thus, we obtain

$$\begin{aligned}\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}(|Z_{n,m} - Z_n| \ge \epsilon) &\le \lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big(\max_{k\le n/m} |I_n(k)| \ge \epsilon\Big)\\ &\le \lim_{m\to\infty} \mathbb{P}\Big(\sup_{\lambda<1/m} |\sigma W^{(0)}(\lambda)| \ge \epsilon\Big) = 0\end{aligned}$$

in the final step. □

7.3. *Proof of Theorem 4.2*

Define

$$\begin{aligned}I_n^\gamma(k) &:= \frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^\gamma} I_n(k),\\ J_n^\gamma(k) &:= \frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^\gamma} J_n(k),\end{aligned}$$

where $I_n(k)$ and $J_n(k)$ are defined as in the proof of Theorem 3.1. Then we have $G_n^\gamma(k) = I_n^\gamma(k) + J_n^\gamma(k)$. We proceed analogously to the proof of Theorem 4.1 and define

$$\begin{aligned}Z_n^\gamma &:= \max_{1\le k<n} |I_n^\gamma(k) + J_n^\gamma(k)|,\\ Z_{n,m}^\gamma &:= \max\left\{\max_{k\le n/m} |J_n^\gamma(k)|, \max_{k>n/m} |I_n^\gamma(k) + J_n^\gamma(k)|\right\},\\ Z_{(m)}^\gamma &:= \max\left\{c^{1-\gamma}u(\Delta), \sup_{1/m\le\lambda\le 1} \frac{\sigma}{(\lambda(1-\lambda))^\gamma}\big|W^{(0)}(\lambda)\big|\right\}\\ Z^\gamma &:= \max\left\{c^{1-\gamma}u(\Delta), \sup_{0\le\lambda\le 1} \frac{\sigma}{(\lambda(1-\lambda))^\gamma}\big|W^{(0)}(\lambda)\big|\right\}.\end{aligned}$$

To prove that $Z_n^\gamma \xrightarrow{\mathcal{D}} Z^\gamma$, for $n\to\infty$, we show

$$Z_{n,m}^\gamma \xrightarrow{\mathcal{D}} Z_{(m)}^\gamma, \text{ as } n\to\infty, \tag{15}$$

$$Z^{\gamma}_{(m)} \xrightarrow{\mathcal{D}} Z^{\gamma}, \text{ as } m \to \infty, \tag{16}$$

$$\lim_{m\to\infty} \limsup_{n\to\infty} P(|Z^{\gamma}_{n,m} - Z^{\gamma}_n| \geq \varepsilon) = 0. \tag{17}$$

First, we show (15). From Lemma 7.3 we know that $\max_{1\leq k<n} |\mathbb{E}(J^{\gamma}_n(k)) - J^{\gamma}_n(k)| \to 0$, for $n \to \infty$. Thus, in order to show $\max_{k\leq n/m} |J^{\gamma}_n(k)| \to c^{1-\gamma}u(\Delta)$, it suffices to show that $\max_{k\leq n/m} |\mathbb{E}(J^{\gamma}_n(k))| \to c^{1-\gamma}u(\Delta)$. With $k^*_n = cn^{\kappa}$ we obtain

$$\mathbb{E}(J_n(k)) = \frac{\phi_n(k)}{n^{3/2}}u(\Delta) = \begin{cases} \frac{k(n-cn^{\kappa})}{n^{3/2}}u(\Delta) & \text{for } k \leq k^*_n = cn^{\kappa} \\ \frac{cn^{\kappa}(n-k)}{n^{3/2}}u(\Delta) & \text{for } k \geq k^*_n = cn^{\kappa}. \end{cases}$$

Define $\psi^{\gamma}_n(k) := \mathbb{E}(J^{\gamma}_n(k)) = \frac{1}{(\frac{k}{n}(1-\frac{k}{n}))^{\gamma}}\mathbb{E}(J_n(k)) = \frac{n^{2\gamma}}{k^{\gamma}(n-k)^{\gamma}}\mathbb{E}(J_n(k))$. Then we have

$$\psi^{\gamma}_n(k) = \begin{cases} \frac{k^{1-\gamma}}{(n-k)^{\gamma}}\, n^{2\gamma-1/2}(1-cn^{\kappa-1})u(\Delta) & \text{for } k \leq k^*_n = cn^{\kappa} \\ \frac{(n-k)^{1-\gamma}}{k^{\gamma}}\, cn^{2\gamma+\kappa-3/2}u(\Delta) & \text{for } k \geq k^*_n = cn^{\kappa}. \end{cases} \tag{18}$$

$\psi^{\gamma}_n(k)$ is monotonically increasing for $k \leq k^*_n$ and monotonically decreasing for $k \geq k^*_n$, i.e., it takes its maximum at $k = k^*_n \approx cn^{\kappa}$ and

$$\begin{aligned} \psi^{\gamma}_n(cn^{\kappa}) &= (1-cn^{\kappa-1})^{1-\gamma}\, c^{1-\gamma}\, n^{\gamma+\kappa-\kappa\gamma-1/2}u(\Delta) \\ &= (1-cn^{\kappa-1})^{1-\gamma}\, c^{1-\gamma}u(\Delta) \longrightarrow c^{1-\gamma}u(\Delta), \end{aligned}$$

as $\gamma+\kappa-\kappa\gamma-1/2 = 0$ and $\kappa - 1 < 0$. For n so large that $\frac{n}{m} > cn^{\kappa} = k^*_n$, we obtain with the definition of $\psi^{\gamma}_n(k) = \mathbb{E}(J^{\gamma}_n(k))$ in (18), for $n \to \infty$,

$$\max_{k<n/m} \left|\psi^{\gamma}_n(k)\right| = \left|\psi^{\gamma}_n(k^*_n)\right| \longrightarrow c^{1-\gamma}u(\Delta), \tag{19}$$

$$\max_{k\geq n/m} \left|\psi^{\gamma}_n(k)\right| = \psi_n\Big(\frac{n}{m}\Big) = \frac{(n-\frac{n}{m})^{1-\gamma}}{(\frac{n}{m})^{\gamma}}cn^{2\gamma+\kappa-3/2}u(\Delta) \tag{20}$$

$$= c\big(1-\frac{1}{m}\big)^{1-\gamma}m^{\gamma}n^{\kappa-\frac{1}{2}}u(\Delta) \longrightarrow 0. \tag{21}$$

From Theorem 3 in [5] we can deduce

$$\max_{k>n/m} \left|I^{\gamma}_n(k)\right| \xrightarrow{\mathcal{D}} \sup_{1/m\leq\lambda\leq 1} \frac{1}{(\lambda(1-\lambda))^{\gamma}}|W^{(0)}(\lambda)|, \text{ as } n \to \infty.$$

Together with (19), (21), and Lemma 7.3, this implies (15). Additionally, (16) follows from the continuity of the process $(W^{(0)}(\lambda)/(\lambda(1-\lambda))^{\gamma})_{0\leq\lambda\leq 1}$. It remains to show (17). For this, note that

$$\left|Z^{\gamma}_{n,m} - Z^{\gamma}_n\right| \leq \max_{k\geq n/m} \left|I^{\gamma}_n(k)\right| + \max_{k\leq n/m} \left|J^{\gamma}_n(k)\right|.$$

The convergence to zero of the first summand is guaranteed by (21). Using Theorem 3 in [5], there is a sequence of Brownian bridges $W^{(n)}$, such that

$$P\Big(\Big|\max_{l\le n/m}\big|I_n^\gamma(l)\big| - \sup_{\lambda\le 1/m}\frac{1}{(\lambda(1-\lambda))^\gamma}|W^{(n)}(\lambda)|\Big| > \frac{\varepsilon}{2}\Big)$$
$$\le P\Big(\sup_{\lambda\le 1/m}\Big|I_n^\gamma([n\lambda]) - \frac{1}{(\lambda(1-\lambda))^\gamma}W^{(n)}(\lambda)\Big| > \frac{\varepsilon}{2}\Big) \longrightarrow 0, \text{ as } n\to\infty.$$

So we can conclude that

$$\limsup_{n\to\infty} P\left(|Z_{n,m}^\gamma - Z_n^\gamma| > \varepsilon\right)$$
$$\le \limsup_{n\to\infty}\mathbb{P}\left(\sup_{\lambda\le 1/m}\frac{1}{(\lambda(1-\lambda))^\gamma}|W^{(n)}(\lambda)| > \frac{\varepsilon}{2}\right) \xrightarrow{m\to\infty} 0,$$

because $W^{(n)}$ has the same distribution as $W^{(1)}$ and

$$\sup_{\lambda\le 1/m}\frac{1}{(\lambda(1-\lambda))^\gamma}|W^{(1)}(\lambda)|) \xrightarrow{m\to\infty} 0$$

almost surely. □

Lemma 7.3. *Under the conditions of Theorem 4.1 and 4.2 it holds for* $0\le\gamma<1/2$

$$\max_{1\le k<n}\frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma}|J_n(k) - \mathbb{E}(J_n(k))| \xrightarrow{\mathcal{P}} 0, \text{ as } n\to\infty,$$

where $J_n(k)$ *is defined as in the proof of Lemma 7.1.*

Proof of Lemma 7.3. As in the proof of Lemma 7.1, we decompose the kernel $h(x,y) = g(y-x+\Delta) - g(y-x)$ via Hoeffding's decomposition. Then we have for $k\le k_n^*$

$$\frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma}\left(J_n(k) - \mathbb{E}(J_n(k))\right)$$
$$=\frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma}\Big(\frac{n-k_n^*}{n^{3/2}}\sum_{i=1}^k h_1(\xi_i) + \frac{k}{n^{3/2}}\sum_{i=k_n^*+1}^n h_2(\xi_i) + \frac{1}{n^{3/2}}\sum_{i=1}^k\sum_{j=k_n^*+1}^n\Psi(\xi_i,\xi_j)\Big).$$

We show that the maximum of each term on the right hand side converges in probability to zero, as n goes to infinity. Recall that $k_n^*\approx cn^\kappa$, $\kappa = \frac{1-2\gamma}{2(1-\gamma)}$ and $0<\gamma<\frac{1}{2}$, i.e., $0<\kappa<\frac{1}{2}$.

Regarding the first term, we get

$$\max_{1\le k\le k_n^*}\frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma}\frac{n-k_n^*}{n^{3/2}}\Big|\sum_{i=1}^k h_1(\xi_i)\Big|$$

$$
\begin{aligned}
&= \max_{1\le k\le k_n^*} \frac{1}{\left(1-\frac{k}{n}\right)^{\gamma}} \frac{n^{\gamma}}{k^{\gamma}} \frac{(n-cn^{\kappa})}{n^{3/2}} \Big|\sum_{i=1}^{k} h_1(\xi_i)\Big| \\
&= \max_{1\le k\le k_n^*} \frac{n^{\gamma}}{\left(1-\frac{k}{n}\right)^{\gamma}} \frac{(1-cn^{\kappa-1})}{\sqrt{n}} \frac{1}{k^{\gamma}} \Big|\sum_{i=1}^{k} h_1(\xi_i)\Big| \\
&\le \frac{n^{\gamma}}{\left(1-\frac{k_n^*}{n}\right)^{\gamma}} \frac{(1-cn^{\kappa-1})}{\sqrt{n}} \max_{1\le k\le k_n^*} \frac{1}{k^{\gamma}} \Big|\sum_{i=1}^{k} h_1(\xi_i)\Big| \\
&= (1-cn^{\kappa-1})^{1-\gamma}\, n^{\gamma-1/2} \max_{1\le k\le k_n^*} \frac{1}{k^{\gamma}} \Big|\sum_{i=1}^{k} h_1(\xi_i)\Big|.
\end{aligned}
$$

As $((1/k)^{\gamma})_{1\le k\le k_n^*}$ is decreasing, we may apply the Hájek-Rényi Inequlity and obtain

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{1\le k\le k_n^*} \frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^{\gamma}} \frac{n-k_n^*}{n^{3/2}} \Big|\sum_{i=1}^{k} h_1(\xi_i)\Big| \ge \varepsilon\Big) \\
\le &\mathbb{P}\Big(\max_{1\le k\le k_n^*} \frac{1}{k^{\gamma}} \Big|\sum_{i=1}^{k} h_1(\xi_i)\Big| \ge \varepsilon (1-cn^{\kappa-1})^{\gamma-1}\, n^{1/2-\gamma}\Big) \\
\le &\frac{1}{\varepsilon^2 (1-cn^{\kappa-1})^{2\gamma-2}\, n^{1-2\gamma}} \sum_{i=1}^{k_n^*} \frac{1}{i^{2\gamma}} \mathrm{Var}(h_1(\xi_i)).
\end{aligned}
$$

As the $(\xi_i)_{i\ge1}$ are identically distributed and as $\sum_{i=1}^{k_n^*} \frac{1}{i^{2\gamma}} \le \int_0^{k_n^*} \frac{1}{x^{2\gamma}} dx$, we have

$$
\begin{aligned}
\sum_{i=1}^{k_n^*} \frac{1}{i^{2\gamma}} \mathrm{Var}(h_1(\xi_i)) &\le \mathrm{Var}(h_1(\xi_1)) \int_0^{k_n^*} \frac{1}{x^{2\gamma}} dx \\
&= \mathrm{Var}(h_1(\xi_1)) (k_n^*)^{-2\gamma+1} = \mathrm{Var}(h_1(\xi_1)) (cn^{\kappa})^{-2\gamma+1},
\end{aligned}
$$

where $\mathrm{Var}(h_1(\xi_1))$ is constant. Thus,

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{1\le k\le k_n^*} \frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^{\gamma}} \frac{n-k_n^*}{n^{3/2}} \Big|\sum_{i=1}^{k} h_1(\xi_i)\Big| \ge \varepsilon\Big) \\
\le &\frac{(cn^{\kappa})^{-2\gamma+1}}{\varepsilon^2 (1-cn^{\kappa-1})^{2\gamma-2}\, n^{1-2\gamma}} \mathrm{Var}(h_1(\xi_1)) \\
= &\frac{c^{-2\gamma+1}}{\varepsilon^2} \frac{1}{(1-cn^{\kappa-1})^{2\gamma-2}\, n^{2\gamma\kappa-\kappa-2\gamma+1}} \mathrm{Var}(h_1(\xi_1)) \longrightarrow 0, \text{ as } n\to\infty,
\end{aligned}
$$

since by our choice of κ we have $\kappa-1<0$ and $2\gamma\kappa-\kappa-2\gamma+1=\frac{1}{2(\gamma-1)}+1>0$ for $0<\gamma<1/2$. For the second term we have

$$
\max_{1\le k\le k_n^*} \frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^{\gamma}} \frac{k}{n^{3/2}} \Big|\sum_{i=k_n^*+1}^{n} h_2(\xi_i)\Big| \le \frac{(k_n^*)^{1-\gamma} n^{\gamma}}{\left(1-\frac{k_n^*}{n}\right)^{\gamma} n^{3/2}} \Big|\sum_{i=k_n^*+1}^{n} h_2(\xi_i)\Big|
$$

$$=\frac{(cn^{\kappa})^{1-\gamma}n^{\gamma}}{(1-cn^{\kappa-1})^{\gamma}n^{3/2}}\Big|\sum_{i=k_n^*+1}^{n}h_2(\xi_i)\Big| = \frac{(cn^{\kappa})^{1-\gamma}}{(1-cn^{\kappa-1})^{\gamma}n^{3/2-\gamma}}\Big|\sum_{i=k_n^*+1}^{n}h_2(\xi_i)\Big|.$$

With Chebychev's inequality, we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\max_{1\le k\le k_n^*}\frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^{\gamma}}\frac{k}{n^{3/2}}\Big|\sum_{i=k_n^*+1}^{n}h_2(\xi_i)\Big|\ge\varepsilon\Big)\\
\le&\mathbb{P}\Big(\Big|\sum_{i=k_n^*+1}^{n}h_2(\xi_i)\Big|\ge\varepsilon\frac{(1-cn^{\kappa-1})^{\gamma}n^{3/2-\gamma}}{(cn^{\kappa})^{1-\gamma}}\Big)\\
\le&\frac{1}{\varepsilon^2}\frac{(cn^{\kappa})^{2-2\gamma}}{(1-cn^{\kappa-1})^{2\gamma}n^{3-2\gamma}}\mathrm{Var}\Big(\sum_{i=k_n^*+1}^{n}h_2(\xi_i)\Big)\\
=&\frac{1}{\varepsilon^2}\frac{(cn^{\kappa})^{2-2\gamma}(n-k_n^*)}{(1-cn^{\kappa-1})^{2\gamma}n^{3-2\gamma}}\mathrm{Var}(h_2(\xi_1))=\frac{1}{\varepsilon^2}\frac{(cn^{\kappa})^{2-2\gamma}\,n(1-cn^{\kappa-1})}{(1-cn^{\kappa-1})^{2\gamma}n^{3-2\gamma}}\mathrm{Var}(h_2(\xi_1))\\
=&\frac{1}{\varepsilon^2}\frac{(cn^{\kappa})^{2-2\gamma}}{n^{2-2\gamma}}(1-cn^{\kappa-1})^{1-2\gamma}\mathrm{Var}(h_2(\xi_1))\\
=&\frac{1}{\varepsilon^2}(cn^{\kappa-1})^{2-2\gamma}(1-cn^{\kappa-1})^{1-2\gamma}\mathrm{Var}(h_2(\xi_1)).
\end{aligned}$$

This converges to zero for $n\to\infty$, as $\mathrm{Var}(h_2(\xi_i))$ is constant and as $(\kappa-1)(2-2\gamma)=-1<0$. For the third term, we use analogous arguments as in the proof of Lemma 7.1. Here we have $(\frac{k}{n}(1-\frac{k}{n}))^{\gamma}\ge n^{-\gamma}(1-\frac{k_n^*}{n})^{\gamma}=(1-cn^{\kappa-1})^{\gamma}n^{-\gamma}$, and hence

$$\frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^{\gamma}n^{3/2}}\le\frac{n^{\gamma-3/2}}{(1-cn^{\kappa-1})^{\gamma}}.$$

We apply Doob's maximal inequality and obtain

$$\begin{aligned}
&\mathbb{P}\Big(\max_{1\le k\le k_n^*}\frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^{\gamma}n^{3/2}}\Big|\sum_{i=1}^{k}\sum_{j=k_n^*}^{n}\Psi(\xi_i,\xi_j)\Big|\ge\varepsilon\Big)\\
\le&\mathbb{P}\Big(\max_{1\le k\le k_n^*}\frac{n^{\gamma-3/2}}{(1-cn^{\kappa-1})^{\gamma}}\Big|\sum_{i=1}^{k}\sum_{j=k_n^*}^{n}\Psi(\xi_i,\xi_j)\Big|\ge\varepsilon\Big)\\
\le&\mathbb{P}\Big(\max_{1\le k\le k_n^*}\Big|\sum_{i=1}^{k}\sum_{j=k_n^*}^{n}\Psi(\xi_i,\xi_j)\Big|\ge\varepsilon\frac{(1-cn^{\kappa-1})^{\gamma}}{n^{\gamma-3/2}}\Big)\\
\le&\frac{n^{2\gamma-3}}{\varepsilon^2(1-cn^{\kappa-1})^{2\gamma}}\mathrm{Var}\Big(\sum_{i=1}^{k_n^*}\sum_{j=k_n^*}^{n}\Psi(\xi_i,\xi_j)\Big)\\
=&\frac{n^{2\gamma-3}}{\varepsilon^2(1-cn^{\kappa-1})^{2\gamma}}k_n^*(n-k_n^*)\mathrm{Var}(\Psi(\xi_1,\xi_2))\\
=&\frac{n^{2\gamma-3}\,cn^{\kappa+1}(1-cn^{\kappa-1})}{\varepsilon^2(1-cn^{\kappa-1})^{2\gamma}}\mathrm{Var}(\Psi(\xi_1,\xi_2))
\end{aligned}$$

$$=\frac{c}{\varepsilon^2}(1-cn^{\kappa-1})^{1-2\gamma}\, n^{2\gamma+\kappa-2}\mathrm{Var}(\Psi(\xi_1,\xi_2)) \longrightarrow 0,$$

for $n\to\infty$, as $2\gamma+\kappa-2<0$ and $\kappa-1<0$. Altogether we have shown that

$$\max_{1\le k\le k_n^*} \frac{1}{(\frac{k}{n}(1-\frac{k}{n}))^\gamma}|J_n(k)-\mathbb{E}J_n(k)| \to 0.$$

In an analogous way, we show that $\max_{k_n^*\le k<n} \frac{1}{(\frac{k}{n}(1-\frac{k}{n}))^\gamma}|J_n(k)-\mathbb{E}J_n(k)| \to 0$. For $k_n^*\le k\le n-1$, we have $J_n(k)=\frac{1}{n^{3/2}}\sum_{i=1}^{k_n^*}\sum_{j=k+1}^{n} h(\xi_i,\xi_j)$ and

$$\begin{aligned}&\frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma}\left(J_n(k)-\mathbb{E}(J_n(k))\right)\\ =&\frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma n^{3/2}}\Big((n-k)\sum_{i=1}^{k_n^*}h_1(\xi_i)+k_n^*\sum_{i=k+1}^{n}h_2(\xi_i)+\sum_{i=1}^{k_n^*}\sum_{j=k+1}^{n}\Psi(\xi_i,\xi_j)\Big).\end{aligned}$$

Regarding the coefficient of the first term, we obtain for $k_n^*\le k\le n-1$

$$\begin{aligned}&\frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma n^{3/2}}(n-k)=\frac{n^\gamma}{k^\gamma}\frac{n^\gamma}{(n-k)^\gamma}\frac{1}{n^{3/2}}(n-k)=\frac{n^{2\gamma-3/2}}{k^\gamma}(n-k)^{1-\gamma}\\ &\le\frac{n^{2\gamma-3/2}}{(k_n^*)^\gamma}n^{1-\gamma}=\frac{n^{2\gamma-3/2}}{(cn^\kappa)^\gamma}n^{1-\gamma}=\frac{1}{c^\gamma}n^{\gamma-\kappa\gamma-1/2}.\end{aligned}$$

Thus, together with Chebyshev's inequality we obtain

$$\begin{aligned}&\mathbb{P}\Big(\max_{k_n^*\le k<n}\frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma n^{3/2}}(n-k)\Big|\sum_{i=1}^{k_n^*}h_1(\xi_i)\Big|\ge\varepsilon\Big)\\ &\le\mathbb{P}\Big(\frac{1}{c^\gamma}n^{\gamma-\kappa\gamma-1/2}\Big|\sum_{i=1}^{k_n^*}h_1(\xi_i)\Big|\ge\varepsilon\Big)\le\frac{1}{\varepsilon^2}\frac{1}{c^{2\gamma}}n^{2\gamma-2\kappa\gamma-1}\mathrm{Var}\Big(\Big|\sum_{i=1}^{k_n^*}h_1(\xi_i)\Big|\Big)\\ &=\frac{1}{\varepsilon^2}\frac{1}{c^{2\gamma}}n^{2\gamma-2\kappa\gamma-1}k_n^*\mathrm{Var}(h_1(\xi_1))=\frac{1}{\varepsilon^2}\frac{1}{c^{2\gamma-1}}n^{2\gamma-2\kappa\gamma-1+\kappa}\mathrm{Var}(h_1(\xi_1)).\end{aligned}$$

This converges to zero for $n\to\infty$, as $2\gamma-2\kappa\gamma-1+\kappa<0$ for $\gamma<1/2$ and $\kappa<1$. Regarding the second and third term, note that for $k_n^*\le k\le n-1$

$$\Big(\frac{k}{n}\big(1-\frac{k}{n}\big)\Big)^\gamma\ge\Big(\frac{k_n^*}{n}\cdot\frac{1}{n}\Big)^\gamma=c^\gamma n^{\kappa\gamma-2\gamma}$$

and hence

$$\frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma n^{3/2}}\le\frac{1}{c^\gamma n^{\kappa\gamma-2\gamma}n^{3/2}}=\frac{1}{c^\gamma}n^{2\gamma-\kappa\gamma-3/2}. \tag{22}$$

Then we get with Kolmogorov's maximal inequality

$$\mathbb{P}\Big(\max_{k_n^*\le k<n}\frac{1}{\left(\frac{k}{n}(1-\frac{k}{n})\right)^\gamma n^{3/2}}k_n^*\Big|\sum_{i=k+1}^{n}h_2(\xi_i)\Big|\ge\varepsilon\Big)$$

$$
\begin{aligned}
&\le \mathbb{P}\Big(\max_{k_n^* \le k < n} \frac{1}{c^\gamma} n^{2\gamma-\kappa\gamma-3/2} k_n^* \Big|\sum_{i=k+1}^{n} h_2(\xi_i)\Big| \ge \varepsilon\Big)\\
&= \mathbb{P}\Big(\max_{k_n^* \le k < n} c^{1-\gamma} n^{2\gamma-\kappa\gamma-3/2+\kappa} \Big|\sum_{i=k+1}^{n} h_2(\xi_i)\Big| \ge \varepsilon\Big)\\
&= \mathbb{P}\Big(\max_{1 \le k \le n-k_n^*} c^{1-\gamma} n^{2\gamma-\kappa\gamma-3/2+\kappa} \Big|\sum_{i=1}^{k} h_2(\xi_i)\Big| \ge \varepsilon\Big)\\
&\le \frac{1}{\varepsilon^2} c^{2-2\gamma}\, n^{4\gamma-2\kappa\gamma-3+2\kappa} \mathrm{Var}\Big(\Big|\sum_{i=1}^{n-k_n^*} h_2(\xi_i)\Big|\Big)\\
&= \frac{1}{\varepsilon^2} c^{2-2\gamma}\, n^{4\gamma-2\kappa\gamma-3+2\kappa} (n-k_n^*) \mathrm{Var}(h_2(\xi_1))\\
&= \frac{1}{\varepsilon^2} c^{2-2\gamma} (1-cn^{\kappa-1})\, n^{4\gamma-2\kappa\gamma-2+2\kappa} \mathrm{Var}(h_2(\xi_1)).
\end{aligned}
$$

This converges to zero for $n \to \infty$, as $4\gamma - 2\kappa\gamma - 2 + 2\kappa < 0$ for $\gamma < 1/2$. Regarding the last term, we use (22) and Doob's maximal inequality to obtain

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{k_n^* \le k < n} \frac{1}{\big(\frac{k}{n}(1-\frac{k}{n})\big)^\gamma n^{3/2}} \Big|\sum_{i=1}^{k_n^*}\sum_{j=k+1}^{n} \Psi(\xi_i,\xi_j)\Big| \ge \varepsilon\Big)\\
\le&\mathbb{P}\Big(\max_{k_n^* \le k < n} \frac{1}{c^\gamma} n^{2\gamma-\kappa\gamma-3/2} \Big|\sum_{i=1}^{k_n^*}\sum_{j=k+1}^{n} \Psi(\xi_i,\xi_j)\Big| \ge \varepsilon\Big)\\
=&\mathbb{P}\Big(\max_{1 \le k < n-k_n^*} \frac{1}{c^\gamma} n^{2\gamma-\kappa\gamma-3/2} \Big|\sum_{i=1}^{k_n^*}\sum_{j=1}^{k} \Psi(\xi_i,\xi_{k_n^*+j})\Big| \ge \varepsilon\Big)\\
\le&\frac{1}{\varepsilon^2 c^{2\gamma}} n^{4\gamma-2\kappa\gamma-3} \mathrm{Var}\Big|\sum_{i=1}^{k_n^*}\sum_{j=1}^{n-k_n^*} \Psi(\xi_i,\xi_{k_n^*+j})\Big|\\
=&\frac{1}{\varepsilon^2 c^{2\gamma}} n^{4\gamma-2\kappa\gamma-3} k_n^*(n-k_n^*) \mathrm{Var}(\Psi(\xi_1,\xi_{k_n^*+1}))\\
=&\frac{1}{\varepsilon^2 c^{2\gamma}} n^{4\gamma-2\kappa\gamma-3} cn^\kappa(n-cn^\kappa) \mathrm{Var}(\Psi(\xi_1,\xi_{k_n^*+1}))\\
=&\frac{1}{\varepsilon^2 c^{2\gamma-1}} (1-cn^{\kappa-1}) n^{4\gamma-2\kappa\gamma-2+\kappa} \mathrm{Var}(\Psi(\xi_1,\xi_{k_n^*+1})).
\end{aligned}
$$

This goes to zero for $n \to \infty$, as $4\gamma - 2\kappa\gamma - 2 + \kappa < 0$ for $\gamma < 1/2$. □

References

[1] Berkes, I., Gombay, E. and Horváth, L. (2009). Testing for changes in the covariance structure of linear processes. *Statist. Plann. Inference* **139** 2044–2063. MR2497559

[2] Billingsley, P. (1999). *Convergence of Probability Measures*, second ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. John Wiley & Sons, Inc., New York A Wiley-Interscience Publication. https://doi.org/10.1002/9780470316962 MR1700749

[3] Csörgő, M. and Horváth, L. (1988). Invariance Principles for Changepoint Problems. *J. Multivariate Anal.* **27** 151–168. https://doi.org/10.1016/0047-259X(88)90122-4 MR971179

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

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

[6] 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

[7] 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

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

[9] Dehling, H., Rooch, A. and Taqqu, M. (2017). Power of change-point tests for long-range dependent data. *Electronic Journal of Statistics* **11** 2168–2198. MR3654823

[10] Dehling, H., Vuk, K. and Wendler, M. (2022). Change-point detection based on weighted two-sample U-statistics. *Electronic Journal of Statistics* **16** 862–891. MR4372100

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

[12] Gombay, E. (2000). Comparison of U-statistics in the change-point problem and in sequential change detection. *Period. Math. Hungar.* **41** 157–166. MR1812802

[13] Hájek, J. and Rényi, A. (1955). Generalization of an inequality of Kolmogorov. *Acta Mathematica Hungarica* **6** 281–283. MR0076207

[14] Horváth, L., Miller, C. and Rice, G. (2020). A new class of change point test statistics of Rényi type. *Journal of Business & Economic Statistics* **38** 570–579. MR4115418

[15] Horváth, L., Rice, G. and Zhao, Y. (2021). Change point analysis of covariance functions: A weighted cumulative sum approach. *J. Multivariate Anal.* 104877. MR4384119

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

[17] Robbins, M., Gallagher, C., Lund, R. and Aue, A. (2011). Mean shift testing in correlated data. *J. Time Series Anal.* **32** 498–511. MR2835683
[18] Szyszkowicz, B. (1991). Changepoint problems and contiguous alternatives. *Statist. Probab. Lett.* **11** 299–308. MR1108352
[19] Xie, H., Li, D. and Xiong, L. (2014). Exploring the ability of the Pettitt method for detecting change point by Monte Carlo simulation. *Stoch. Environ. Res. Risk Assess.* **28** 1643–1655.