SFB 823

Robust estimation of change-point location

Carina Gerstenberger

Discussion Paper

Nr. 2/2017

SFB 823

Robust Estimation of Change-Point Location

Carina Gerstenberger*

We introduce a robust estimator of the location parameter for the change-point in the mean based on the Wilcoxon statistic and establish its consistency for L_1 near epoch dependent processes. It is shown that the consistency rate depends on the magnitude of change. A simulation study is performed to evaluate finite sample properties of the Wilcoxon-type estimator in standard cases, as well as under heavy-tailed distributions and disturbances by outliers, and to compare it with a CUSUM-type estimator. It shows that the Wilcoxon-type estimator is equivalent to the CUSUM-type estimator in standard cases, but outperforms the CUSUM-type estimator in presence of heavy tails or outliers in the data.

KEYWORDS: Wilcoxon statistic; change-point estimator; near epoch dependence

1 Introduction

In many applications it can not be assumed that observed data have a constant mean over time. Therefore, extensive research has been done in testing for change-points in the mean, see e.g. Giraitis *et al.* (1996), Csörgö and Horváth (1997), Ling (2007), and others. A number of papers deal with the problem of estimation of the change-point location. Bai (1994) estimates the unknown location point for the break in the mean of a linear process by the method of least squares. Antoch *et al.* (1995) and Csörgö and Horváth (1997) established the consistency rates for CUSUM-type estimators for independent data, while Csörgö and Horváth (1997) considered weakly dependent variables. Horváth and Kokoszka (1997) established consistency of CUSUM-type estimators of location of change-point for strongly dependent variables. Kokoszka and Leipus (1998, 2000) discussed CUSUM-type estimators for dependent observations and ARCH models. In spite of numerous studies on testing for changes and estimating for change-points, however, just a few procedures are robust against outliers in the data. In a recent work Dehling *et al.* (2015) address the robustness problem of testing for change-points by introducing a Wilcoxon-type test which is applicable under short-range dependence (see also Dehling *et al.* (2013) for the long-range dependence case).
In this paper we suggest a robust Wilcoxon-type estimator for the change-point location based on the idea of Dehling *et al.* (2015) and applicable for L_1 near epoch dependent

Date: January 9, 2017.
*Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany

processes. The Wilcoxon change-point test statistic is defined as

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

and counts how often an observation of the second part of the sample, $X_{k+1}, \ldots, X_n$, exceeds an observation of the first part, $X_1, \ldots, X_k$. Assuming a change in mean happens at the time k^*, the absolute value of $W_n(k^*)$ is expected to be large. Hence, the Wilcoxon-type estimator for the location of the change-point,

$$\hat{k} = \min \left\{ k : \max_{1 \le l < n} \left| W_n(l) \right| = \left| W_n(k) \right| \right\}, \tag{2}$$

can be defined as the smallest k for which the Wilcoxon test statistic $W_n(k)$ attains its maximum. Since the Wilcoxon test statistic is a rank-type statistic, outliers in the observed data can not affect the test statistic significantly. On the contrary, the CUSUM-type test statistic

$$C_n(k) = \frac{1}{k} \sum_{i=1}^{k} X_i - \frac{1}{n} \sum_{i=1}^{n} X_i,$$

which compares the difference of the sample mean of the first k observations and the sample mean over all observations, can be significantly disturbed by a single outlier.
The outline of the paper is as follows. In Section 2 we discuss the consistency and the rates of the estimator $\hat{k}$ in (2). Section 3 contains the simulation study. Section 4 provides useful properties of the Wilcoxon test statistic and the proof of the main result. Sections 5 and 6 contain some auxiliary results.

2 Definitions, assumptions and main results

Assume the random variables $X_1, \ldots, X_n$ follow the change-point model

$$X_i = \begin{cases} Y_i + \mu, & 1 \le i \le k^* \\ Y_i + \mu + \Delta_n, & k^* < i \le n, \end{cases} \tag{3}$$

where the process (Y_j) is a stationary zero mean short-range dependent process, k^* denotes the location of the unknown change-point and μ and $\mu + \Delta_n$ are the unknown means. We assume that Y_1 has a continuous distribution function F with bounded second derivative and that the distribution functions of $Y_1 - Y_k$, $k \ge 1$ satisfy

$$\mathrm{P}(x \le Y_1 - Y_k \le y) \le C|y - x|, \tag{4}$$

for all $0 \le x \le y \le 1$, where C does not depend on k and x, y. We allow the magnitude of the change Δ_n vary with the sample size n.

Assumption 2.1. a) The change-point $k^* = [n\theta]$, $0 < \theta < 1$, is proportional to the sample size n.

b) The magnitude of change Δ_n depends on the sample size n, and is such that

$$\Delta_n \to 0, \qquad n\Delta_n^2 \to \infty, \qquad n \to \infty. \tag{5}$$

Next we specify the assumptions on the underlying process (Y_j). The following definition introduces the concept of an absolutely regular process which is also known as β-mixing.

Definition 2.1. *A stationary process $(Z_j)_{j\in\mathbb{Z}}$ is called absolutely regular if*

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

as $k \to \infty$, where $\mathcal{F}_a^b$ is the σ-field generated by random variables $Z_a, \ldots, Z_b$.

The coefficients β_k are called mixing coefficients. For further information about mixing conditions see Bradley (2002). The concept of absolute regularity covers a wide range of processes. However, important processes like linear processes or AR processes might not be absolutely regular. To overcome this restriction, in this paper we discuss functionals of absolutely regular processes, i.e. instead of focusing on the absolute regular process (Z_j) itself, we consider process (Y_j) with $Y_j = f(Z_j, Z_{j-1}, Z_{j-2}, \ldots)$, where $f : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is a measurable function. The following near epoch dependence condition ensures that Y_j mainly depends on the near past of (Z_j).

Definition 2.2. *We say that stationary process (Y_j) is L_1 near epoch dependent (L_1 NED) on a stationary process (Z_j) with approximation constants a_k, $k \geq 0$, if conditional expectations $\mathrm{E}(Y_1|\mathcal{G}_{-k}^k)$, where $\mathcal{G}_{-k}^k$ is the σ-field generated by $Z_{-k}, \ldots, Z_k$, have property*

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

and $a_k \to 0$, $k \to \infty$.

Note that L_1 NED is a special case of more general L_r near epoch dependence, where approximation constants are defined using L_r norm: $\mathrm{E}\left|Y_1 - \mathrm{E}(Y_1|\mathcal{G}_{-k}^k)\right|^r \leq a_k$, $r \geq 1$. L_r NED processes are also called r-approximating functionals. In testing problems considered in this paper we allow for heavy-tailed distributions. Hence, we deal with L_1 near epoch dependence, which assumes existence of only the first moment $\mathrm{E}\,|Y_1|$. The concept of near epoch dependence is applicable e.g. to GARCH(1,1) processes, see Hansen (1991), and linear processes, see Example 2.1 below. Borovkova *et al.* (2001) provide additional examples and information about properties of L_r near epoch dependent process.

Example 2.1. Let (Y_j) be a linear process, i.e. $Y_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, where (Z_j) is white-noise process and the coefficients ψ_j, $j \geq 0$, are absolutely summable. Since (Z_j) is stationary and Z_{t-j} is $\mathcal{G}_{-k}^k$ measurable for $|t-j| \leq k$, we get

$$\mathrm{E}\,|Y_t - \mathrm{E}(Y_t|\mathcal{G}_{-k}^k)| \leq \sum_{j=k+1}^{\infty} |\psi_j|\,\mathrm{E}\,|Z_{t-j} - \mathrm{E}(Z_{t-j}|\mathcal{G}_{-k}^k)|$$

$$\leq 2 \sum_{j=k+1}^{\infty} |\psi_j|\,\mathrm{E}\,|Z_{t-j}| = 2\,\mathrm{E}\,|Z_1| \sum_{j=k+1}^{\infty} |\psi_j|.$$

Thus, the linear process (Y_j) is L_1 NED on (Z_j) with approximation constants $a_k = 2\,\mathrm{E}\,|Z_1| \sum_{j=k+1}^{\infty} |\psi_j|$.

We will assume that the process (Y_j) in (3) is L_1 near epoch dependent on some absolutely regular process (Z_j). In addition, we impose the following condition on the decay of the mixing coefficients β_k and approximation constants a_k:

$$\sum_{k=1}^{\infty} k^2 \left(\beta_k + \sqrt{a_k}\right) < \infty. \tag{6}$$

The next theorem states the rates of consistency of the Wilcoxon-type change-point estimator $\hat{k}$ given in (2) and the estimator $\hat{\theta} = \hat{k}/n$ of the true location parameter θ for the change-point $k^* = [n\theta]$.

Theorem 2.1. *Let $X_1, \ldots, X_n$ follow the change-point model (3) and Assumption 2.1 be satisfied. Assume that (Y_j) is a stationary zero mean L_1 near epoch dependent process on some absolutely regular process (Z_j) and (6) holds. Then,*

$$\left|\hat{k} - k^*\right| = O_P\Big(\frac{1}{\Delta_n^2}\Big), \tag{7}$$

and

$$\left|\hat{\theta} - \theta\right| = O_P\Big(\frac{1}{n\Delta_n^2}\Big). \tag{8}$$

The rate of consistency of $\hat{\theta}$ in (8) is given by $n\Delta_n^2$. The assumption $n\Delta_n^2 \to \infty$ in (5) implies $\hat{k} - k^* = o_P(k^*)$ and yields consistency of the estimator: $\hat{\theta} \to_p \theta$. In particular, for $\Delta_n \geq n^{-1/2+\epsilon}$, $\epsilon > 0$, the rate of consistency in (8) is $n^{2\epsilon}$: $\left|\hat{\theta} - \theta\right| = O_P\left(n^{-2\epsilon}\right)$.
The same consistency rate n^{ϵ} for the CUSUM-type change-point location estimator $\tilde{\theta}_C = \tilde{k}_C/n$, given by

$$\tilde{k}_C = \min\left\{k : \max_{1\le i\le n}\left|\sum_{j=1}^{i} X_j - \frac{i}{n}\sum_{j=1}^{n} X_j\right| = \left|\sum_{j=1}^{k} X_j - \frac{k}{n}\sum_{j=1}^{n} X_j\right|\right\}, \tag{9}$$

was established by Antoch *et al.* (1995) for independent data and by Csörgö and Horváth (1997) for weakly dependent data.

3 Simulation results

In this simulation study we compare the finite sample properties of the Wilcoxon-type change-point estimator $\hat{k}$, given in (2), with the CUSUM-type estimator $\tilde{k}_C$, given in (9). We refer to the Wilcoxon-type change-point estimator by W and to the CUSUM-type estimator by C.

We generate the sample of random variables $X_1, \ldots, X_n$ using the model

$$X_i = \begin{cases} Y_i + \mu & , 1 \le i \le k^* \\ Y_i + \mu + \Delta & , k^* < i \le n \end{cases} \tag{10}$$

where $Y_i = \rho Y_{i-1} + \epsilon_i$ is an AR(1) process. In our simulations we consider $\rho = 0.4$, which yields a moderate positive autocorrelation in X_i. The innovations ϵ_i are generated from a standard normal distribution and a Student's t-distribution with 1 degree of freedom. We consider the time of change $k^* = [n\theta]$, $\theta = 0.25, 0.5, 0.75$, the magnitude of change $\Delta = 0.5, 1, 2$ and the sample sizes $n = 50, 100, 200, 500$. All simulation results are based on 10.000 replications. Note that we report estimation results not for $\hat{k}$ and $\tilde{k}_C$, but $\hat{\theta} = \hat{k}/n$ and $\tilde{\theta}_C = \tilde{k}_C/n$.

Figure 1 contains the histogram based on the sample of 10.000 values of Wilcoxon-type estimator $\hat{\theta}$ and the CUSUM-type estimator $\tilde{\theta}_C$, for the model (10) with $\Delta = 1$, $\theta = 0.5$, $n = 50$ and independent standard normal innovations ϵ_i. Both estimation methods give very similar histograms.

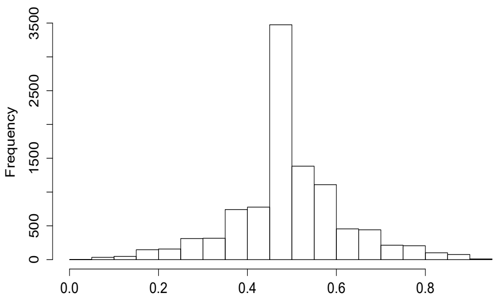
Table 1 reports the sample mean and the sample standard deviation based on 10.000 values of $\hat{\theta}$ and $\tilde{\theta}_C$ for other choices of parameters Δ and θ. It shows that performance of both estimators improves when the sample size n and the magnitude of change Δ are rising, and when the change happens in the middle of the sample. In general, Wilcoxon-type estimator performs in all experiments as good as the CUSUM-type estimator.

Figure 2 shows the histogram based on 10.000 values of $\hat{\theta}$ and $\tilde{\theta}_C$, for the model (10) with t_1-distributed heavy-tailed iid innovations ϵ_i, $\Delta = 1$, $\theta = 0.5$ and $n = 500$. For heavy-tailed innovations ϵ_i, both estimators deviate from the true value of the parameter θ more significantly than under normal innovations. Nevertheless, the Wilcoxon-type estimator seems to outperform the CUSUM-type estimator.

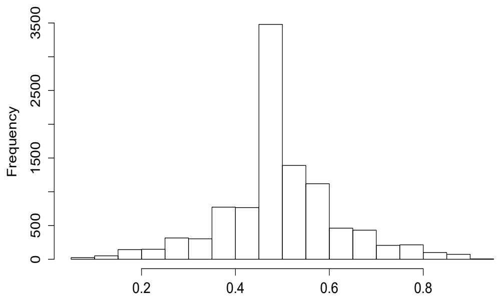
Figure 3 shows the histogram based on 10.000 values for $\hat{\theta}$ and $\tilde{\theta}_C$ when the data $X_1, \ldots, X_n$ is generated by (10) with $\Delta = 1$, $\theta = 0.5$, $n = 200$ and $\epsilon_i \sim \text{NIID}(0, 1)$ and contains outliers. The outliers are introduced by multiplying observations $X_{[0.2n]}$, $X_{[0.3n]}$, $X_{[0.6n]}$ and $X_{[0.8n]}$ by the constant $M = 50$. The histogram shows that the Wilcoxon-type estimator is rarely affected by the outliers, whereas the CUSUM-type estimator suffers large distortions.

Table 2 reports the sample mean and the sample standard deviation based on 10.000 values of $\hat{\theta}$ and $\tilde{\theta}_C$ for $\Delta = 1$ and $\theta = 0.5$ for sample size $n = 50, 100, 200, 500$ in the case of the normal, normal with outliers and t_1-distributed innovations. Figures 1, 2 and 3 presents results for $n = 50, 200, 500$.

In general, we conclude that the Wilcoxon-type change-point location estimator performs equally well as the CUSUM-type change-point estimator in standard situations, but outperforms the CUSUM-type estimator in presence of heavy tails and outliers.



(a) CUSUM $\tilde{\theta}_C$

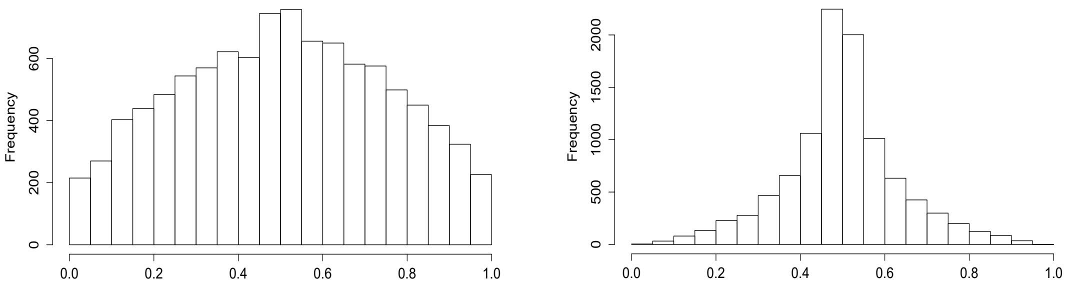


(b) Wilcoxon $\hat{\theta}$

Figure 1: Histogram based on 10.000 values for the Wilcoxon-type estimator $\hat{\theta}$ and the CUSUM-type estimator $\tilde{\theta}_C$. X_i follows the model (10) with $\Delta = 1$, $\theta = 0.5$, $n = 50$ and normal innovations $\epsilon_i \sim \text{NIID}(0, 1)$.

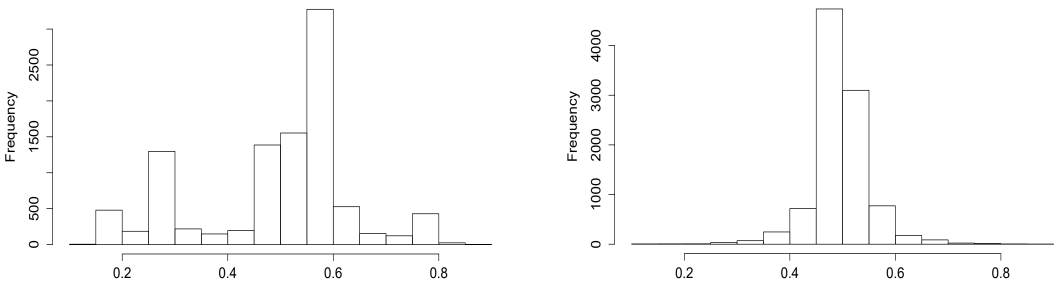
			n=50		n=100		n=200		n=500	
Δ	θ		C	W	C	W	C	W	C	W
0.5	0.25	mean	0.46	0.46	0.43	0.44	0.40	0.40	0.34	0.34
		sd	0.21	0.21	0.20	0.20	0.18	0.18	0.13	0.13
	0.50	mean	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
		sd	0.18	0.18	0.16	0.16	0.13	0.13	0.08	0.08
	0.75	mean	0.54	0.54	0.57	0.56	0.61	0.61	0.66	0.66
		sd	0.20	0.20	0.20	0.20	0.18	0.18	0.13	0.13
1	0.25	mean	0.39	0.39	0.35	0.35	0.31	0.31	0.28	0.28
		sd	0.18	0.18	0.14	0.14	0.10	0.10	0.05	0.06
	0.50	mean	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
		sd	0.12	0.12	0.09	0.09	0.05	0.05	0.02	0.02
	0.75	mean	0.61	0.60	0.65	0.65	0.69	0.69	0.72	0.72
		sd	0.17	0.17	0.15	0.15	0.10	0.10	0.05	0.06
2	0.25	mean	0.30	0.31	0.28	0.29	0.27	0.28	0.26	0.26
		sd	0.10	0.10	0.06	0.07	0.04	0.04	0.02	0.02
	0.50	mean	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
		sd	0.05	0.05	0.03	0.03	0.02	0.01	0.01	0.01
	0.75	mean	0.69	0.68	0.72	0.71	0.73	0.73	0.74	0.74
		sd	0.09	0.10	0.06	0.07	0.04	0.04	0.02	0.02

Table 1: Sample mean and the sample standard deviation based on 10.000 values of $\hat{\theta}$ and $\tilde{\theta}_C$. X_i follows the model (10) with normal innovations $\epsilon_i \sim \text{NIID}(0, 1)$.



(a) CUSUM $\tilde{\theta}_C$ (b) Wilcoxon $\hat{\theta}$

Figure 2: Histogram of CUSUM-type estimator $\tilde{\theta}_C$ and Wilcoxon-type estimator $\hat{\theta}$ based on 10.000 values of $\tilde{\theta}_C$ and $\hat{\theta}$ for the model (10) with iid t_1-distributed innovations, $\Delta = 1$, $\theta = 0.5$ and $n = 500$.



(a) CUSUM $\tilde{\theta}_C$ (b) Wilcoxon $\hat{\theta}$

Figure 3: Histogram based on 10.000 values of $\tilde{\theta}_C$ and $\hat{\theta}$ for the model (10) with normal innovations $\epsilon_i \sim \text{NIID}(0, 1)$, $\Delta = 1$, $\theta = 0.5$, $n = 200$ and outliers.

		n=50		n=100		n=200		n=500	
Innovations		C	W	C	W	C	W	C	W
normal	mean	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	sd	0.12	0.12	0.09	0.09	0.05	0.05	0.02	0.02
t_1	mean	0.52	0.50	0.51	0.50	0.51	0.50	0.50	0.50
	sd	0.23	0.20	0.24	0.19	0.24	0.17	0.25	0.14
normal with	mean	0.50	0.49	0.50	0.50	0.50	0.50	0.51	0.50
outliers	sd	0.17	0.13	0.16	0.09	0.15	0.06	0.09	0.02

Table 2: Sample mean and the sample standard deviation of $\hat{\theta}$ and $\tilde{\theta}_C$ based on 10.000 replications for the normal, normal with outliers and t_1-distributed innovations, $\Delta = 1$ and $\theta = 0.5$.

4 Useful properties of the Wilcoxon test statistic and proof of Theorem 2.1

This section presents some useful properties of the Wilcoxon test statistic and the proof of Theorem 2.1.
Throughout the paper without loss of generality, we assume that $\mu = 0$ and $\Delta_n > 0$. We let C denote a generic non-negative constant, which may vary from time to time. The notation $a_n \sim b_n$ means that two sequences a_n and b_n of real numbers have property $a_n/b_n \to c$, as $n \to \infty$, where $c \neq 0$ is a constant. $\|g\|_\infty = \sup_x |g(x)|$ stands for the supremum norm of function g. By $\xrightarrow{d}$ we denote the convergence in distribution, by $\to_p$ the convergence in probability and by $\overset{d}{=}$ we denote equality in distribution.

4.1 U-statistics and Hoeffding decomposition

The Wilcoxon test statistic $W_n(k)$ in (1) under the change-point model (3) can be decomposed into two terms

$$\begin{aligned} W_n(k) &= \sum_{i=1}^{k} \sum_{j=k+1}^{n} (1_{\{X_i \le X_j\}} - 1/2) \\ &= \begin{cases} \sum_{i=1}^{k} \sum_{j=k+1}^{n} (1_{\{Y_i \le Y_j\}} - 1/2) + \sum_{i=1}^{k} \sum_{j=k^*+1}^{n} 1_{\{Y_j < Y_i \le Y_j + \Delta_n\}}, & 1 \le k \le k^* \\ \sum_{i=1}^{k} \sum_{j=k+1}^{n} (1_{\{Y_i \le Y_j\}} - 1/2) + \sum_{i=1}^{k^*} \sum_{j=k+1}^{n} 1_{\{Y_j < Y_i \le Y_j + \Delta_n\}}, & k^* < k \le n, \end{cases} \\ &= \begin{cases} U_n(k) + U_n(k, k^*), & 1 \le k \le k^* \\ U_n(k) + U_n(k^*, k), & k^* < k \le n, \end{cases} \end{aligned} \tag{11}$$

where

$$U_n(k) = \sum_{i=1}^{k} \sum_{j=k+1}^{n} (1_{\{Y_i \le Y_j\}} - 1/2), \qquad 1 \le k \le n, \tag{12}$$

$$U_n(k, k^*) = \sum_{i=1}^{k} \sum_{j=k^*+1}^{n} 1_{\{Y_j < Y_i \le Y_j + \Delta_n\}}, \qquad 1 \le k \le k^*, \tag{13}$$

$$U_n(k^*, k) = \sum_{i=1}^{k^*} \sum_{j=k+1}^{n} 1_{\{Y_j < Y_i \le Y_j + \Delta_n\}}, \qquad k^* < k \le n. \tag{14}$$

The first term $U_n(k)$ depends only on the underlying process (Y_j), while the terms $U_n(k, k^*)$ and $U_n(k^*, k)$ depend in addition on the change-point time k^* and the magnitude Δ_n of the change in the mean.
The term $U_n(k)$ can be written as a second order U-statistic

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

with the kernel function $h(x,y) = 1_{\{x \le y\}}$ and the constant $\Theta = \mathrm{E}\, h(Y_1', Y_2') = 1/2$, where Y_1' and Y_2' are independent copies of Y_1.
We apply to $U_n(k)$ Hoeffding's decomposition of U-statistics established by Hoeffding (1948). It allows to write the kernel function as the sum

$$h(x,y) = \Theta + h_1(x) + h_2(y) + g(x,y), \tag{15}$$

where

$$h_1(x) = \mathrm{E}\, h\left(x, Y_2'\right) - \Theta = 1/2 - \mathrm{F}(x), \qquad h_2(y) = \mathrm{E}\, h\left(Y_1', y\right) - \Theta = \mathrm{F}(y) - 1/2,$$
$$g(x,y) = h(x,y) - h_1(x) - h_2(y) - \Theta.$$

By definition of h_1 and h_2, $\mathrm{E}\, h_1(Y_1) = 0$ and $\mathrm{E}\, h_2(Y_1) = 0$. Hence, $\mathrm{E}\, g(x, Y_1) = \mathrm{E}\, g(Y_1, y) = 0$, i.e. $g(x,y)$ is a degenerate kernel.
The term $U_n(k,k^*)$ in (13) (and $U_n(k^*,k)$ in (14)) can be written as a U-statistic

$$U_n(k,k^*) = \sum_{i=1}^{k} \sum_{j=k^*+1}^{n} h_n(Y_i, Y_j), \qquad 1 \le k \le k^*,$$

with the kernel $h_n(x,y) = h(x, y+\Delta_n) - h(x,y) = 1_{\{y < x \le y + \Delta_n\}}$. The Hoeffding decomposition allows to write the kernel as

$$h_n(x,y) = \Theta_{\Delta_n} + h_{1,n}(x) + h_{2,n}(y) + g_n(x,y), \tag{16}$$

with $\Theta_{\Delta_n} = \mathrm{E}\, 1_{\{Y_2' \le Y_1' \le Y_2' + \Delta_n\}}$,

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

By assumption the distribution function F of Y_1 has bounded probability density f and bounded second derivative. This allows to specify the asymptotic behaviour of Θ_{Δ_n}, as $n \to \infty$,

$$\begin{aligned} \Theta_{\Delta_n} &= \mathrm{E}\, 1_{\{Y_2' < Y_1' \le Y_2' + \Delta_n\}} = \mathrm{P}\left(Y_2' < Y_1' \le Y_2' + \Delta_n\right) \\ &= \int_{\mathbb{R}} \left(\mathrm{F}(y + \Delta_n) - \mathrm{F}(y)\right) dF(y) = \Delta_n \left(\int_{\mathbb{R}} f^2(y)\, dy + o(1) \right). \end{aligned} \tag{17}$$

Note that $\mathrm{E}\, h_{1,n}(Y_1) = 0$ and $\mathrm{E}\, h_{2,n}(Y_1) = 0$. Therefore, $g_n(x,y)$ is a degenerate kernel, i.e. $\mathrm{E}\, g_n(x, Y_1) = \mathrm{E}\, g_n(Y_1, y) = 0$. Furthermore, $\|h_{1,n}\|_\infty \to 0$, as $n \to \infty$, since

$$|h_{1,n}(x)| \le |\,\mathrm{F}(x) - \mathrm{F}(x - \Delta_n) - \Theta_{\Delta_n}| \le C\Delta_n + \Theta_{\Delta_n} \le C\Delta_n, \tag{18}$$

where $C > 0$ is a constant and $\Delta_n \to 0$, as $n \to \infty$.

4.2 1-continuity property of kernel functions h and h_n

Asymptotic properties of near epoch dependent processes (Y_j) introduced in Section 2 are well investigated in the literature, see e.g. Borovkova *et al.* (2001). In the context of change-point estimation we are interested in asymptotic properties of the variables $h(Y_i, Y_j)$, where $h(x, y) = 1_{\{x \le y\}}$ is the Wilcoxon kernel, and also in properties of the terms $h_1(Y_j)$ and $h_{1,n}(Y_j)$ of the Hoeffding decomposition of the kernels in (15) and (16). We will need to show that the variables $(h(Y_i, Y_j))$, $(h_1(Y_j))$ and $(h_{1,n}(Y_j))$ retain some properties of (Y_j). To derive them, we will use the fact that the kernels h in (15) and h_n in (16) satisfy the 1-continuity condition introduced by Borovkova *et al.* (2001).

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

$$\begin{aligned} \mathrm{E}\left(\left|h\left(Y_1, Y_k\right) - h\left(Y_1', Y_k\right)\right| 1_{\left\{\left|Y_1 - Y_1'\right| \le \epsilon\right\}}\right) &\le \phi(\epsilon), \\ \mathrm{E}\left(\left|h\left(Y_k, Y_1\right) - h\left(Y_k, Y_1'\right)\right| 1_{\left\{\left|Y_1 - Y_1'\right| \le \epsilon\right\}}\right) &\le \phi(\epsilon), \end{aligned} \tag{19}$$

and

$$\begin{aligned} \mathrm{E}\left(\left|h\left(Y_1, Y_2'\right) - h\left(Y_1', Y_2'\right)\right| 1_{\left\{\left|Y_1 - Y_1'\right| \le \epsilon\right\}}\right) &\le \phi(\epsilon), \\ \mathrm{E}\left(\left|h\left(Y_2', Y_1\right) - h\left(Y_2', Y_1'\right)\right| 1_{\left\{\left|Y_1 - Y_1'\right| \le \epsilon\right\}}\right) &\le \phi(\epsilon), \end{aligned} \tag{20}$$

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

For a univariate function $g(x)$ we define the 1-continuity property as follows.

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

$$\mathrm{E}\left(\left|g\left(Y_1\right) - g\left(Y_1'\right)\right| 1_{\left\{\left|Y_1 - Y_1'\right| \le \epsilon\right\}}\right) \le \phi(\epsilon), \tag{21}$$

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

Corollary 4.1 below establishes the 1-continuity of functions $h(x, y) = 1_{\{x \le y\}}$ and $h_n(x, y) = 1_{\{y < x \le y + \Delta_n\}}$, $n \ge 1$. For h_n, $n \ge 1$ we assume that (19) and (20) hold with the same $\phi(\epsilon)$ for all $n \ge 1$. We start the proof by showing the 1-continuity of the more general kernel function $h(x, y; t) = 1_{\{x - y \le t\}}$.

Lemma 4.1. *Let (Y_j) be a stationary process, Y_1 have distribution function F which has bounded first and second derivative and $Y_1 - Y_k$, $k \ge 1$ satisfy (4). Then the function $h(x, y; t) = 1_{\{x - y \le t\}}$ is 1-continuous with respect to the distribution function of (Y_j).*

Proof. The proof is similar to the proof of 1-continuity of the kernel function $h(x, y; t) = 1_{\{|x-y|\le t\}}$ given in Example 2.2 of Borovkova *et al.* (2001).
Note that $1_{\{Y_1-Y_k\le t\}} - 1_{\{Y_1'-Y_k\le t\}} = 0$ if $Y_1 - Y_k \le t$ and $Y_1' - Y_k \le t$; or $Y_1 - Y_k > t$ and $Y_1' - Y_k > t$. The difference is not zero if $Y_1 - Y_k \le t$ and $Y_1' - Y_k > t$; or $Y_1 - Y_k > t$ and $Y_1' - Y_k \le t$. Let $|Y_1 - Y_1'| \le \epsilon$, where $\epsilon > 0$. Then $Y_1 - Y_k < t - \epsilon$ implies $Y_1' - Y_k < t$, and $Y_1 - Y_k > t + \epsilon$ implies $Y_1' - Y_k > t$.
Hence, $|1_{\{Y_1-Y_k\le t\}} - 1_{\{Y_1'-Y_k\le t\}}|1_{\{|Y_1-Y_1'|\le\epsilon\}} \le 1_{\{t-\epsilon\le Y_1-Y_k\le t+\epsilon\}}$. Therefore,

$$\mathrm{E}\left(\left|1_{\{Y_1-Y_k\le t\}} - 1_{\{Y_1'-Y_k\le t\}}\right| 1_{\{|Y_1-Y_1'|\le\epsilon\}}\right) \le \mathrm{P}\left(t-\epsilon \le Y_1 - Y_k \le t+\epsilon\right) \le C_1\epsilon, \quad (22)$$

because of assumption (4). Similar argument yields

$$\begin{aligned}
&\mathrm{E}\left(\left|1_{\{Y_k-Y_1\le t\}} - 1_{\{Y_k-Y_1'\le t\}}\right| 1_{\{|Y_1-Y_1'|\le\epsilon\}}\right) \le \mathrm{P}\left(t-\epsilon \le Y_1 - Y_k \le t+\epsilon\right) \le C_1\epsilon,\\
&\mathrm{E}\left(\left|1_{\{Y_1-Y_2'\le t\}} - 1_{\{Y_1'-Y_2'\le t\}}\right| 1_{\{|Y_1-Y_1'|\le\epsilon\}}\right) \le \mathrm{P}(t-\epsilon \le Y_1 - Y_2' \le t+\epsilon) \le C_2\epsilon,\\
&\mathrm{E}\left(\left|1_{\{Y_2'-Y_1\le t\}} - 1_{\{Y_2'-Y_1'\le t\}}\right| 1_{\{|Y_1-Y_1'|\le\epsilon\}}\right) \le \mathrm{P}(t-\epsilon \le Y_1 - Y_2' \le t+\epsilon) \le C_2\epsilon,
\end{aligned}$$

where Y_2' is an independent copy of Y_1, noting that by the mean value theorem and $|d\,\mathrm{F}(y)/dy| \le C$,

$$\begin{aligned}
\mathrm{P}(t-\epsilon \le Y_1 - Y_2' \le t+\epsilon) &= \int_{\mathbb{R}} \left(\mathrm{F}\left(y+t+\epsilon\right) - \mathrm{F}\left(y+t-\epsilon\right)\right) dF\left(y\right)\\
&\le C\epsilon \int_{\mathbb{R}} f(y)dy = C_2\epsilon.
\end{aligned}$$

These bounds imply (19) and (20) with $\phi(\epsilon) = C\epsilon$, where C does not depend on t. This completes the proof. □

Corollary 4.1. *Assume that assumptions of Lemma* 4.1 *are satisfied. Then,*

(i) Function $h(x, y) = 1_{\{x\le y\}}$ *is* 1*-continuous with respect to the distribution function of* (Y_j)*.*

(ii) Function $h_n(x, y) = 1_{\{y<x\le y+\Delta_n\}}$ *is* 1*-continuous with respect to the distribution function of* (Y_j)*.*

Proof. (i) follows from Lemma 4.1, noting that $1_{\{x\le y\}} = h(x, y; 0)$.
(ii) We need to verify (19) and (20). Write $h_n(x, y) = h(x, y) - h(x, y + \Delta_n) = 1_{\{x\le y\}} - 1_{\{x\le y+\Delta_n\}}$. Then by (22),

$$\begin{aligned}
\mathrm{E}\left(|h_n(Y_1, Y_k) - h_n(Y_1', Y_k)|1_{\{|Y_1-Y_1'|\le\epsilon\}}\right) &\le \mathrm{E}\left(|1_{\{Y_1\le Y_k\}} - 1_{\{Y_1'\le Y_k\}}|1_{\{|Y_1-Y_1'|\le\epsilon\}}\right)\\
&+ \mathrm{E}\left(|1_{\{Y_1\le Y_k+\Delta_n\}} - 1_{\{Y_1'\le Y_k+\Delta_n\}}|1_{\{|Y_1-Y_1'|\le\epsilon\}}\right) \le C\epsilon.
\end{aligned}$$

Similar argument yields $\mathrm{E}\left(|h_n(Y_k, Y_1) - h_n(Y_k, Y_1')| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) \le C\epsilon$,

$$\mathrm{E}\left(|h_n(Y_1, Y_2') - h_n(Y_1', Y_2')| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) \le C\epsilon,$$
$$\mathrm{E}\left(|h_n(Y_2', Y_1) - h_n(Y_2', Y_1')| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) \le C\epsilon.$$

Hence, (19) and (20) hold with $\phi(\epsilon) = C\epsilon$. □

Note that condition (4) is satisfied if variables (Y_1, Y_k), $k \ge 1$, have joint probability densities that are bounded by the same constant C for all k. If the joint density does not exist, for examples of verification of condition (4) see pages 4315, 4316 of Borovkova *et al.* (2001).

Lemma 2.15 of Borovkova *et al.* (2001) yields that if a general function $h(x, y)$ is 1-continuous, i.e. satisfies (19) and (20) with function $\phi(\epsilon)$ then $\mathrm{E}\, h(x, Y_2')$, where Y_2' is an independent copy of Y_1, is also 1-continuous and satisfies the condition in (21) with the same function $\phi(\epsilon)$. Hence, $h_i(x)$ and $h_{i,n}(x)$, $i = 1, 2$ are 1-continuous and satisfy the condition in (21) with $\phi(\epsilon) = C\epsilon$.

Next we turn to 1-continuity property of $g(x, y)$. By Hoeffding decomposition (15), $g(x, y) = h(x, y) - \Theta - h_1(x) - h_2(y)$. Since $h(x, y)$, $h_1(x)$ and $h_2(x)$ in (15) are 1-continuous and satisfy (19), (20) and (21) with the same function $\phi(\epsilon) = C\epsilon$, then $g(x, y)$ is also 1-continuous with function $\phi(\epsilon) = C\epsilon$. Indeed,

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

and similarly, $\mathrm{E}\left(|g(Y_k, Y_1) - g(Y_k, Y_1')| 1_{\{|Y_1 - Y_1'| \le \epsilon\}}\right) \le 2\phi(\epsilon)$.
Using the same argument, it follows that the function $g_n(x, y) = h_n(x, y) - \Theta_{\Delta_n} - h_{1,n}(x) - h_{2,n}(x)$ in the Hoeffding decomposition (16) is also 1-continuous and satisfies (19), (20) with $\phi(\epsilon) = C\epsilon$.

4.3 NED property of $(h_1(Y_j))$ and $(h_{1,n}(Y_j))$

In Proposition 2.11 of Borovkova *et al.* (2001) it is shown that if (Y_j) is L_1 NED on a stationary absolutely regular process (Z_j) with approximation constants a_k and $g(x)$ is 1-continuous with function ϕ, then $(g(Y_j))$ is also L_1 NED on (Z_j) with approximation constants $\phi\left(\sqrt{2a_k}\right) + 2\sqrt{2a_k}||g||_\infty$.
Thus, the processes $(h_1(Y_j))$ and $(h_2(Y_j))$ in (15) and $(h_{1,n}(Y_j))$ and $(h_{2,n}(Y_j))$ in (16) are L_1 NED processes with approximation constants $a_k' = C\sqrt{a_k}$.

Corollary 3.2 of Wooldridge and White (1988) provides a functional central limit theorem for partial sum process $\sum_{i=1}^{k} \tilde{Y}_i$, $k \ge 1$, where $(\tilde{Y}_j)$ is L_2 NED on a strongly mixing process $(\tilde{Z}_j)$. To apply this result to $(h_1(Y_j))$ which is L_1 NED on (Z_j) with approximation constants a'_k, we need to show that $(h_1(Y_j))$ is also L_2 NED process. Note that the variables $\eta_k := h_1(Y_1) - \mathrm{E}(h_1(Y_1)|\mathcal{G}_{-k}^k)$ have property

$$\begin{aligned}\mathrm{E}\,\eta_k^2 &= \mathrm{E}\left(\eta_k^2 1_{\left\{|\eta_k|\le a_k'^{-\frac{1}{2}}\right\}}\right) + \mathrm{E}\left(\eta_k^2 1_{\left\{|\eta_k|> a_k'^{-\frac{1}{2}}\right\}}\right)\\ &\le a_k'^{-\frac{1}{2}}\,\mathrm{E}\,|\eta_k| + \sqrt{a'_k}\,\mathrm{E}\,|\eta_k|^4 \le \sqrt{a'_k} + \sqrt{a'_k}C =: a''_k.\end{aligned}$$

The last inequality holds, because by Definition 2.2 of L_1 near epoch dependence, $\mathrm{E}\,|h_1(Y_1) - \mathrm{E}(h_1(Y_1)|\mathcal{G}_{-k}^k)| \le a'_k$ and because $|h_1(Y_1)| \le 1/2$. Therefore the process $(h_1(Y_j))$ is L_2 NED on (Z_j) with approximation constant a''_k. Since absolute regular process (Z_j) is strongly mixing process, from Corollary 3.2 of Wooldridge and White (1988), we obtain

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

where $W(t)$ is a Brownian motion and $\sigma^2 = \sum_{k=-\infty}^{\infty} \mathrm{Cov}(h_1(Y_0), h_1(Y_k))$.
Since $h_2(x) = -h_1(x)$, all properties of $(h_1(Y_j))$ remain valid also for $(h_2(Y_j))$.

4.4 Proof of Theorem 2.1

First we show consistency property $|k^* - \hat{k}| = o_P(k^*)$ of the estimate $\hat{k} = \mathrm{argmax}_{1\le k\le n} |W_n(k)|$. To prove it, we verify that for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left(|k^* - \hat{k}| \le \epsilon k^*\right) = 1. \tag{23}$$

This means that the estimated value $\hat{k}$ with probability tending to 1 is in a neighbourhood of the true value k^*:

$$\mathrm{P}\left(\hat{k} \in [k^*(1-\epsilon), k^*(1+\epsilon)]\right) \to 1.$$

We will show that as $n \to \infty$,

$$\mathrm{P}\left(\max_{k:|k-k^*|\ge\epsilon k^*} |W_n(k)| < |W_n(k^*)|\right) \to 1. \tag{24}$$

Since $|W_n(k^*)| \le \max_{k:|k^*-k|\le\epsilon k^*} |W_n(k)|$, this proves (23).

By (11),

$$W_n(k) = \begin{cases} U_n(k) + U_n(k, k^*), & 1 \le k \le k^* \\ U_n(k) + U_n(k^*, k), & k^* < k \le n. \end{cases}$$

Theorem 6.1 implies $\max_{1\le k\le n} |U_n(k)| = O_P\left(n^{3/2}\right)$ and Proposition 5.1 below yields

$$\begin{aligned}\max_{1\le k\le k^*} |U_n(k, k^*) - k\,(n-k^*)\,\Theta_{\Delta_n}| &= o_P\left(n^{3/2}\right),\\ \max_{k^*\le k\le n} |U_n(k^*, k) - k^*\,(n-k)\,\Theta_{\Delta_n}| &= o_P\left(n^{3/2}\right).\end{aligned}$$

Hence,

$$
\begin{aligned}
W_n(k^*) &= k^*(n-k^*)\Theta_{\Delta_n} + (U_n(k^*,k^*) - k^*(n-k^*)\Theta_{\Delta_n}) + U_n(k^*) \\
&= k^*(n-k^*)\Theta_{\Delta_n} + O_P(n^{3/2}), \\
\max_{1\le k\le k^*(1-\epsilon)} |W_n(k)| &\le (1-\epsilon)k^*(n-k^*)\Theta_{\Delta_n} + O_P(n^{3/2}), \\
\max_{(1+\epsilon)k^*\le k\le n} |W_n(k)| &\le k^*(n-(1+\epsilon)k^*)\Theta_{\Delta_n} + O_P(n^{3/2}).
\end{aligned}
$$

Thus,

$$
|W_n(k^*)| - \max_{k:|k^*-k|\ge\epsilon k^*} |W_n(k)| \ge \epsilon\delta_n + O_P(n^{3/2}),
$$

where $\delta_n = k^* \min(n-k^*,k^*)\Theta_{\Delta_n}$.
By definition $k^* = [n\theta] \sim n\theta$, and by (17) and (5), $\sqrt{n}\Theta_{\Delta_n} \sim c\sqrt{n}\Delta_n \to \infty$. Hence, $\delta_n^{-1} = o(n^{-3/2})$ and $\epsilon\delta_n + O_P(n^{3/2}) = \epsilon\delta_n(1 + O_P(n^{3/2}\delta_n^{-1})) = \epsilon\delta_n(1+o_P(1))$ which proves (24).

Next we establish the rate of convergence in (7), $k^* - k = O_P(1/\Delta_n^2)$. Set $a(n) = \frac{M}{\Delta_n^2}$. Then for fixed $M > 0$, $a(n) \to \infty$, as $n \to \infty$. We will verify that

$$
\lim_{n\to\infty} \mathrm{P}\left(|k^* - \hat{k}| \le a(n)\right) \to 1, \qquad \text{as } M \to \infty,
$$

which implies (7). As in (24), we prove this by showing

$$
\lim_{n\to\infty} \mathrm{P}\left(\max_{k:|k-k^*|\ge a(n)} |W_n(k)| < |W_n(k^*)|\right) \to 1, \qquad \text{as } M \to \infty. \tag{25}
$$

Define $V_k := W_n^2(k) - W_n^2(k^*)$. If $|W_n(k)|$ attains its maximum at k', it is easy to see that V_k attains its maximum at the same k'. Hence, $\hat{k} = \min\{k : \max_{1\le l\le n}|W_n(l)| = |W_n(k)|\} = \min\{k : \max_{1\le l\le n} V_l = V_k\}$. Thus, instead of (25) it remains to show that

$$
\lim_{n\to\infty} \mathrm{P}\left(\max_{k:|k-k^*|\ge a(n)} V_k < 0\right) \to 1, \qquad M \to \infty. \tag{26}
$$

Define $\tilde{k} := \min\{k : |k-k^*| \le \epsilon k^*; V_k = \max_{n\alpha\le l\le n\beta} V_l\}$. Since by (23) $\hat{k}$ is a consistent estimator of k^*, it holds $\lim_{n\to\infty} \mathrm{P}(\hat{k} = \tilde{k}) = 1$.
So, in the proof of (26) it suffices to consider max over k, such that $|k-k^*| \le \epsilon k^*$, $|k-k^*| \ge a(n)$, which corresponds to $(1-\epsilon)k^* \le k \le k^* - a(n)$ and $k^* + a(n) < k \le (1+\epsilon)k^*$.

Let us start with $(1-\epsilon)k^* \le k \le k^* - a(n)$. Since $k^* - k > 0$, relation (26) holds for such k, if

$$
\lim_{n\to\infty} \mathrm{P}\left(\max_{(1-\epsilon)k^*\le k\le k^*-a(n)} \frac{V_k}{(n(k^*-k))^2} < 0\right) \to 1, \qquad M \to \infty. \tag{27}
$$

Note that

$$\frac{-V_k}{(n(k^*-k))^2} = \frac{W_n^2(k^*) - W_n^2(k)}{(n(k^*-k))^2}$$
$$= -\Big(\frac{W_n(k^*) - W_n(k)}{n(k^*-k)}\Big)^2 + 2\frac{W_n(k^*) - W_n(k)}{n(k^*-k)}\frac{W_n(k^*)}{n(k^*-k)}. \qquad (28)$$

By (11), $W_n(k) = U_n(k) + U_n(k, k^*)$. Then,

$$\frac{W_n(k^*) - W_n(k)}{n(k^*-k)} = \frac{n-k^*}{n}\Theta_{\Delta_n} + \delta_{1,k} + \delta_{2,k},$$

where

$$\delta_{1,k} = \frac{U_n(k^*) - U_n(k)}{n(k^*-k)}, \qquad \delta_{2,k} = \frac{U_n(k^*, k^*) - U_n(k, k^*)}{n(k^*-k)} - \frac{n-k^*}{n}\Theta_{\Delta_n}.$$

Observe that by (17), $\Theta_{\Delta_n} \sim c_*\Delta_n$, $c_* > 0$, and $k^*/n \to \theta$. Therefore, $(n-k^*)/n\Theta_{\Delta_n} \sim c_0\Delta_n$, where $c_0 = (1-\theta)c_*$. Moreover, $\max_{1\le k\le k^*-a(n)} |\delta_{i,k}| = o_P(\Delta_n)$, $i = 1, 2$, by (47) and (48) of Lemma 5.4. Hence,

$$\frac{W_n(k^*) - W_n(k)}{n(k^*-k)} = c_0\Delta_n(1 + o_P(1)), \qquad \Big(\frac{W_n(k^*) - W_n(k)}{n(k^*-k)}\Big)^2 = c_0^2\Delta_n^2(1 + o_P(1)). \qquad (29)$$

In turn,

$$\frac{W_n(k^*)}{n(k^*-k)} = \frac{U_n(k^*, k^*)}{n(k^*-k)} + \frac{U_n(k^*)}{n(k^*-k)}$$

and

$$\frac{U_n(k^*, k^*)}{n} = \frac{k^*(n-k^*)\Theta_{\Delta_n}}{n} + \frac{\delta_{3,k}}{n},$$

where $\delta_{3,k} = U_n(k^*, k^*) - k^*(n-k^*)\Theta_{\Delta_n}$. By Proposition 5.1, $\max_{1\le k\le k^*} \delta_{3,k}/n = o_P(n^{1/2})$. Since

$$\frac{k^*(n-k^*)\Theta_{\Delta_n}}{n} \sim k^*c_0\Delta_n \sim \theta c_0 n\Delta_n$$

and $\sqrt{n} = o(n\Delta_n)$, this implies

$$\frac{U_n(k^*, k^*)}{n} = k^*c_0\Delta_n(1 + o_P(1)).$$

Next, by Theorem 6.1 below, $U_n(k^*) = O_P(n^{3/2})$, and hence, $U_n(k^*)/n = O_P(n^{1/2})$. Therefore, $W_n(k^*)/n = k^*c_0\Delta_n(1 + o_P(1))$. Hence, for $(1-\epsilon)k^* \le k \le k^* - a(n)$,

$$\frac{W_n(k^*)}{n(k^*-k)} = \frac{k^*c_0\Delta_n(1 + o_P(1))}{k^*-k} \ge \frac{1}{\epsilon}c_0\Delta_n(1 + o_P(1)). \qquad (30)$$

Using (29) and (30) in (28), it follows

$$-\frac{V_k}{(n(k^*-k))^2} \ge \frac{2}{\epsilon}c_0^2\Delta_n^2(1 + o_P(1)) - c_0^2\Delta_n^2(1 + o_P(1)) \ge \Big(\frac{2}{\epsilon} - 1\Big)(c_0\Delta_n)^2(1 + o_P(1)) > 0.$$

This proves (27). Similar argument yields

$$\lim_{n\to\infty} \mathrm{P}\left(\max_{k^*+a(n)\le k\le k^*(1+\epsilon)} V_k < 0\right) \to 1, \qquad M\to\infty,$$

which completes the proof of (26) and the theorem. □

5 Auxiliary results

This section contains auxiliary results used in the proof of Theorem 2.1.
We establish asymptotic properties of the quantities $U_n(k)$, $U_n(k,k^*)$ and $U_n(k^*,k)$ defined in (12)-(14) and appearing in the decomposition (11) of $W_n(k)$.

The following lemma derives a Hájek-Rényi type inequality for L_1 NED random variables.

Lemma 5.1. *Let (Y_j) be a stationary L_1 near epoch dependent process on some absolutely regular process (Z_j), satisfying (6). Assume that $\mathrm{E}\, Y_j = 0$ and $|Y_j| \le K \le \infty$ a.s. for some $K \ge 0$. Then, for all fixed $\epsilon > 0$, for all $1 \le m \le n$,*

$$\mathrm{P}\left(\max_{m\le k\le n} \frac{1}{k}\left|\sum_{i=1}^{k} Y_i\right| > \epsilon\right) \le \frac{1}{\epsilon^2}\frac{C}{\sqrt{m}}, \tag{31}$$

where $C > 0$ does not depend on m, n, ϵ.

Proof. To prove (31), we use the Hájek-Rényi type inequality of Theorem 6.3 established in Kokoszka and Leipus (2000),

$$\begin{aligned}\epsilon^2\,\mathrm{P}\left(\max_{m\le k\le n} \frac{1}{k}\left|\sum_{i=1}^{k} Y_i\right| > \epsilon\right) &\le \frac{1}{m^2}\,\mathrm{E}\left(\sum_{i=1}^{m} Y_i\right)^2 + \sum_{k=m}^{n}\left|\frac{1}{(k+1)^2}-\frac{1}{k^2}\right|\mathrm{E}\left(\sum_{i=1}^{k} Y_i\right)^2 \\ &\quad + 2\sum_{k=m}^{n}\frac{1}{(k+1)^2}\,\mathrm{E}\left(|Y_{k+1}|\left|\sum_{j=1}^{k} Y_j\right|\right) + \sum_{k=m}^{n}\frac{1}{(k+1)^2}\,\mathrm{E}\, Y_{k+1}^2. \end{aligned}\tag{32}$$

First we bound $\mathrm{E}\left(\sum_{i=1}^{k} Y_i\right)^2$. Under assumptions of this lemma, by Lemma 6.1 below, for $i, j \ge 0$

$$|\mathrm{Cov}\,(Y_i, Y_{i+j})| = |\mathrm{E}\,(Y_i Y_{i+j})| \le 4K a_{\lfloor \frac{j}{3}\rfloor} + 2K^2 \beta_{\lfloor \frac{j}{3}\rfloor} \le C(a_{\lfloor \frac{j}{3}\rfloor} + \beta_{\lfloor \frac{j}{3}\rfloor}). \tag{33}$$

By stationarity of (Y_j),

$$|\mathrm{E}\,(Y_i Y_j)| = |\mathrm{Cov}\,(Y_i, Y_j)| = \left|\mathrm{Cov}\left(Y_0, Y_{|i-j|}\right)\right|.$$

Hence,

$$\mathrm{E}\Big(\sum_{i=1}^{k} Y_i\Big)^2 = \sum_{i,j=1}^{k} \mathrm{E}\,(Y_iY_j) \le \sum_{i,j=1}^{k} \big|\mathrm{Cov}\,\big(Y_0, Y_{|i-j|}\big)\big|$$
$$\le C\sum_{i,j=1}^{k}(a_{\lfloor\frac{|i-j|}{3}\rfloor} + \beta_{\lfloor\frac{|i-j|}{3}\rfloor}) \le C\sum_{i=1}^{k}\sum_{k=0}^{\infty}(a_{\lfloor\frac{k}{3}\rfloor} + \beta_{\lfloor\frac{k}{3}\rfloor}) \le Ck,$$

by (33) and (6). Since $|Y_j| \le K$, then

$$\mathrm{E}\left(|Y_{k+1}|\Big|\sum_{j=1}^{k} Y_j\Big|\right) \le K\,\mathrm{E}\left(\Big|\sum_{j=1}^{k} Y_j\Big|\right) \le K\Big(\mathrm{E}\left(\sum_{i=1}^{k} Y_i\right)^2\Big)^{1/2} \le C\sqrt{k}.$$

Using these bounds in (32) together with

$$\frac{1}{(k+1)^2} \le \frac{1}{k^2}, \qquad \Big|\frac{1}{(k+1)^2} - \frac{1}{k^2}\Big| \le \frac{1+2k}{(k+1)^2k^2} \le \frac{4}{k^3},$$

we obtain (31):

$$\epsilon^2\,\mathrm{P}\left(\max_{m\le k\le n}\frac{1}{k}\Big|\sum_{i=1}^{k} Y_i\Big| > \epsilon\right) \le C\left[\frac{1}{m} + \sum_{k=m}^{n}\frac{1}{k^2} + \sum_{k=m}^{n}\frac{1}{k^{3/2}}\right] \le \frac{C}{\sqrt{m}}.$$

□

The next lemma establishes asymptotic bounds of the sums

$$S_k^{(1)} = \sum_{i=1}^{k} h_{1,n}(Y_i), \qquad S_k^{(2)} = \sum_{j=1}^{k} h_{2,n}(Y_j). \tag{34}$$

Lemma 5.2. *Assume that (Y_j) is a stationary zero mean L_1 near epoch dependent process on some absolutely regular process (Z_j) and (6) holds. Furthermore, let Assumption 2.1 be satisfied and $S_k^{(i)}$, $i = 1, 2$, be as in (34). Then*

$$\max_{1\le k\le n} n^{-1/2}\Big|S_k^{(i)}\Big| = o_P(1), \qquad i = 1, 2. \tag{35}$$

Proof. To show (35) for $i = 1$, we will use the inequality given in Theorem 6.2. Define $S_k = \sum_{i=1}^{k} n^{-1/2}h_{1,n}(Y_i)$, $k \ge 1$, and set $S_0 = 0$. We need to evaluate $\mathrm{E}(S_l - S_k)^4$ for $1 \le k < l \le n$. Note that

$$\mathrm{E}\,(S_l - S_k)^4 = n^{-2}\,\mathrm{E}\left|\sum_{i=k+1}^{l} h_{1,n}(Y_i)\right|^4 = n^{-2}\,\mathrm{E}\left|\sum_{i=1}^{l-k} h_{1,n}(Y_i)\right|^4,$$

where the last equality holds because $(h_{1,n}(Y_j))$ is a stationary process. Since $(h_{1,n}(Y_j))$ is L_1 NED on an absolutely regular process, see Section 4.3, $\mathrm{E}\, h_{1,n}(Y_0) = 0$ and $|h_{1,n}(x)| \leq C\Delta_n$ by (18), then by Lemma 6.1 and the comment below

$$\mathrm{E}\left|\sum_{i=1}^{l-k} h_{1,n}(Y_i)\right|^4 \leq C(l-k)^2\Delta_n^2,$$

where C does not depend on l, k or n. Thus,

$$\mathrm{P}(|S_l - S_k| \geq \lambda) \leq \frac{1}{\lambda^4}\,\mathrm{E}\,|S_l - S_k|^4 \leq \frac{C(l-k)^2\Delta_n^2}{\lambda^4 n^2} = \frac{1}{\lambda^4}\left(\sum_{i=k+1}^{l} u_{n,i}\right)^2,$$

where $u_{n,i} = C^{1/2}\Delta_n n^{-1}$. Hence, S_j satisfies assumption (53) of Theorem 6.2 with $\beta = 4$, $\alpha = 2$. Therefore, by (54), for any fixed $\epsilon > 0$, as $n \to \infty$,

$$\mathrm{P}\left(\max_{1\leq k\leq n} n^{-1/2}\left|S_k^{(1)}\right| \geq \epsilon\right) \leq \frac{K}{\epsilon^4}\left(\sum_{i=1}^{n} u_{n,i}\right)^2 = \frac{KC\Delta_n^2}{\epsilon^4} \to 0,$$

since $\Delta_n \to 0$. The proof of (35) for $i = 2$ follows using a similar argument as in the proof for $i = 1$. □

Proposition 5.1. *Assume that (Y_j) is L_1 near epoch dependent process on some absolutely regular process (Z_j) and (6) holds. Furthermore, let Assumption 2.1 be satisfied. Then*

$$\max_{1\leq k\leq k^*} n^{-3/2}\left|U_n(k,k^*) - k\,(n-k^*)\,\Theta_{\Delta_n}\right| = o_P\,(1) \tag{36}$$

and

$$\max_{k^*\leq k\leq n} n^{-3/2}\left|U_n(k^*,k) - k^*\,(n-k)\,\Theta_{\Delta_n}\right| = o_P\,(1)\,, \tag{37}$$

where Θ_{Δ_n} is the same as in (17).

Proof. By the Hoeffding decomposition (16),

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

Hence,

$$\begin{aligned} U_n(k,k^*) - k(n-k^*)\Theta_{\Delta_n} &= \sum_{i=1}^{k}\sum_{j=k^*+1}^{n}\big(h_{1,n}\,(Y_i) + h_{2,n}\,(Y_j) + g_n\,(Y_i,Y_j)\big) \\ &= (n-k^*)\sum_{i=1}^{k} h_{1,n}\,(Y_i) + k\sum_{j=k^*+1}^{n} h_{2,n}\,(Y_j) + \sum_{i=1}^{k}\sum_{j=k^*+1}^{n} g_n\,(Y_i,Y_j)\,. \end{aligned}$$

Denote

$$U_n^{(g)}(k,k^*) = \sum_{i=1}^{k}\sum_{j=k^*+1}^{n} g_n\left(Y_i,Y_j\right), \qquad U_n^{(g)}(k^*,k) = \sum_{i=1}^{k^*}\sum_{j=k+1}^{n} g_n\left(Y_i,Y_j\right). \tag{38}$$

Since $|n-k^*| \le n$, $k^* \le n$ and $\sum_{i=k^*+1}^{n} h_{2,n}(Y_j) = S_n^{(2)} - S_{k^*}^{(2)}$, then

$$\begin{aligned}
\Big|U_n(k,k^*) - k(n-k^*)\Theta_{\Delta_n}\Big| &\le n\Big(\big|S_k^{(1)}\big| + \big|S_n^{(2)}\big| + \big|S_{k^*}^{(2)}\big|\Big) + \Big|U_n^{(g)}(k,k^*)\Big|,\\
\Big|U_n(k^*,k) - k^*(n-k)\Theta_{\Delta_n}\Big| &\le n\Big(\big|S_{k^*}^{(1)}\big| + \big|S_n^{(2)}\big| + \big|S_k^{(2)}\big|\Big) + \Big|U_n^{(g)}(k^*,k)\Big|,
\end{aligned}$$

where $S_k^{(i)}$, $i = 1,2$ are defined in (34). Therefore,

$$\begin{aligned}
&\max_{1\le k\le k^*} n^{-3/2}\Big|U_n(k,k^*) - k(n-k^*)\Theta_{\Delta_n}\Big| \\
&\qquad \le \max_{1\le k\le n} n^{-1/2}\Big(\big|S_k^{(1)}\big| + \big|S_n^{(2)}\big| + \big|S_{k^*}^{(2)}\big|\Big) + \max_{1\le k\le k^*} n^{-3/2}\Big|U_n^{(g)}(k,k^*)\Big|.
\end{aligned}\tag{39}$$

The degenerate kernel g_n is bounded and 1-continuous, see Subsections 4.1 and 4.2. Thus, by Proposition 6.1 below,

$$\begin{aligned}
&\max_{1\le k\le k^*} n^{-3/2}\Big|U_n^{(g)}(k,k^*)\Big| \\
&\le \max_{1\le k\le n} n^{-3/2}\Big|\sum_{i=1}^{k}\sum_{j=k+1}^{n} g_n(Y_i,Y_j)\Big| + \max_{1\le k\le k^*} n^{-3/2}\Big|\sum_{i=1}^{k}\sum_{j=k+1}^{k^*} g_n(Y_i,Y_j)\Big| = o_P(1).
\end{aligned}\tag{40}$$

Similar argument implies

$$\max_{k^*<k\le n} n^{-3/2}\Big|U_n^{(g)}(k^*,k)\Big| = o_P(1).$$

Using in (39) the bounds (40) and (35) of Lemma 5.2 we obtain

$$\max_{1\le k\le k^*} n^{-3/2}\Big|U_n(k,k^*) - k(n-k^*)\Theta_{\Delta_n}\Big| = o_P(1)$$

which proves (36). The proof of (37) follows using similar argument. □

Denote

$$\tilde{U}_n^{(g)}(k) = \sum_{i=1}^{k}\sum_{j=k+1}^{n} g(Y_i,Y_j). \tag{41}$$

Lemma 5.3. *Assume that (Y_j) is L_1 near epoch dependent process on some absolutely regular process (Z_j) and (6) holds. Furthermore, let Assumption 2.1 be satisfied and let*

$a(n)=M/\Delta_n^2$, $M>0$, and $\tilde U_n^{(g)}(k)$, $U_n^{(g)}(k,k^*)$ and $U_n^{(g)}(k^*,k)$ are defined as in (41), (38). Then there exists $C>0$ such that for any $\epsilon>0$,

$$\mathrm{P}\left(\max_{k:|k-k^*|\ge a(n)}\Big|\frac{\tilde U_n^{(g)}(k)}{k^*-k}\Big|>\epsilon\right)\le\frac{C}{\epsilon^2}\Big(\frac{n^2}{a(n)}+\frac{1}{n}\Big), \tag{42}$$

$$\mathrm{P}\left(\max_{1\le k\le k^*-a(n)}\Big|\frac{U_n^{(g)}(k,k^*)}{k^*-k}\Big|>\epsilon\right)\le\frac{C}{\epsilon^2}\Big(\frac{n^2}{a(n)}+\frac{1}{n}\Big), \tag{43}$$

$$\mathrm{P}\left(\max_{k^*+a(n)\le k\le n}\Big|\frac{U_n^{(g)}(k^*,k)}{k-k^*}\Big|>\epsilon\right)\le\frac{C}{\epsilon^2}\Big(\frac{n^2}{a(n)}+\frac{1}{n}\Big),$$

where C does not depend on ϵ, n and $a(n)$.

Proof. Recall $\{k:|k-k^*|\ge a(n)\}=\{k\le k^*-a(n)\}\cup\{k\ge k^*+a(n)\}$. We consider only the case $\max_{1\le k\le k^*-a(n)}$ since the proof for $\max_{k^*+a(n)\le k\le n}$ is similar.
Proof of (42). Define $R_k=\tilde U_n^{(g)}(k)-\tilde U_n^{(g)}(k-1)$, $k\ge1$, $\tilde U_n^{(g)}(0)=0$ and $R_0=0$. Then $\tilde U_n^{(g)}(k)=\sum_{i=1}^k R_i$. Inequality (55) of Theorem 6.3, applied to the random variables R_i with $c_k=1/(k^*-k)$ yields

$$\begin{aligned}\rho_n &:= \epsilon^2\,\mathrm{P}\left(\max_{1\le k\le k^*-a(n)}\frac{1}{k^*-k}\Big|\sum_{i=1}^k R_i\Big|>\epsilon\right)\\ &\le \frac{1}{(k^*-1)^2}\,\mathrm{E}\,R_1^2+\sum_{k=1}^{k^*-a(n)}\Big|\frac{1}{(k^*-k-1)^2}-\frac{1}{(k^*-k)^2}\Big|\,\mathrm{E}\left(\sum_{i=1}^k R_i\right)^2\\ &+2\sum_{k=1}^{k^*-a(n)}\frac{1}{(k^*-k-1)^2}\,\mathrm{E}\left(|R_{k+1}|\Big|\sum_{j=1}^k R_j\Big|\right)+\sum_{k=1}^{k^*-a(n)}\frac{1}{(k^*-k-1)^2}\,\mathrm{E}\,R_{k+1}^2.\end{aligned} \tag{44}$$

In Subsections 4.1 and 4.2, we showed that kernel function $g(x,y)$ is bounded and 1-continuous. Therefore, by Lemma 6.2 below

$$\mathrm{E}\left[(\tilde U_n^{(g)}(k))^2\right]=\mathrm{E}\left(\sum_{i=1}^k R_i\right)^2\le Ck(n-k),\qquad k=1,\dots,n. \tag{45}$$

Lemma 6.2 also yields

$$\mathrm{E}\,R_{k+1}^2=\mathrm{E}\left(\tilde U_n^{(g)}(k+1)-\tilde U_n^{(g)}(k)\right)^2\le n^3C\frac{(k+1)-k}{n^2}=Cn,\qquad k=1,\dots,n. \tag{46}$$

Then,

$$\mathrm{E}\left(|R_{k+1}|\Big|\sum_{j=1}^k R_j\Big|\right)\le\left(\mathrm{E}\,R_{k+1}^2\right)^{1/2}\left(\mathrm{E}\Big(\sum_{j=1}^k R_j\Big)^2\right)^{1/2}\le C\sqrt{n}\sqrt{k(n-k)}.$$

From (44), (45) and (46), using $\frac{1}{(k^*-k-1)^2} - \frac{1}{(k^*-k)^2} \leq \frac{2}{(k^*-k-1)^3}$, we obtain

$$\rho_n \leq C\Big[\frac{n-1}{(k^*-1)^2} + \sum_{k=1}^{k^*-a(n)} \Big\{\frac{k(n-k)}{(k^*-k-1)^3} + \frac{\sqrt{n}\sqrt{k(n-k)}+n}{(k^*-k-1)^2}\Big\}\Big].$$

Noting that $\sqrt{k(n-k)} \leq n$, $(k^*-k-1)^{-3} \leq (k^*-k-1)^{-2}$, it follows

$$\rho_n \leq C\Big(\frac{1}{n} + \sum_{k=1}^{k^*-a(n)} \frac{n^2}{(k^*-k-1)^2}\Big) \leq C\Big(\frac{1}{n} + \frac{n^2}{a(n)}\Big).$$

Proof of (43). It follows a similar line to the proof of (42). Denote $\tilde{R}_k = U_n^{(g)}(k, k^*) - U_n^{(g)}(k-1, k^*)$. We verified in Subsections 4.1 and 4.2 that function $g_n(x, y)$ is bounded and 1-continuous. Therefore, by Lemma 6.2 below,

$$\mathrm{E}\left[(U_n^{(g)}(k, k^*))^2\right] = \mathrm{E}\left(\sum_{i=1}^{k} \tilde{R}_i\right)^2 \leq Ck(n-k^*), \qquad k = 1, \ldots, k^*$$

and

$$\begin{aligned}\mathrm{E}\,\tilde{R}_{k+1}^2 &= \mathrm{E}\left(U_n^{(g)}(k+1, k^*) - U_n^{(g)}(k, k^*)\right)^2 \\ &= \mathrm{E}\Big(\sum_{j=k^*+1}^{n} g_n(Y_{k+1}, Y_j)\Big)^2 \leq C(n-k^*), \qquad k = 1, \ldots, k^*.\end{aligned}$$

Combining both bounds, we obtain

$$\mathrm{E}\left(\left|\tilde{R}_{k+1}\right|\left|\sum_{j=1}^{k} \tilde{R}_j\right|\right) \leq \left(\mathrm{E}\,\tilde{R}_{k+1}^2\right)^{1/2}\left(\mathrm{E}\Big(\sum_{j=1}^{k} \tilde{R}_j\Big)^2\right)^{1/2} \leq C(n-k^*)\sqrt{k}.$$

Using the same argument as in the proof of (42), we obtain

$$\mathrm{P}\left(\max_{1\leq k\leq k^*-a(n)} \frac{1}{k^*-k}\left|U_n^{(g)}(k, k^*)\right| > \epsilon\right) \leq \frac{C}{\epsilon^2}\Big(\frac{1}{n} + \frac{n^2}{a(n)}\Big).$$

This completes proof of (43) and the lemma. □

Lemma 5.4. *Assume that (Y_j) is a stationary zero mean L_1 near epoch dependent process on some absolutely regular process (Z_j) and (6) holds. Furthermore, let Assumption 2.1 be satisfied and let $a(n) = M/\Delta_n^2$, $M > 0$. Then, as $n \to \infty$, $M \to \infty$,*

(i) For any $\epsilon > 0$,

$$\mathrm{P}\left(\frac{1}{n\Delta_n} \max_{k:|k-k^*|\geq a(n)} \left|\frac{U_n(k^*) - U_n(k)}{k^*-k}\right| > \epsilon\right) \to 0. \tag{47}$$

(ii) For any $\epsilon > 0$,

$$\mathrm{P}\left(\frac{1}{n\Delta_n}\max_{1\le k\le k^*-a(n)}\left|\frac{U_n(k^*,k^*)-U_n(k,k^*)}{k^*-k}-(n-k^*)\Theta_{\Delta_n}\right|>\epsilon\right)\to 0, \tag{48}$$

and

$$\mathrm{P}\left(\frac{1}{n\Delta_n}\max_{k^*+a(n)\le k\le n}\left|\frac{U_n(k^*,k^*)-U_n(k^*,k)}{k-k^*}-k^*\Theta_{\Delta_n}\right|>\epsilon\right)\to 0,$$

where Θ_{Δ_n} is the same as in (17).

Proof. Notice that $\{k : |k-k^*| \ge a(n)\} = \{k \le k^*-a(n)\} \cup \{k \ge k^*+a(n)\}$. We will prove relations (47) and (48) for $\max_{1\le k\le k^*-a(n)}$. The proof for $\max_{k^*+a(n)\le k\le n}$ is similar.
Notice, that $a(n) = \frac{Mn}{\Delta_n^2 n} = o(Mn)$ since $n\Delta_n^2 \to \infty$ by assumption (5). Therefore, for a fixed M, $a(n) = o(k^*)$ and $k^*-a(n) > 1$ as $n \to \infty$.
(i) Denote

$$\tilde{S}_k = \sum_{i=1}^{k} h_1(Y_i).$$

By Hoeffding's decomposition (15), for $k \le k^*$, and using $h_1(x) = -h_2(x)$, it follows

$$\begin{aligned} U_n(k) &= \sum_{i=1}^{k}\sum_{j=k+1}^{n}\big(h_1(Y_i)+h_2(Y_j)+g(Y_i,Y_j)\big) \\ &= (n-k)\sum_{i=1}^{k}h_1(Y_i) - k\sum_{j=k+1}^{n}h_1(Y_j) + \sum_{i=1}^{k}\sum_{j=k+1}^{n}g(Y_i,Y_j) = n\tilde{S}_k - k\tilde{S}_n + \tilde{U}_n^{(g)}(k), \end{aligned}$$

where $\tilde{U}_n^{(g)}(k)$ is defined in (41). Hence,

$$\Big|U_n(k^*)-U_n(k)\Big| = \Big|n\big(\tilde{S}_{k^*}-\tilde{S}_k\big) - (k^*-k)\tilde{S}_n + \tilde{U}_n^{(g)}(k^*) - \tilde{U}_n^{(g)}(k)\Big|.$$

Therefore, for $1 \le k \le k^*-a(n)$,

$$\begin{aligned} &\frac{1}{n\Delta_n}\left|\frac{U_n(k^*)-U_n(k)}{k^*-k}\right| \\ &\qquad \le \frac{1}{\Delta_n}\frac{\big|\tilde{S}_{k^*}-\tilde{S}_k\big|}{k^*-k} + \frac{1}{\Delta_n}\frac{\big|\tilde{S}_n\big|}{n} + \frac{\big|\tilde{U}_n^{(g)}(k^*)\big|}{n\Delta_n a(n)} + \frac{1}{n\Delta_n}\frac{\big|\tilde{U}_n^{(g)}(k)\big|}{k^*-k} \\ &\qquad =: \rho_k^{(1)} + \rho_k^{(2)} + \rho_k^{(3)} + \rho_k^{(4)}. \end{aligned}$$

It suffices to show that for any $\epsilon > 0$, as $n \to \infty$, for $l = 1, \dots, 4$,

$$\mathrm{P}\left(\max_{1\le k\le k^*-a(n)}\rho_k^{(l)} > \epsilon\right) \to 0, \quad M \to \infty, \tag{49}$$

which proves (47) for $\max_{1\le k\le k^*-a(n)}$.

For $l = 1$, stationarity of the process $(h_1(Y_j))$ yields

$$\Big\{\big|\tilde{S}_{k^*} - \tilde{S}_k\big| = \big|\sum_{i=k+1}^{k^*} h_1(Y_i)\big|,\ 1 \le k \le k^* - a(n)\Big\} \stackrel{d}{=} \Big\{\big|\tilde{S}_{k^*-k}\big|,\ 1 \le k \le k^* - a(n)\Big\}.$$

Therefore,

$$\max_{1\le k\le k^*-a(n)} \rho_k^{(1)} \stackrel{d}{=} \frac{1}{\Delta_n} \max_{k\ge 1:\, k^*-k\ge a(n)} \frac{\big|\tilde{S}_{k^*-k}\big|}{k^*-k} \stackrel{d}{=} \frac{1}{\Delta_n} \max_{a(n)\le j\le n} \frac{\big|\tilde{S}_j\big|}{j}.$$

Since $(h_1(Y_j))$ is L_1 NED on an absolutely regular process (Z_j), $\mathrm{E}\, h_1(Y_1) = 0$ and $|h_1(x)| \le 1/2$, then by Lemma 5.1,

$$\max_{a(n)\le k\le n} \frac{1}{k}\big|\tilde{S}_k\big| = O_P\Big(\frac{1}{\sqrt{a(n)}}\Big). \tag{50}$$

Thus,

$$\max_{1\le k\le k^*-a(n)} \rho_k^{(1)} = O_P\Big(\frac{1}{\Delta_n\sqrt{a(n)}}\Big) = O_P\Big(\frac{1}{\sqrt{M}}\Big) = o_P(1), \quad \text{as } M \to \infty,$$

which proves (49) for $l = 1$.
For $l = 2$, by (50), $|\tilde{S}_n|/n = O_P(n^{-1/2})$. Thus,

$$\rho_k^{(2)} = \frac{1}{\Delta_n}\frac{\tilde{S}_n}{n} = O_P\Big(\frac{1}{\Delta_n\sqrt{n}}\Big) = o_P(1),$$

since $\Delta_n\sqrt{n} \to \infty$ by (5), which proves (49) for $l = 2$.
To show (49) for $l = 3$, recall that $g(x, y) \le 3/2$ is 1-continuous, see Subsection 4.3. Therefore, by Lemma 6.2,

$$\mathrm{E}\Big(\frac{\tilde{U}_n^{(g)}(k^*)}{\sqrt{k^*(n-k^*)}}\Big)^2 \le C,$$

which implies that

$$\frac{\tilde{U}_n^{(g)}(k^*)}{\sqrt{k^*(n-k^*)}} = O_P(1).$$

Thus,

$$\begin{aligned}\rho_k^{(3)} = \frac{\big|\tilde{U}_n^{(g)}(k^*)\big|}{n\Delta_n a(n)} &= O_P\Big(\frac{\sqrt{k^*(n-k^*)}}{n\Delta_n a(n)}\Big)\\ &= O_P\Big(\frac{1}{\Delta_n a(n)}\Big) = O_P\Big(\frac{\Delta_n}{M}\Big) = O_P\Big(\frac{1}{M}\Big) = o_P(1), \quad \text{as } M \to \infty,\end{aligned}$$

which proves (49) for $l = 3$.

Finally, for $l = 4$, by Lemma 5.3,

$$\mathrm{P}\Big(\max_{1\le k\le k^*-a(n)} \rho_k^{(4)} > \epsilon\Big) = \mathrm{P}\Big(\max_{1\le k\le k^*-a(n)} \frac{\tilde{U}_n^{(g)}(k)}{k^*-k} > \epsilon n\Delta_n\Big)$$
$$\le \frac{C}{(\epsilon n\Delta_n)^2}\Big(\frac{n^2}{a(n)} + \frac{1}{n}\Big) = \frac{C}{\epsilon^2}\Big(\frac{1}{M} + \frac{1}{n^3\Delta_n^2}\Big) \to 0, \quad \text{as } M\to\infty,$$

which proves (49) for $l = 4$ and completes the proof of *(i)*.

(ii) Let $S_k^{(1)}$, $S_k^{(2)}$ and $U_n^{(g)}(k,k^*)$ be defined as in (34) and (38). By Hoeffding's decomposition (16), for $k \le k^*$,

$$U_n(k,k^*) - k(n-k^*)\Theta_{\Delta_n} = \sum_{i=1}^{k}\sum_{j=k^*+1}^{n}\big(h_{1,n}(Y_i) + h_{2,n}(Y_j) + g_n(Y_i,Y_j)\big)$$
$$= (n-k^*)\sum_{i=1}^{k} h_{1,n}(Y_i) + k\sum_{j=k^*+1}^{n} h_{2,n}(Y_j) + \sum_{i=1}^{k}\sum_{j=k^*+1}^{n} g_n(Y_i,Y_j)$$
$$= (n-k^*)S_k^{(1)} + k(S_n^{(2)} - S_{k^*}^{(2)}) + U_n^{(g)}(k,k^*).$$

Hence,

$$\big|U_n(k^*,k^*) - U_n(k,k^*) - (k^*-k)(n-k^*)\Theta_{\Delta_n}\big|$$
$$= \big|(n-k^*)(S_{k^*}^{(1)} - S_k^{(1)}) + (k^*-k)(S_n^{(2)} - S_{k^*}^{(2)}) + U_n^{(g)}(k^*,k^*) - U_n^{(g)}(k,k^*)\big|.$$

Therefore, for $1 \le k \le k^* - a(n)$,

$$\frac{1}{n\Delta_n}\Big|\frac{U_n(k^*,k^*) - U_n(k,k^*)}{k^*-k} - (n-k^*)\Theta_{\Delta_n}\Big|$$
$$\le \frac{1}{\Delta_n}\frac{\big|S_{k^*}^{(1)} - S_k^{(1)}\big|}{k^*-k} + \frac{1}{\Delta_n}\frac{\big|S_n^{(2)} - S_{k^*}^{(2)}\big|}{n} + \frac{\big|U_n^{(g)}(k^*,k^*)\big|}{n\Delta_n a(n)} + \frac{1}{n\Delta_n}\frac{\big|U_n^{(g)}(k,k^*)\big|}{k^*-k}$$
$$=: \nu_k^{(1)} + \nu_k^{(2)} + \nu_k^{(3)} + \nu_k^{(4)}.$$

It suffices to show that for any $\epsilon > 0$, as $n \to \infty$, for $l = 1, \dots, 4$,

$$\mathrm{P}\Big(\max_{1\le k\le k^*-a(n)} \nu_k^{(l)} > \epsilon\Big) \to 0, \quad M\to\infty, \tag{51}$$

which proves (48) for $\max_{1\le k\le k^*-a(n)}$. The process $(h_{1,n}(Y_j))$ is stationary and L_1 NED on an absolutely regular process, see Section 4.3. Furthermore, it has zero mean and $|h_{1,n}| \le C\Delta_n$ by (18). Hence, by the same argument as for $\rho_k^{(1)}$, using Lemma 5.1, it follows

$$\max_{1\le k\le k^*-a(n)} \nu_k^{(1)} \overset{d}{=} \frac{1}{\Delta_n}\max_{a(n)\le j\le n}\frac{|S_j^{(1)}|}{j} = O_P\Big(\frac{1}{\Delta_n\sqrt{a(n)}}\Big) = O_P\Big(\frac{1}{\sqrt{M}}\Big) = o_P(1),$$

as $M \to \infty$.
Lemma 5.2 yields $\max_{1\le k\le n} n^{-1/2}|S_k^{(2)}| = o_P(1)$. Therefore,

$$\max_{1\le k\le k^*-a(n)} \nu_k^{(2)} \le \frac{1}{\sqrt{n}\Delta_n}\big(n^{-1/2}|S_n^{(2)}| + n^{-1/2}|S_{k^*}^{(2)}|\big) = o_P(1),$$

since $\sqrt{n}\Delta_n \to \infty$.
We showed in Subsections 4.1 and 4.2 that the function $g_n(x,y)$ is bounded and 1-continuous. Hence, by Lemma 6.2,

$$\mathrm{E}\left(\frac{U_n^{(g)}(k^*,k^*)}{\sqrt{k^*(n-k^*)}}\right)^2 \le C.$$

Therefore, the claim $\max_{1\le k\le k^*-a(n)} \nu_k^{(3)} = o_P(1)$ follows using the same argument as in the proof of (49) for $l=3$.
By Lemma 5.3,

$$\begin{aligned}\mathrm{P}\Big(\max_{1\le k\le k^*-a(n)} \nu_k^{(4)} > \epsilon\Big) &= \mathrm{P}\Big(\max_{1\le k\le k^*-a(n)} \frac{|U_n^{(g)}(k,k^*)|}{k^*-k} > \epsilon n\Delta_n\Big)\\ &\le \frac{C}{(\epsilon n\Delta_n)^2}\Big(\frac{n^2}{a(n)}+\frac{1}{n}\Big) = \frac{C}{\epsilon^2}\Big(\frac{1}{M}+\frac{1}{n^3\Delta_n^2}\Big) \to 0, \quad \text{as } M\to\infty,\end{aligned}$$

which proves (51) for $l=4$. This completes the proof of (48) and the lemma.

□

6 Auxiliary results from the literature

This section contains results from the literature used in the proofs of this paper.
Lemma 6.1 states a correlation and a moment inequality for L_1 NED random variables, established by Borovkova *et al.* (2001).

Lemma 6.1. *(Lemma 2.18 and 2.24, Borovkova et al. (2001)) Let (Y_j) be L_1 near epoch dependent on an absolutely regular, stationary process with mixing coefficients β_k and approximation constants a_k, and such that $|Y_0| \le K \le \infty$ a.s. Then, for all $i,k\ge 0$,*

$$|\mathrm{Cov}(Y_i, Y_{i+k})| \le 4Ka_{\lfloor \frac{k}{3}\rfloor} + 2K^2\beta_{\lfloor \frac{k}{3}\rfloor}.$$

In addition, if $\sum_{k=0}^{\infty} k^2(a_k+\beta_k) < \infty$, then there exists $C>0$ such that for all $n\ge 1$

$$\mathrm{E}\left(\sum_{i=1}^{n}\big(Y_i - \mathrm{E}\,Y_i\big)\right)^4 \le Cn^2. \tag{52}$$

The proof of Lemma 2.24 in Borovkova *et al.* (2001) shows that (52) holds with $C = C_0K^2$, where $C_0>0$ does not depend on K and n.

In Theorem 3 of Dehling *et al.* (2015) the asymptotic distribution of the Wilcoxon test statistic for L_1 NED random process is obtained. We use this result to show the consistency of the Wilcoxon-type estimator $\hat{k}$.

Theorem 6.1. *(Theorem 3, Dehling et al. (2015)) Assume that* (Y_j) *is stationary and* L_1 *near epoch dependent process on some absolutely regular process* (Z_j) *and (6) holds. Then,*

$$\frac{1}{n^{3/2}} \max_{1\le k<n} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left(1_{\{Y_i \le Y_j\}} - 1/2\right) \right| \xrightarrow{d} \sigma \sup_{0\le\tau\le1} |B(\tau)|,$$

where $(B(\tau))_{0\le\tau\le1}$ *is the standard Brownian bridge process,*

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

and F *denotes the distribution function of* Y_j*.*

We use the following results from Dehling *et al.* (2015) to handle the degenerate part $g(x, y)$ of the Hoeffding decomposition (15).

Proposition 6.1. *(Proposition 1, Dehling et al. (2015)) Let* (Y_j) *be stationary and* L_1 *near epoch dependent on an absolutely regular process with mixing coefficients* β_k *and approximation constants* a_k *satisfying*

$$\sum_{k=1}^{\infty} k\left(\beta_k + \sqrt{a_k} + \phi(a_k)\right) < \infty,$$

with $\phi(\epsilon)$ *as in Definition* 4.1*. If* $g(x, y)$ *is a 1-continuous bounded degenerate kernel, then, as* $n \to \infty$*,*

$$\frac{1}{n^{3/2}} \max_{1\le k\le n} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} g(Y_i, Y_j) \right| \to_p 0.$$

Lemma 6.2. *(Lemma 1 and 2, Dehling et al. (2015)) Under assumptions of Proposition* 6.1 *there exists* $C > 0$ *such that for all* $1 \le m \le k \le n$*,* $n \ge 2$*,*

$$\mathrm{E}\left(\sum_{i=1}^{k} \sum_{j=k+1}^{n} g(Y_i, Y_j) \right)^2 \le Ck(n-k),$$

$$\mathrm{E}\left(n^{-3} \left| \sum_{i=1}^{k} \sum_{j=k+1}^{n} g(Y_i, Y_j) - \sum_{i=1}^{m} \sum_{j=m+1}^{n} g(Y_i, Y_j) \right|^2 \right) \le C\frac{k-m}{n^2}.$$

In our proofs we use the maximal inequality of Billingsley (1999), which is valid for stationary/non-stationary and independent/dependent random variables ξ_i.

Theorem 6.2. *(Theorem 10.2, Billingsley (1999)) Let* $\xi_1, \dots, \xi_n$ *be random variables and* $S_k = \sum_{i=1}^{k} \xi_k$*,* $k \ge 1$*,* $S_0 = 0$ *denotes the partial sum. Suppose that there exist* $\alpha > 1$*,* $\beta > 0$ *and non-negative numbers* $u_{n,1}, \dots, u_{n,n}$ *such that*

$$\mathrm{P}\left(|S_j - S_i| \ge \lambda \right) \le \frac{1}{\lambda^\beta} \left(\sum_{l=i+1}^{j} u_{n,l} \right)^\alpha, \tag{53}$$

for $\lambda > 0$, $0 \le i \le j \le n$. Then for all $\lambda > 0$, $n \ge 2$,

$$\mathrm{P}\left(\max_{1\le k\le n} |S_k| \ge \lambda\right) \le \frac{K}{\lambda^\beta}\left(\sum_{l=1}^{n} u_{n,l}\right)^\alpha, \tag{54}$$

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

By the Markov inequality, (53) is satisfied if

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

In the proof of Lemma 5.1 we use a Hájek-Rényi type inequality established by Kokoszka and Leipus (2000).

Theorem 6.3. *(Theorem 4.1, Kokoszka and Leipus (2000)) Let $X_1, ..., X_n$ be any random variables with finite second moments and $c_1, ..., c_n$ be any non-negative constants. Then*

$$\begin{aligned}\epsilon^2\,\mathrm{P}\left(\max_{m\le k\le n} c_k\left|\sum_{i=1}^{k} X_i\right| > \epsilon\right) \le c_m^2 \sum_{i,j=1}^{m} \mathrm{E}\,(X_iX_j) + \sum_{k=m}^{n-1} \left|c_{k+1}^2 - c_k^2\right| \sum_{i,j=1}^{k} \mathrm{E}\,(X_iX_j)\\ + 2\sum_{k=m}^{n-1} c_{k+1}^2\,\mathrm{E}\left(|X_{k+1}|\left|\sum_{j=1}^{k} X_j\right|\right) + \sum_{k=m}^{n-1} c_{k+1}^2\,\mathrm{E}\,X_{k+1}^2.\end{aligned} \tag{55}$$

Acknowledgement

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

References

Antoch, J., Hušková, M. and Veraverbeke, N. (1995). Change-point problem and bootstrap. *J. Nonparametr. Stat.* **5**, 123-144.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *J. Time Series Anal.* **15**, 453-472.

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

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

Bradley, R.C. (2002). *Introduction to Strong Mixing Conditions.* Kendrick Press, Heber City.

Csörgö, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* Wiley, New York.

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

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

Giraitis, L., Leipus, R. and Surgailis, D. (1996). The change-point problem for dependent observations. *J. Statist. Plann. Inference* **53**, 297-310.

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

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

Horváth, L. and Kokoszka, P. (1997). The effect of long-range dependence on change-point estimators. *J. Statist. Plann. Inference* **64**, 57-81.

Kokoszka, P. and Leipus, R. (1998). Change-point in the mean of dependent observations. *Statist. Probab. Lett.* **40**, 385-393.

Kokoszka, P. and Leipus, R. (2000). Change-point estimation in ARCH models. *Bernoulli* **6**, 513-539.

Ling, S. (2007). Testing for change points in time series models and limiting theorems for NED sequences. *Ann. Statist.* **35**, 1213-1237.

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