SFB 823

A robust method for shift detection in time series

Discussion Paper

Herold Dehling, Roland Fried, Martin Wendler

Nr. 16/2015

SFB 823

June 9, 2015

A ROBUST METHOD FOR SHIFT DETECTION IN TIME SERIES

HEROLD DEHLING, ROLAND FRIED, AND MARTIN WENDLER

Abstract. We present a robust test for change-points in time series which is based on the two-sample Hodges-Lehmann estimator. We develop new limit theory for a class of statistics based on the two-sample U-quantile processes, in the case of short range dependent observations. Using this theory we can derive the asymptotic distribution of our test statistic under the null hypothesis. We study the finite sample properties of our test via a simulation study and compare the test with the classical CUSUM test and a test based on the Wilcoxon-Mann-Whitney statistic.

Contents

1. Introduction

Statistical tests for the presence of changes in the structure of a time series are of great importance in a wide range of scientific discussions, e.g. regarding economic, technological and climate data. Many procedures for detecting changes and for estimating change points have been proposed in the statistical literature, see e.g. Csörgő and Horvath (1997) for a detailed exposition. In the case of independent observations with normal tails, the theory is quite satisfactory. For a wide variety of change-point models, many statistical procedures have been proposed and their properties have been investigated. In contrast, the situation is quite different for dependent data, such as encountered in time series models, and for heavy-tailed data. For dependent data, most research has focused on linear procedures, such as CUSUM tests, and there are many open problems when it comes to other types of test procedures, e.g. those based on robust statistics.

In the present paper, we study tests for detecting a level shift in a time series. Specifically, we assume that the sequence of observations $(X_n)_{n\geq 1}$ is generated by the model

$$X_n = \mu_n + Y_n,$$

Key words and phrases. Change-point tests, shift detection, Hodges-Lehmann estimator, time series, weakly dependent data, two-sample U-statistics, two-sample U-process, two-sample U-quantiles, functional central limit theorem.

where $(\mu_n)_{n\geq 1}$ is a sequence of unknown constants and $(Y_n)_{n\geq 1}$ is a stationary process with mean zero. We will focus on the case when $(Y_n)_{n\geq 1}$ is a weakly dependent process, in a sense that we will specify below. As examples, we will be able to treat most standard models of time series analysis, such as ARMA processes and GARCH processes, but our theory is not restricted to such concrete models.

Given observations $X_1, \ldots, X_n$, we want to test the null hypothesis that the process is stationary, i.e.

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

against the alternative that there is a level shift at some unknown point in time, i.e.

$$A: \text{ there exists } k \in \{1, \ldots, n-1\} \text{ such that } \mu_1 = \ldots = \mu_k \neq \mu_{k+1} = \ldots = \mu_n.$$

Note that in case the change-point k is known in advance, the test problem becomes a standard two-sample problem, where the first sample is $X_1, \ldots, X_k$ and the second sample is $X_{k+1}, \ldots, X_n$. This test problem is obviously much simpler than the change-point problem studied here. At the same time, tests for the two-sample problem often serve as guideline for finding tests for the problem of detecting a change at an unknown point in time.

The standard test statistic for the above change-point problem is the CUSUM statistic, which is defined as

$$\max_{k=1,\ldots,n} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{k} X_i - \frac{k}{n} \sum_{i=1}^{n} X_i \right|.$$

The asymptotic distribution of this test statistic under the null hypothesis can be derived from a functional central limit theorem for the partial sum process $(\frac{1}{\sqrt{n}} \sum_{i=1}^{[n\lambda]} Y_i)_{0\leq\lambda\leq 1}$. In the case when the noise process $(Y_i)_{i\geq 1}$ is i.i.d. with finite variance, this is Donsker's invariance principle. Similar results have been obtained for a wide range of short range dependent processes $(Y_i)_{i\geq 1}$; see e.g. Ibragimov and Linnik (1961), Bradley (2007) and Dedecker et al. (2007) for various functional central limit theorems for a large class of weakly dependent processes. In this case, the partial sum process will converge in distribution to a Brownian motion $(\sigma W(\lambda))_{0\leq\lambda\leq 1}$, where $\sigma^2 = \mathrm{Var}(Y_1) + 2\sum_{k=2}^{\infty} \mathrm{Cov}(Y_1, Y_k)$ is the long-run variance. As an application of the continuous mapping theorem, we thus find that

$$\max_{k=1,\ldots,n} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{k} X_i - \frac{k}{n} \sum_{i=1}^{n} X_i \right| \xrightarrow{d} \sigma \sup_{0\leq\lambda\leq 1} |W(\lambda) - \lambda W(1)|.$$

Hence we can calculate approximate critical values for the CUSUM test from tables of the Kolmogorov-Smirnov distribution, i.e. the distribution of the supremum of the Brownian bridge process $(W(\lambda) - \lambda W(1))_{0\leq\lambda\leq 1}$, provided we have a consistent estimator of the asymptotic variance σ^2. Such estimators have been proposed in the literature; see e.g. Dehling, Fried, Sharipov, Vogel and Wornowizki (2013) for a recent result under dependence conditions relevant for this paper.

The CUSUM test is based on partial sums and is thus not robust to outliers in the data. In this paper, we will propose a robust alternative to the CUSUM test and investigate its properties. Our test will be valid without any moment assumptions on the underlying data, and can thus be applied to arbitrarily heavy-tailed data. In order to motivate our test, we note that using some elementary algebra, we obtain the following alternative representation of the CUSUM test statistic,

$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^{k} X_i - \frac{k}{n} \sum_{i=1}^{n} X_i \right) = \sqrt{n} \frac{k}{n} \left(1 - \frac{k}{n}\right) \left(\frac{1}{k} \sum_{i=1}^{k} X_i - \frac{1}{n-k} \sum_{i=k+1}^{n} X_i \right).$$

On the right hand side we have the term $\frac{1}{k}\sum_{i=1}^k X_i - \frac{1}{n-k}\sum_{i=k+1}^n X_i$, which is the standard estimator for the difference of the expected values of the two samples $X_1,\ldots,X_k$ and $X_{k+1},\ldots,X_n$, in a Gaussian two sample model of i.i.d. data.

In our paper, we propose a test that is based on the Hodges-Lehmann two sample estimator, in the same way that the CUSUM test is based on the difference of the arithmetic means of the two samples. In a classical two-sample problem with independent samples $X_1,\ldots,X_{n_1}$ and $Y_1,\ldots,Y_{n_2}$, the Hodges-Lehmann estimator is defined as

$$\mathrm{med}\{(Y_j - X_i) : 1 \le i \le n_1, 1 \le j \le n_2\};$$

see Hodges and Lehmann (1963). The Hodges-Lehmann estimator is robust, while having a high efficiency in the case of Gaussian observations. Asymptotic normality of the Hodges-Lehmann estimator has been established under a wide range of assumptions. Recently, Fried and Dehling (2011) explored the good robustness properties of two-sample tests based on this estimator, and Dehling and Fried (2012) proved its asymptotic normality in the case of short range dependent observations.

We propose the Hodges-Lehmann change-point test statistic, which we define as

$$T_n := \sqrt{n}\max_{1\le k\le n}\frac{k}{n}\left(1-\frac{k}{n}\right)\left|\mathrm{med}\{(X_j - X_i) : 1 \le i \le k, k+1 \le j \le n\}\right|.$$

The Hodges-Lehmann change-point test will reject the null hypothesis for large positive or large negative values of T_n. In this paper, we will derive asymptotic distribution theory for a class of test statistics, of which the Hodges-Lehmann test statistic is a special example. As an application of our general results, we can determine the asymptotic distribution of the Hodges-Lehmann change-point test under very general conditions.

Theorem 1.1. *Let $(Y_i)_{i\ge1}$ be a stationary process that is a near epoch dependent functional of an absolutely regular process $(Z_n)_{n\in\mathbb{Z}}$ with mixing coefficients $(\beta(n))_{n\ge1}$ and approximating constants $(a_n)_{n\ge1}$ satisfying $\beta(n) = O(n^{-8})$ and $a_n = O(n^{-12})$. Moreover, let Y_1 have an absolutely continuous distribution with density $f(x)$ and assume that $u(x) = \int f(y)f(x+y)dy$ is $\frac{1}{2}$-Hölder continuous. Then, under the null hypothesis of no change, we obtain*

$$\sqrt{n}\max_{1\le k\le n}\frac{k}{n}(1-\frac{k}{n})\left|\mathrm{med}\{(X_j - X_i) : 1 \le i \le k, k+1 \le j \le n\}\right| \xrightarrow{\mathcal{D}} \frac{\sigma}{u(0)}\sup_{0\le\lambda\le1}\left|W^{(0)}(\lambda)\right|,$$

where

$$\sigma^2 = \sum_{k=-\infty}^{\infty}\mathrm{Cov}(F(X_0), F(X_k)),$$

and where $(W^{(0)}(\lambda))_{0\le\lambda\le1}$ denotes a standard Brownian bridge process.

In order to apply the above theorem, we have to provide consistent estimators for the nuisance parameters σ^2 and $u(0)$. We use a subsampling estimator for σ^2 that was proposed by Dehling and Fried (2012). We choose a block length $l = l_n$ in such a way that $\frac{l}{\sqrt{n}} + \frac{1}{l} = o(1)$. We then define the estimator

$$\hat{\sigma}^2 = \frac{\sqrt{\pi}}{\sqrt{2l_n}\lfloor n/l_n\rfloor}\sum_{i=1}^{\lfloor n/l\rfloor}|\hat{S}_i(l)|,$$

where $\hat{S}_i(l) = \sum_{j=(i-1)l+1}^{il}(F_n(X_i) - \frac{1}{2})$. Dehling, Fried, Sharipov, Vogel and Wornowizki (2013) have established consistency of this estimator under the same assumptions as made in the present paper. We use a kernel density estimator for $u(0)$. Observing that $u(x)$ is the density of $X - Y$, where X and Y are independent random variables with the same distribution as X_1, we use a kernel density estimator based on the pairwise differences $X_i - X_j$, $1 \le i < j \le n$. In this way we get

$$\hat{u}(0) = \frac{2}{n(n-1)b} \sum_{1 \le i < j \le n} K\left(\frac{X_i - X_j}{b}\right)$$

for a symmetric, Lipschitz-continuous kernel function K which is integrates to 1. Below, we will show that $\hat{u}(0)$ is a consistent estimator of $u(0)$ under H, provided that the bandwidth $b = b_n$ is chosen appropriately.

Corollary 1.2. *Under the same assumptions as in the above theorem, we obtain under the null hypothesis*

$$\sqrt{n}\frac{\hat{u}(0)}{\hat{\sigma}} \max_{1 \le k \le n} \frac{k}{n}\left(1 - \frac{k}{n}\right) |\mathrm{med}\{(X_j - X_i) : 1 \le i \le k, k+1 \le j \le n\}|$$

converges in distribution to $\sup_{0 \le \lambda \le 1} \left|W^{(0)}(\lambda)\right|$, *where* $W^{(0)}(\lambda)$ *denotes a standard Brownian bridge process.*

The asymptotic distribution of the Hodges-Lehmann test statistic can be derived from a study of the process $\sqrt{n}\lambda(1-\lambda)\,\mathrm{med}\{(X_j - X_i) : 1 \le i \le [n\lambda], [n\lambda]+1 \le j \le n\}$, $0 \le \lambda \le 1$. More generally, we are lead to the study of the process of quantiles of the values

$$g(X_i, X_j),\ 1 \le i \le [n\lambda],\ [n\lambda]+1 \le j \le n,$$

indexed by $0 \le \lambda \le 1$, where $g(x, y)$ is a given function of two variables. In the present paper we investigate the asymptotic distribution of this process in the case of short range dependent data.

2. Main Theoretical Results

2.1. **Near Epoch Dependent Processes.** We will derive the asymptotic results in this paper under the assumption of short range dependence. In the literature, there is a wide range of notions that formally capture the idea of short range dependent processes. The classical approach is to impose mixing conditions, such as strong mixing, absolute regularity and uniform mixing, also known as α-mixing, β-mixing and ϕ-mixing, respectively. This approach was initiated by the seminal paper of Rosenblatt (1956), where strongly mixing processes were introduced and where a central limit theorem for partial sums of strongly mixing processes was proved. For a survey of various mixing concepts and associated limit theorems for partial sums, see the monographs by Doukhan (1994) and Rio (2000), as well as the encyclopedic three volume monograph by Bradley (2007).

Mixing concepts provided a unifying structure that allowed establishing limit theorems for a wide range of stochastic processes, that previously could be treated only by ad hoc methods. Among these processes are, e.g. ARMA-processes with a continuous innovation distribution, stationary ergodic Markov Chains, the process of digits in a continued fraction expansion. At the same time there are very important processes that are not mixing. One of the simplest examples are AR(1)-processes with discrete innovations, as was pointed out by Andrews (1984). The largest class of non-mixing processes are deterministic dynamical systems, i.e. processes defined as $X_n = T(X_{n-1})$, where $T : \mathcal{X} \to \mathcal{X}$ is a map on some state space $\mathcal{X}$ and where X_0 is a random variable. Such processes do not satisfy any of the classical mixing conditions, but yet under some assumptions on the map T and the distribution of X_0, they satisfy many of the classical limit theorems. In order to overcome these obvious shortcomings of mixing concepts, Doukhan and Louhichi (1999) suggested various new notions of weakly dependent processes, which are based on covariance inequalities for suitable functions of blocks of random variables that are separated in time. In their paper and in subsequent publications of various authors, a large variety of limit theorems was established for processes satisfying these notions. For a comprehensive survey, see the monograph by Dedecker et al (2007).

In the present paper, we follow a more classical approach that has been used already by Billingsley (1968) and Ibragimov and Linnik (1971). We assume that the noise process $(Y_i)_{i\geq 1}$ is near epoch dependent (NED) on an absolutely regular process.

Definition 2.1. *(i) Let $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ be two σ-fields on the probability space $(\Omega, \mathcal{F}, P)$. We define the absolute regularity coefficient $\beta(\mathcal{A}, \mathcal{B})$ by*

$$\beta(\mathcal{A}, \mathcal{B}) = E(\sup_{A\in\mathcal{A}} |P(A|\mathcal{B}) - P(A)|)$$

(ii) For a stationary process $(Z_n)_{n\in\mathbb{Z}}$ we define the absolute regularity coefficients

$$\beta(k) = \sup_{n\geq 1} \beta(\mathcal{G}_1^n, \mathcal{G}_{n+k}^\infty),$$

where $\mathcal{G}_k^l$ denotes the σ-field generated by the random variables $Z_k, \ldots, Z_l$. The process $(Z_n)_{n\in\mathbb{Z}}$ is called absolutely regular if $\beta(k) \to 0$ as $k \to \infty$.
(iii) Let $((X_n, Z_n))_{n\in\mathbb{Z}}$ be a stationary process. We say that $(X_n)_{n\geq 0}$ is L^1-near epoch dependent on the process $(Z_n)_{n\in\mathbb{Z}}$ with approximating constants $(a_l)_{l\geq 1}$, if

$$E|X_0 - E(X_0|\mathcal{G}_{-l}^l)| \leq a_l,$$

and $\lim_{l\to\infty} a_l = 0$.

If the process $(X_n)_{n\geq 0}$ is near epoch dependent on the process $(Z_n)_{n\in\mathbb{Z}}$, we get by definition the representation $X_0 = f((Z_n)_{n\in\mathbb{Z}})$ for some measurable function $f : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$. By stationarity, we thus obtain a representation

$$X_k = f((Z_{n+k})_{n\in\mathbb{Z}}).$$

Thus, a process that is NED on an absolutely regular process is also called a functional of an absolutely regular process. The class of processes that are NED on an absolutely regular process contains all relevant processes from time series analysis as well as many dynamical systems; see e.g. Borovkova, Burton and Dehling (2001) for a detailed list of examples.

2.2. **Two-sample empirical U-quantile process.** In this paper, we will investigate the two-sample empirical quantile process associated with the kernel $g(x, y)$. We will now formally define this process, as well as the related two-sample empirical U-process, both in a slightly more general setup of empirical processes indexed by classes of functions.

Definition 2.2. *Let $h : \mathbb{R}^2 \times \mathbb{R} \to [0, 1]$ be a measurable function, and let $(X_i)_{i\geq 1}$ be a stochastic process.*
(i) We define the two-sample empirical U-process

$$U_n(\lambda, t) = \frac{1}{[n\lambda](n - [n\lambda])} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} h(X_i, X_j, t),\ 0 \leq \lambda \leq 1, t \in \mathbb{R}.$$

(ii) Given $p \in [0, 1]$, we define the two-sample empirical U-quantile process

$$Q_n(\lambda, p) = \inf\{t : U_n(\lambda, t) \geq p\},\ 0 \leq \lambda \leq 1.$$

Remark 2.3. (i) Given a kernel $g(x, y)$, we can define

$$h(x, y, t) = 1_{\{g(x,y)\leq t\}}.$$

Then, $U_n(\lambda, \cdot)$ is the empirical distribution function of the data $g(X_i, X_j)$, $1 \leq i \leq [n\lambda]$, $[n\lambda]+1 \leq j \leq n$, and $Q_n(\lambda)$ is the p-th quantile of the same data.
(ii) For fiixed t, the process $(U_n(\lambda, t))_{0\leq\lambda\leq1}$ is a two-sample U-process that has been introduced and investigated by Dehling, Fried, Garcia and Wendler (2013).

In this paper, we will study the asymptotic distribution of the two-sample empirical U-quantile process $(Q_n(\lambda))_{0\leq\lambda\leq1}$ in the case of weakly dependent data $(X_i)_{i\geq1}$. Before we can formulate the results, we have to make some further definitions.

Definition 2.4. *Let $h(x, y, t)$ be a measurable function, and let X, Y be independent random variables with the same distribution as X_i. Then we define the functions $U(t), h_1(x, t)$, and $h_2(y, t)$ by*

$$U(t) = Eh(X, Y, t) \tag{1}$$
$$h_1(x, t) = Eh(x, Y, t) - U(t) \tag{2}$$
$$h_2(y, t) = Eh(X, y, t) - U(t). \tag{3}$$

Moreover, we define the quantile function

$$Q(p) = \inf\{t : U(t) \geq p\},$$

and the p-th quantile $t_p = Q(p)$.

Our theorems will require various technical assumptions regarding the process $(X_i)_{i\geq1}$ and the kernel $h(x, y, t)$, which we list now.

(A1) $(X_n)_{n\geq1}$ is a near epoch dependent functional of an absolutely regular process $(Z_n)_{n\in\mathbb{Z}}$ with mixing coefficients $\beta(n)_{n\geq1}$ and approximation constants $(a_n)_{n\geq1}$ satisfying

$$\begin{aligned} \beta(n) &= O(n^{-\beta}) \\ a_n &= O(n^{-(\beta+3)}) \end{aligned}$$

for some constant $\beta > 3$.

(A2) $(U(t))$, as defined in (1) is differentiable in a neighborhood of t_p. Moreover, $u(t) = U'(t)$ satisfies $u(t_p) > 0$, and

$$|U(t) - p - u(t_p)(t - t_p)| = O(|t - t_p|^{3/2}),$$

as $t \to t_p$.

(A3) The kernel $h : \mathbb{R}^2 \times \mathbb{R}$ is a bounded measurable function. Moreover, $t \mapsto h(x, y, t)$ is nondecreasing, and $(x, y) \mapsto h(x, y, t)$ is uniformly p-Lipschitz continuous in a neighborhood of t_p. I.e., there exists a neighborhood of t_p and a constant $L > 0$ such that

$$\begin{aligned} E\left(|h(X, Y, t) - h(X', Y, t)|1_{\{|X-X'|\leq\epsilon\}}\right) &\leq L\,\epsilon \\ E\left(|h(X, Y, t) - h(X, Y', t)|1_{\{|Y-Y'|\leq\epsilon\}}\right) &\leq L\,\epsilon \end{aligned}$$

holds for all t in this neighborhood, for all $\epsilon > 0$, and for all quadruples X, Y, X', Y' of random variables such that (X, Y) has joint distribution $P_{X_1} \times P_{X_1}$ or P_{X_1,X_k}, for some k, and such that X' and Y' each have the same marginal distribution as X_i.

Theorem 2.5. *Let $(X_i)_{i\geq 1}$ be a near epoch dependent functional of an absolutely regular process such that assumption (A1) is satisfied. Moreover, let $h : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ be a measurable kernel such assumptions (A2) and (A3) hold. Then*

$$\sqrt{n}\,(\lambda(1-\lambda)(Q_n(\lambda,p)-Q(p)))_{0\leq\lambda\leq 1} \xrightarrow{\mathcal{D}} ((1-\lambda)W_1(\lambda)+\lambda(W_2(1)-W_2(\lambda)))_{0\leq\lambda\leq 1},$$

where $(W_1(\lambda), W_2(\lambda))$ is a two-dimensional Brownian motion with covariance structure

$$\mathrm{Cov}(W_i(\mu), W_j(\lambda)) = (\mu \wedge \lambda)\frac{1}{u^2(Q(p))}\sum_{k\in\mathbb{Z}} E(h_i(X_0; Q(p)), h_j(X_k; Q(p))).$$

Here $u(t) = \frac{d}{dt}U(t)$.

An important ingredient in the proof of the limit theorem for the two-sample U-quantile process is the Bahadur-Kiefer representation of the U-quantiles, see Bahadur (1966). The Bahadur-representation for two-sample U-quantiles (with fixed λ) have been studied by Inagaki (1973) for independent data and by Dehling and Fried (2012) for dependent data. To the best of our knowledge, there are no results for the process indexed by λ. There is much more literature on one-sample U-quantiles, beginning with Geertsema (1970). In this case, better rates of the Bahadur representation are known, see Dehling et al. (1987), Choudhury and Serfling (1988) and Arcones (1996) for the independent case, Wendler (2011) for the dependent case.

Theorem 2.6. *Under the same assumptions as in Theorem 2.5, we obtain*

$$\sup_{0\leq\lambda\leq 1} \lambda(1-\lambda)\left(Q_n(\lambda,p) - Q(p) + \frac{U_n(\lambda, Q(p)) - p}{u(Q(p))}\right) = O_P(n^{-\frac{5}{9}}).$$

Here $u(t) = \frac{d}{dt}U(t)$.

3. Simulation Results

The practical value of the theoretical results presented above is illustrated in a simulation study. We generate time series of length $n = 200$ from first order autoregressive models with parameter $\phi \in \{0, 0.4, 0.8\}$ and the innovations stemming from scaled t-distributions with $\nu \in \{2, 3, \infty\}$ degrees of freedom. In case of $\nu = \infty$ this is a standard normal distribution, while $\nu = 3$ and $\nu = 2$ correspond to the cases where the variance just exists or does not exist. All t-distributions are scaled to have $F_\nu(1) = 0.8413447$, like for the standard normal, to ease comparison.

For estimation of the long-run variance σ^2, needed for our change-point test based on the Hodges-Lehmann estimator (HLE), we consider the fixed block length $l_n = [(3n)^{1/3} + 1)]$, corresponding to $l_n = 9$ when $n = 200$. This agrees well with the findings of Dehling et al. (2013) for ARMA(1,1) processes. Additionally, we consider the adaptive block length

$$l_n = \max(\lceil n^{1/3}(2\phi/(1-\phi^2))^{2/3}\rceil, 1). \tag{4}$$

Carlstein (1986) proved that this block length minimizes the MSE of the estimator for the long-run variance of the CUSUM test asymptotically in case an AR(1)-process with autoregression coefficient ϕ. Dehling et al. (2013) obtained good results also when applying this adaptive block length for subsampling estimation of the long-run variance of their change-point test based on the Wilcoxon-Mann-Whitney (WMW) statistic, replacing ϕ by the lag-one sample autocorrelation of the series $F_n(x_t), t = 1, \ldots, n$.

The sizes of the different tests under the different error distributions are assessed by applying the tests with the critical value 1.36, corresponding to an asymptotic 5% significance level, to 4000 time series without shift. Then we generate 400 time series for shifts with each

TABLE 1. Empirical sizes of the different tests for differently strong autocorrelations (ϕ) and heaviness of the tails (ν).

		fixed block length			adaptive block length		
ϕ	ν	CUSUM	WMW	HLE	CUSUM	WMW	HLE
0.0	∞	3.9	2.9	3.6	3.8	2.2	2.8
0.0	3	3.5	2.9	3.7	3.1	2.4	2.8
0.0	2	3.4	3.1	4.8	2.5	2.2	3.4
0.4	∞	4.9	3.1	3.8	6.0	3.9	4.3
0.4	3	3.8	3.0	3.9	4.9	3.3	4.0
0.4	2	3.6	3.0	4.5	4.2	3.8	5.1
0.8	∞	10.6	6.5	7.1	4.0	2.5	2.9
0.8	3	10.5	7.0	7.7	3.9	2.8	3.1
0.8	2	8.8	6.7	10.7	3.7	2.3	5.7

of different heights $h = 0.1, 0.2, \ldots, 1$ in case of $\phi = 0$, $h = 0.2, 0.4, \ldots, 2$ under $\phi = 0.4$ and $h = 0.4, 0.8, \ldots, 4$ under $\phi = 0.8$ to compare the power of the different tests.

Figure 1 illustrates the power of the different tests in case of a shift of increasing height in time series with normal innovations. As expected, the CUSUM is usually the most powerful test under normality. In case of a shift in the center of the time series, the WMW and the HLE test are close competitors, with the latter providing good power also when the shift is not in the center. The adaptive block length for the subsampling increases the power in case of small or moderate positive autocorrelations, particularly for a shift outside the center, and it stabilizes the size of the tests in case of large positive autocorrelations. Note that the HLE test with adaptive block length performs even better than the corresponding CUSUM test in case of strong positive autocorrelations. Estimation of the asymptotic variance is less vulnerable to shifts in case of the HLE test than for the CUSUM, since we deal with autocovariances of random variables which are transformed to the bounded interval $[0, 1]$.

Figure 2 compares the power of the different tests in case of a shift of increasing height in time series with t_3-distributed innovations. The asymptotic theory underlying the CUSUM still applies for this heavy-tailed distribution, since the variance exists. Nevertheless, our implementations of the CUSUM test provide smaller power than the WMW and HLE tests. Again we find little difference between the WMW and the HLE test if a shift occurs in the center of the series, and a substantial advantage of the HLE test if the shift occurs outside the center. The increased power in case of small positive autocorrelations and the better size preservation in case of large positive autocorrelations resulting from the adaptive block length is also confirmed. We get similar results for the situation of t_2-distributed innovations, see Figure 3. Although the asymptotic theory underlying the CUSUM test does not apply, it still preserves the size if the autocorrelations are moderate, but the advantages of the WMW and even more the HLE test in terms of power increase as the tails get heavier.

Table 1 indicates that the better power of the HLE test as compared to the WMW test is in part due to its size being closer to the nominal significance level. The WMW test is more conservative, particularly if the autocorrelation ϕ is large or the degrees of freedom ν are small.

The tests employing Carlstein's adaptive choice of the block length could be improved further by using a more sophisticated estimate of ρ than the sample autocorrelation coefficient applied here. The latter is positively biased in the presence of a shift, which leads to undesirably large choices of the block length. This negative effect becomes more severe for larger values of ρ, since the plug-in-estimate of the asymptotically MSE-optimal choice of

Figure 1. Power of the tests in case of a shift in the center (left) or after 3/4 (right) of AR(1) time series with $\phi = 0$ (top), $\phi = 0.4$ (middle) or $\phi = 0.8$ (bottom) and normal innovations, length $n = 200$. CUSUM (dotted), WMW (dashed), HLE (solid) with fixed (black) or adaptive (grey) block length for the subsampling.

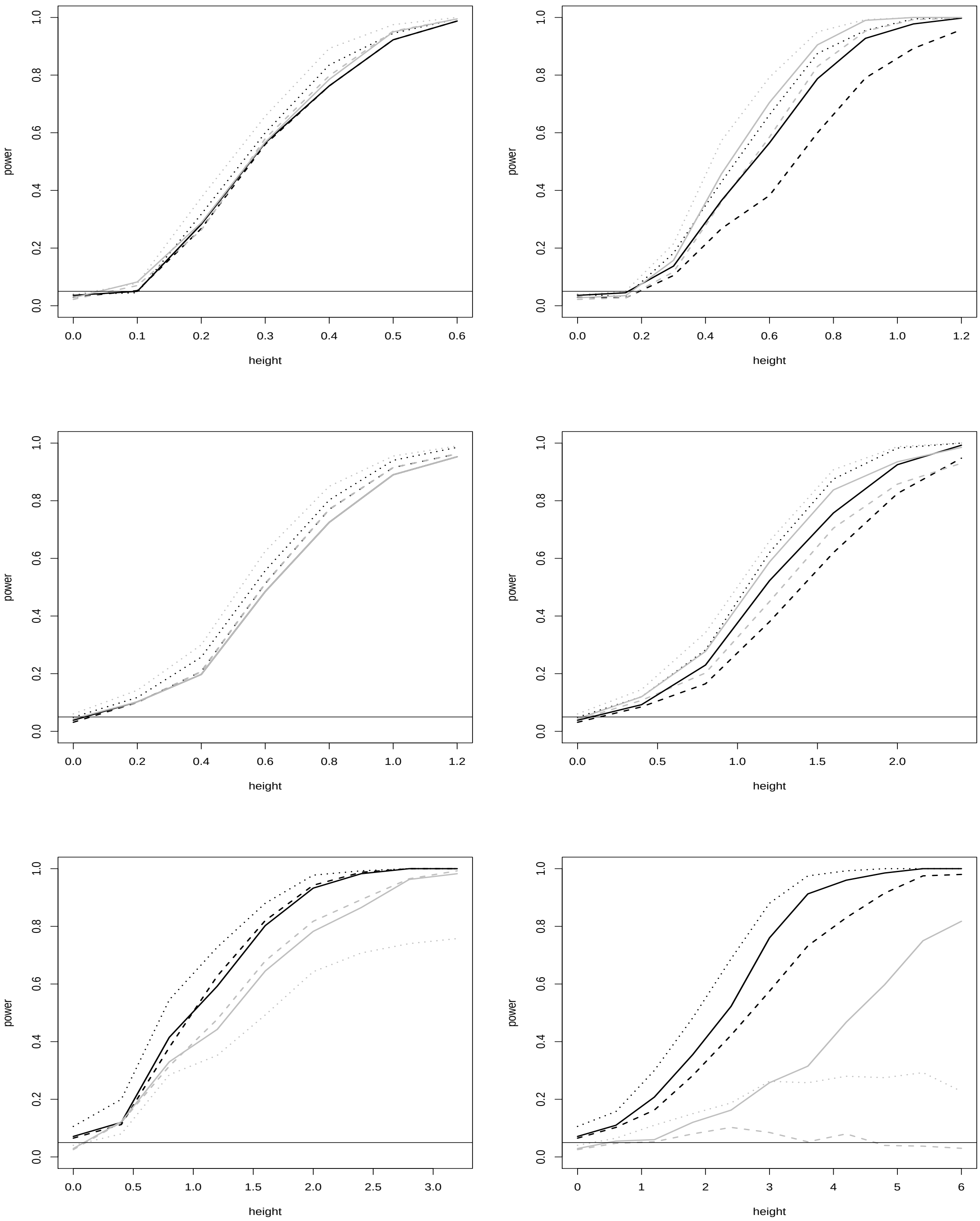


FIGURE 2. Power of the tests in case of a shift in the center (left) or after 3/4 (right) of AR(1) time series with $\phi = 0$ (top), $\phi = 0.4$ (middle) or $\phi = 0.8$ (bottom) and t_3 innovations, length $n = 200$. CUSUM (dotted), WMW (dashed), HLE (solid) with fixed (black) or adaptive (grey) block length for the subsampling.

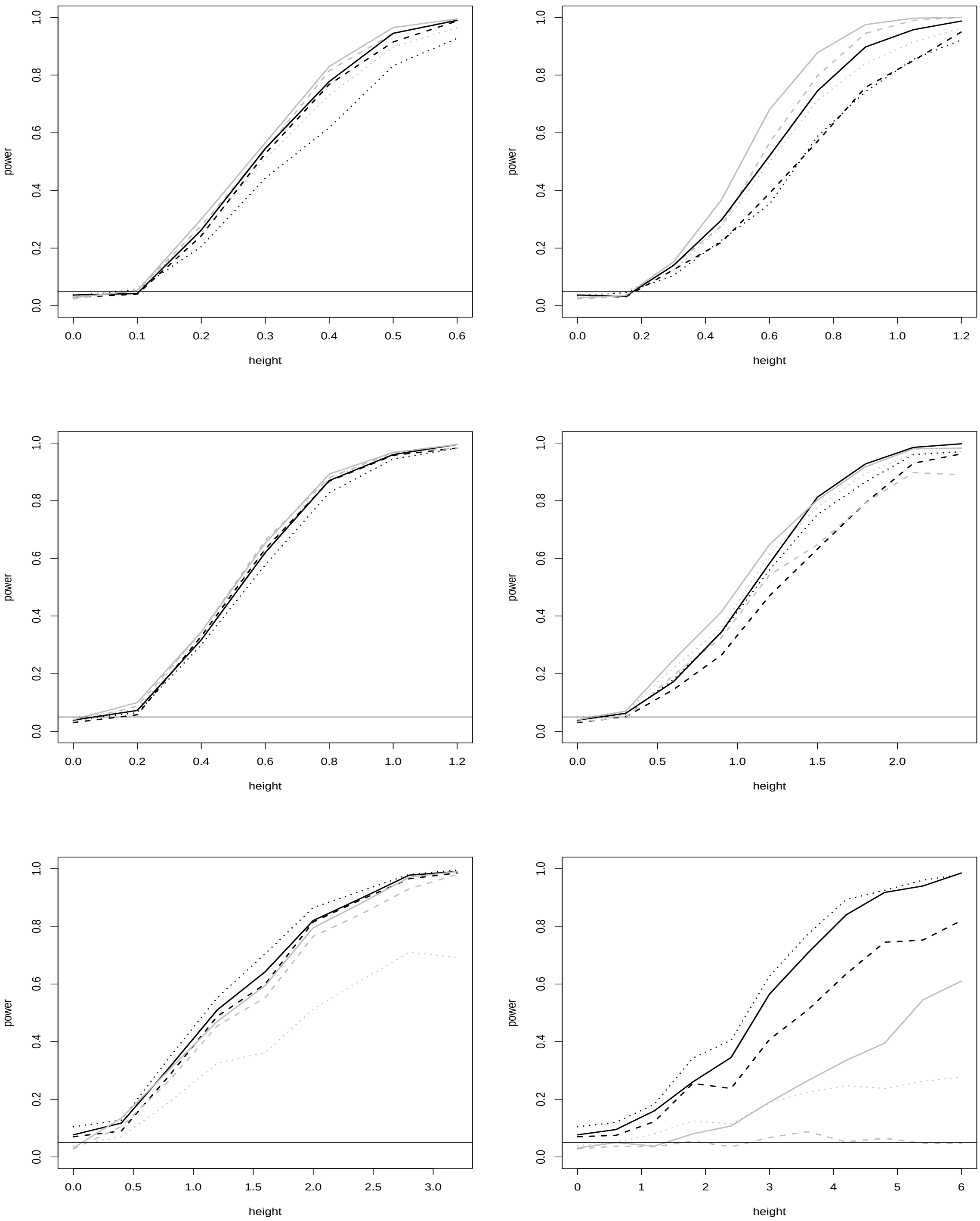
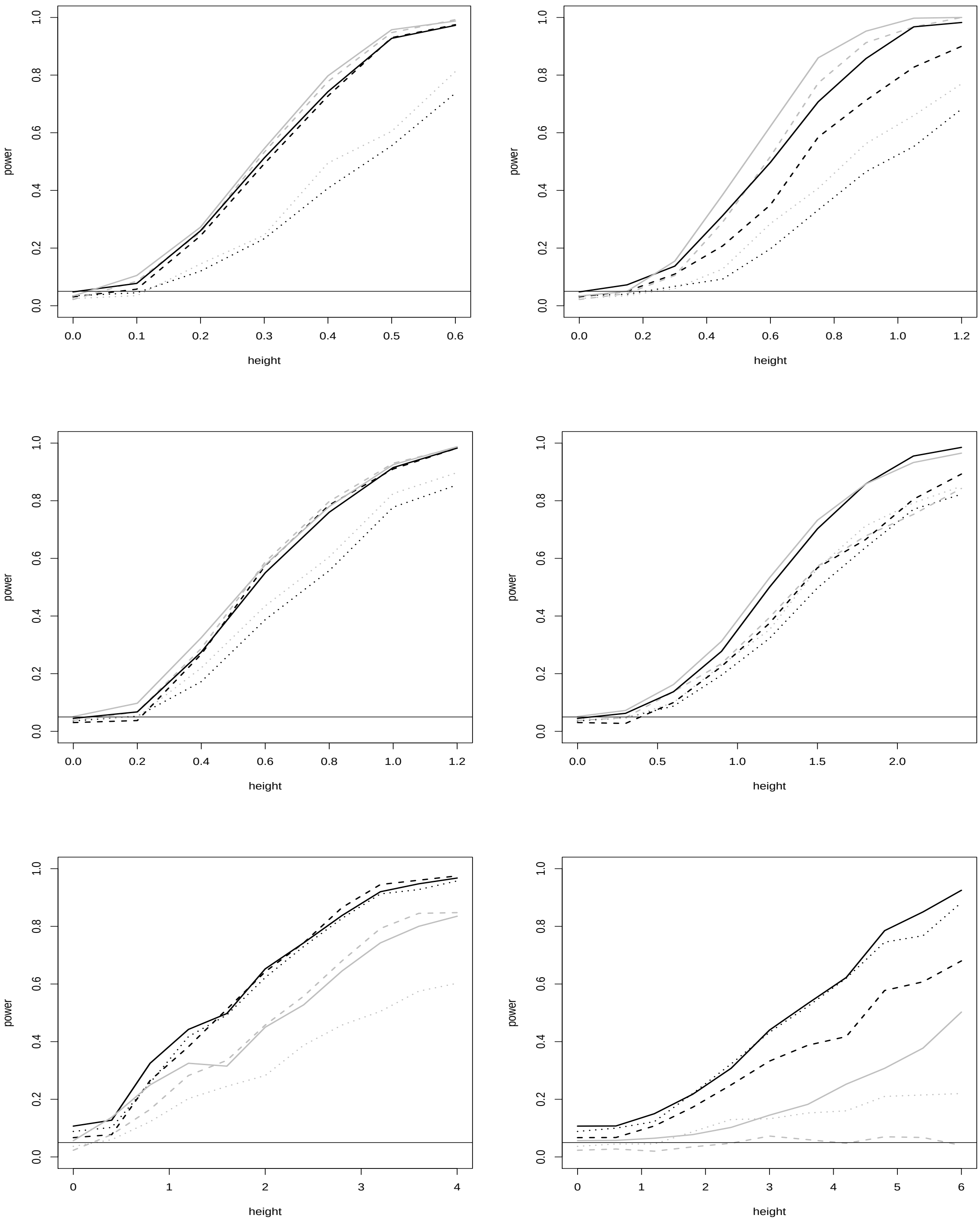


FIGURE 3. Power of the tests in case of a shift in the center (left) or after 3/4 of AR(1) time series with $\phi = 0$ (top), $\phi = 0.4$ (middle) or $\phi = 0.8$ (bottom) and t_2-distributed innovations, length $n = 200$. CUSUM (dotted), WMW (dashed), HLE (solid) with fixed (black) or adaptive (grey) block length for the subsampling.



l_n increases more rapidly if $\hat{\phi}$ is close to 1, while it is rather stable for moderate and small values of $\hat{\phi}$. In our study, in case of $\phi = 0$ the average value chosen for l_n increases from about 2 to about 4 on average as the height of the shift increases, while it is from about 6 to about 10 if $\phi = 0.4$, and even from about 16 to about 24 if $\phi = 0.8$. A robust estimate of the autocorrelation coefficient resisting shifts could be used, but this is left for future research.

We also considered direct estimation of the asymptotic variance after correcting the data for a possible shift as proposed by Huskova and Kirch (2010), but our implementation provided substantially oversized tests. When correcting the sizes of the tests by using the empirical 95% percentile of the absolute values of the test statistics, derived from time series without a shift generated from the same model, the differences between the tests were less pronounced than those presented here. However, such a comparison is not realistic, since in practice we usually know neither the time series model nor the type of innovations and can thus not use such critical values derived from simulations. Instead, bootstrap procedures might be an interesting alternative, but this will not be pursued here.

4. Data analysis

For illustration of the gains in power arising from the HLE test as compared to the CUSUM and the WMW test we analyze the monthly averages of the daily minimum temperatures in Potsdam, Germany, from 1893 to 2010. The 1416 data points have been deseasonalized by subtracting the median value from each calendar month, see Figure 4. We are interested in whether the level of this data set is constant or whether there is a monotonic change. Such a change is likely to show a trend-like behavior and not a jump, but nevertheless a change-point test should detect such a change if its null hypothesis is a constant level.

The empirical autocorrelation and partial autocorrelations suggest a first order autoregressive model with lag-one autocorrelation about 0.25 for the deseasonalized data. The CUSUM and the HLE test statistics take their maximum value in 1987 after time point 1136, while the WMW test takes it in 1943 after time point 595, i.e. rather in the middle of the time series. The resulting p-values are 0.002, 0.002 and below 0.001 for the CUSUM, the WMW and the HLE test with the fixed block length. All p-values are even below 0.001 when using the adaptive block length.

When dividing the data into the times periods before and after the detected change-point, the HLE test with adaptive block length yields a further significant p-value of 0.042 after time point 380 in 1924. For the HLE with fixed block length and the WMW with adaptive block length the p-value is 0.062, while the CUSUM test is far from being significant and does not reject the null hypothesis of constant levels within the subsequences. The multiple change-points detected by the HLE test with adaptive block length might be due to a monotonic trend and can be explained by the superior power of this test in case of heavy tailed data like daily minima. The HLE estimates an increase of the temperatures by 0.355^oC from the first to the second and by another 0.765^oC from the second to the third period.

5. Proofs

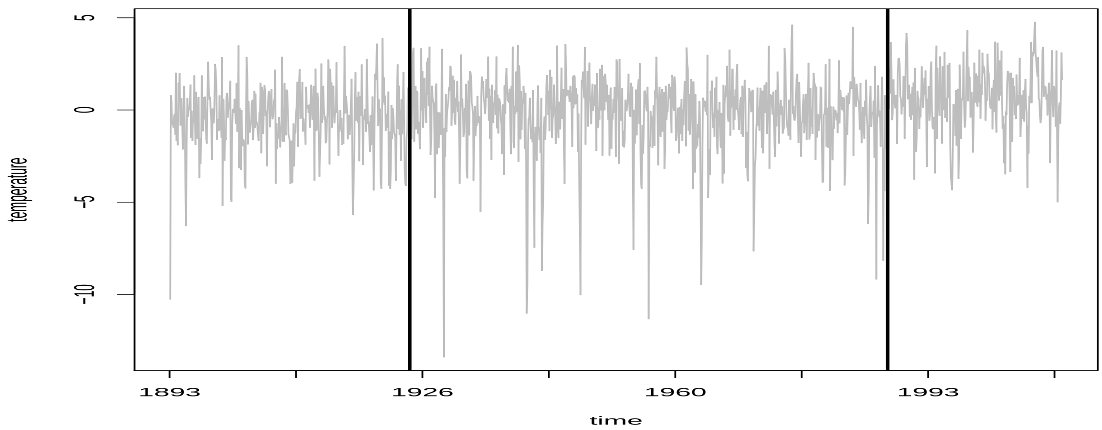
5.1. **Auxiliary Results.** The proofs require some further notations, which we introduce now. Given the kernel $h(x, y, t)$, we define the two-sample empirical U-process

$$U_{n_1,n_2}(t) := \frac{1}{n_1\, n_2} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n_1+n_2} h(X_i, X_j, t),$$

and the two-sample empirical U-quantile process

$$Q_{n_1,n_2}(p) = U_{n_1,n_2}^{-1}(p) = \inf\{t : U_{n_1,n_2}(t) \geq p\}.$$

FIGURE 4. Deseasonalized time series representing the monthly average daily minimum temperatures in Potsdam, Germany, and change-points detected by the HLE test with adaptive block length.



Note that $U_n(\lambda, t) = U_{[n\lambda], n-[n\lambda]}(t)$ and $Q_n(\lambda, p) = Q_{[n\lambda], n-[n\lambda]}(p)$. Moreover, we define

$$g(x, y, t) = h(x, y, t) - h_1(x, t) - h_2(y, t) - U(t),$$

where $h_1(x, t)$, $h_2(y, t)$, and $U(t)$ have been defined in (1), (2) and (3), respectively. Thus, we obtain the Hoeffding decomposition of the two-sample U-statistic as

$$U_{n_1,n_2}(t) := U(t) + \frac{1}{n_1}\sum_{i=1}^{n_1} h_1(X_i, t) + \frac{1}{n_2}\sum_{j=n_1+1}^{n_1+n_2} h_2(X_j, t) + \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n_1+n_2} g(X_i, X_j, t).$$

The next two lemmas will deal with the last sum, which is called degenerate part:

Lemma 5.1. *Under the assumptions (A1) and (A3), there exists a constant C, such that for any integers $0 \le m_1 \le n_1 \le m_2 \le n_2$*

$$E\left(\sum_{i=m_1+1}^{n_1}\sum_{j=m_2+1}^{n_2} g(X_i, X_j, t)\right)^2 \le C(n_1 - m_1)(n_2 - m_2),$$

for all t in the neighborhood refered to in assumption (A3).

Proof. For the special case $m_2 = n_1$, this is Proposition 6.2 of Dehling and Fried (2012). The more general case can be proved with the same arguments, we omit the details. □

Lemma 5.2. *Suppose that the assumptions (A1) and (A3) hold.*
(i) There is a constant C, such that

$$E\left(\max_{0\le m_1\le n_1\le m_2\le n_2\le 2^l}\left|\sum_{i=m_1+1}^{n_1}\sum_{j=m_2+1}^{n_2} g(X_i, X_j, t)\right|\right)^2 \le C 2^{2l} l^4.$$

(ii) As $n \to \infty$, we have

$$\sup_{0\le\lambda\le 1}\left|\sum_{i=1}^{\lfloor \lambda n\rfloor}\sum_{j=\lfloor\lambda n\rfloor+1}^{n} g(X_i, X_j, t)\right| = o(n \log^3 n),$$

almost surely.

Proof. To prove the first part of the lemma, we introduce the notation

$$Q_{m_1,n_1,m_2,n_2}=\sum_{i=m_1+1}^{n_1}\sum_{j=m_2+1}^{n_2}g(X_i,X_j,t),$$

for $m_1\leq n_1\leq m_2\leq n_2$, and $Q_{m_1,n_1,m_2,n_2}=0$ otherwise. These quantities satisfy an addition rule

$$Q_{m_1,n_1,m_2,n_2}+Q_{n_1,n_1',m_2,n_2}=Q_{m_1,n_1',m_2,n_2}$$
$$Q_{m_1,n_1,m_2,n_2}+Q_{m_1,n_1,n_2,n_2'}=Q_{m_1,n_1,m_2,n_2'}$$

Note that

$$\max_{0\leq m_1\leq n_1\leq m_2\leq n_2\leq 2^l}|Q_{m_1,n_1,m_2,n_2}|\leq 2\max_{0\leq m_1\leq n_1\leq m_2\leq 2^l}|Q_{m_1,n_1,m_2,2^l}|\leq 4\max_{0\leq n_1\leq m_2\leq 2^l}|Q_{0,n_1,m_2,2^l}|$$

Now we use a chaining technique. For example,

$$|Q_{0,5,7,16}|\leq|Q_{0,4,7,8}|+|Q_{0,4,8,16}|+|Q_{4,5,7,8}|+|Q_{4,5,8,16}|.$$

We conclude that

$$\max_{0\leq m_1\leq n_1\leq m_2\leq n_2\leq 2^l}|Q_{m_1,n_1,m_2,n_2}|\leq 4\sum_{d_1=0}^{l}\sum_{d_2=0}^{l}\max_{\substack{i=1,\dots,2^{l-d_1}\\ j=1,\dots,2^{l-d_2}}}|Q_{((i-1)2^{d_1},i2^{d_1},(j-1)2^{d_2},j2^{d_2})}|.$$

Note that for any random variables $Y_1,\dots,Y_k$ we have that $E\left(\max_{i=1,\dots,k}Y_k\right)^2\leq\sum_{i=1}^k EY_i^2$. Using this inequality and Lemma 5.1, we conclude that

$$\begin{aligned}
&E\left(\max_{0\leq m_1\leq n_1\leq m_2\leq n_2\leq 2^l}|Q_{m_1,n_1,m_2,n_2}|\right)^2\\
&\quad\leq 16E\left(\sum_{d_1=0}^{l}\sum_{d_2=0}^{l}\max_{\substack{i=1,\dots,2^{l-d_1}\\ j=1,\dots,2^{l-d_2}}}|Q_{((i-1)2^{d_1},i2^{d_1},(j-1)2^{d_2},j2^{d_2})}|\right)^2\\
&\quad\leq 16l^2\sum_{d_1=0}^{l}\sum_{d_2=0}^{l}E\left(\max_{\substack{i=1,\dots,2^{l-d_1}\\ j=1,\dots,2^{l-d_2}}}|Q_{((i-1)2^{d_1},i2^{d_1},(j-1)2^{d_2},j2^{d_2})}|\right)^2\\
&\quad\leq 16l^2\sum_{d_1=0}^{l}\sum_{d_2=0}^{l}\sum_{i=1}^{2^{l-d_1}}\sum_{j=1}^{2^{l-d_2}}E\left(Q_{((i-1)2^{d_1},i2^{d_1},(j-1)2^{d_2},j2^{d_2})}\right)^2\\
&\quad\leq Cl^2\sum_{d_1=0}^{l}\sum_{d_2=0}^{l}\sum_{i=1}^{2^{l-d_1}}\sum_{j=1}^{2^{l-d_2}}2^{d_1}2^{d_2}\leq Cl^4 2^{2l}.
\end{aligned}$$

So the first part of the lemma is proved. For the second part, it suffices to show that

$$\max_{0\leq m_1\leq n_1\leq m_2\leq n_2\leq 2^l}|Q_{m_1,n_1,m_2,n_2}|=o(2^l l^3).$$

Now by the Chebyshev inequality, we obtain

$$\begin{aligned}
&\sum_{l=1}^{\infty}P\left(\frac{1}{2^l l^3}\max_{0\leq m_1\leq n_1\leq m_2\leq n_2\leq 2^l}|Q_{m_1,n_1,m_2,n_2}|\geq\epsilon\right)\\
&\qquad\leq\frac{1}{\epsilon^2}\sum_{l=1}^{\infty}\frac{1}{2^{2l}l^6}E\left(\max_{0\leq m_1\leq n_1\leq m_2\leq n_2\leq 2^l}|Q_{m_1,n_1,m_2,n_2}|\right)^2\leq C\sum_{l=1}^{\infty}\frac{1}{l^2}<\infty.
\end{aligned}$$

The Borel-Cantelli lemma completes the proof. □

In order to prove Theorem 2.6, we need some information about the local behaviour of the empirical U-process. We will first concentrate on the first half of the process, i.e. $\lambda \in [0, 1/2]$:

Lemma 5.3. *Under the assumptions (A1), (A2), and (A3),*

$$\sup_{\substack{\lambda\in[0,1/2] \\ |t-t_p|\leq C\sqrt{\frac{\log\log(\min\{\lambda,1-\lambda\}n)}{\min\{\lambda,1-\lambda\}n}}}} \lambda(1-\lambda)\left|(U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t)-U(t))-(U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p)-p)\right| = O\left(n^{-\frac{5}{9}}\right)$$

almost surely.

Proof. We define $n_1 = \lfloor n\lambda\rfloor$ and $n_2 = n - \lfloor n\lambda\rfloor$, and note that $n_1 + n_2 = n$. We define the sequences $c_{2^l} = 2^{-\frac{5}{9}l}$, and for $n = 2^{l-1}+1, \ldots, 2^l$ we set $c_n = c_{2^l}$. By the monotonicity of U_{n_1,n_2} and U in t, we have that

$$\begin{aligned}
&\sup_{\substack{n_1\leq\frac{n}{2} \\ |t-t_p|\leq C\sqrt{\frac{\log\log n_1}{n_1}}}} \frac{n_1n_2}{n^2}\left|(U_{n_1,n_2}(t)-U(t))-(U_{n_1,n_2}(t_p)-p)\right| \\
&\leq \max_{\substack{n_1\leq\frac{n}{2} \\ t\in c_n\mathbb{Z} \\ |t-t_p|\leq C\sqrt{\frac{\log\log n_1}{n_1}}}} \frac{n_1n_2}{n^2}\left|(U_{n_1,n_2}(t)-U(t))-(U_{n_1,n_2}(t_p)-p)\right| + \max_{\substack{t\in c_n\mathbb{Z} \\ |t-t_p|\leq C\sqrt{\frac{\log\log n_1}{n_1}}}} |U(t)-U(t+c_n)|.
\end{aligned}$$

As U is differentiable in t_p, we get that the second summand is of the order $O(c_n)$. For the first summand, we use the Hoeffding decomposition and get

$$\begin{aligned}
(5)\quad &\max_{\substack{n_1\leq\frac{n}{2} \\ t\in c_n\mathbb{Z} \\ |t-t_p|\leq C\sqrt{\frac{\log\log n_1}{n_1}}}} \frac{n_1n_2}{n^2}\left|(U_{n_1,n_2}(t)-U(t))-(U_{n_1,n_2}(t_p)-p)\right| \\
\leq\ &\max_{\substack{n_1\leq\frac{n}{2} \\ t\in c_n\mathbb{Z} \\ |t-t_p|\leq C\sqrt{\frac{\log\log n_1}{n_1}}}} \left|\frac{1}{n}\sum_{i=1}^{n_1} h_1(X_i,t) - \frac{1}{n}\sum_{i=1}^{n_1} h_1(X_i,t_p)\right| \\
&+ \max_{\substack{n_1\leq\frac{n}{2} \\ t\in c_n\mathbb{Z} \\ |t-t_p|\leq C\sqrt{\frac{\log\log n_1}{n_1}}}} \frac{n_1}{n}\left|\frac{1}{n}\sum_{j=n_1+1}^{n} h_2(X_j,t) - \frac{1}{n}\sum_{j=n_1+1}^{n} h_2(X_j,t_p)\right| \\
&+ \max_{\substack{n_1\leq\frac{n}{2} \\ t\in c_n\mathbb{Z} \\ |t-t_p|\leq C\sqrt{\frac{\log\log n_1}{n_1}}}} \left|\frac{1}{n^2}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n_1+n_2} g(X_i,X_j,t)\right| + \max_{n_1\leq\frac{n}{2}}\left|\frac{1}{n^2}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n_1+n_2} g(X_i,X_j,t_p)\right|
\end{aligned}$$

For the first summand, we refer to (13) in Theorem 1 of Wendler [34] and conclude that it is of size $o(n^{-\frac{5+\gamma}{8}}(\log n)^{\frac{3}{4}}(\log\log n)^{\frac{1}{2}}) = O(n^{-\frac{5}{9}})$ almost surely for a $\gamma > 0$. Note that the continuity condition on the kernel in [34] is different, but the continuity is only needed to guarantee that $(h_1(X_i))_{i\in\mathbb{N}}$ is near epoch dependent. This also holds under our continuity condition by Proposition 2.11 from Borovkova et al. [5].

We split the second summand into two parts, so that for the first part n_1/n is small and for the second part $\sqrt{\frac{\log\log n_1}{n_1}}$:

$$\begin{aligned}
&\max_{\substack{n_1\leq \frac{n}{2}\\ t\in c_n\mathbb{Z}\\ |t-t_p|\leq C\sqrt{\frac{\log\log n_1}{n_1}}}} \frac{n_1}{n}\left|\frac{1}{n}\sum_{j=n_1+1}^{n} h_2(X_j,t)-\frac{1}{n}\sum_{j=n_1+1}^{n} h_2(X_j,t_p)\right|\\
\leq\ & \max_{\substack{n_1\leq n^{\frac{4}{9}}\\ t\in c_n\mathbb{Z}\\ |t-t_p|\leq C\sqrt{\frac{\log\log n_1}{n_1}}}} n^{-5/9}\left|\frac{1}{n}\left(\sum_{j=n_1+1}^{n} h_2(X_j,t)-\sum_{j=n_1+1}^{n} h_2(X_j,t_p)\right)\right|\\
&+\max_{\substack{n_1\leq \frac{n}{2}\\ t\in c_n\mathbb{Z}\\ |t-t_p|\leq C\sqrt{\frac{\log\log n^{4/9}}{n^{4/9}}}}} \left|\frac{1}{n}\sum_{j=n_1+1}^{n} h_2(X_j,t)-\frac{1}{n}\sum_{j=n_1+1}^{n} h_2(X_j,t_p)\right|\\
=:\ & A_1+A_2.
\end{aligned}$$

As h is bounded and therefore h_2 is bounded, we have that $A_1=O(n^{-\frac{5}{9}})$. Along the lines of the proof of Theorem 1 in Wendler [34], we obtain

$$A_2=O(n^{-\frac{2}{9}\frac{1+\gamma}{4}}n^{-\frac{1}{2}}(\log n)^{\frac{3}{4}}(\log\log n)^{\frac{1}{2}})=O(n^{-\frac{5}{9}})$$

almost surely. For the third summand on the r.h.s. of (5), we use the first part of Lemma 5.2, the Chebyshev inequality, and the fact that the second moment of the maximum of random variables is smaller or equal to the sum of second moments. We obtain

$$\begin{aligned}
&\sum_{l=1}^{\infty} P\left(\frac{1}{c_{2^l}}\max_{2^{l-1}\leq n\leq 2^l}\max_{\substack{n_1\leq \frac{n}{2}\\ t\in c_n\mathbb{Z}\\ |t-t_p|\leq C\sqrt{\frac{\log\log n_1}{n_1}}}}\left|\frac{1}{n^2}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n_1+n_2} g(X_i,X_j,t)\right|\geq\epsilon\right)\\
\leq&\sum_{l=1}^{\infty}\frac{1}{c_{2^l}^2\epsilon^2 2^{4(l-1)}}E\left(\max_{2^{l-1}\leq n\leq 2^l}\max_{\substack{n_1\leq \frac{n}{2}\\ t\in c_n\mathbb{Z}\\ |t-t_p|\leq C\sqrt{\frac{\log\log n_1}{n_1}}}}\left|\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n_1+n_2} g(X_i,X_j,t)\right|\right)^2\\
\leq&\sum_{l=1}^{\infty}\sum_{\substack{t\in c_{2^l}\mathbb{Z}\\ |t-t_p|\leq C}}\frac{1}{c_{2^l}^2\epsilon^2 2^{4(l-1)}}E\left(\max_{0\leq n_1\leq n_1+n_2\leq 2^l}\left|\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n_1+n_2} g(X_i,X_j,t)\right|\right)^2\\
\leq& C\sum_{l=1}^{\infty}\frac{1}{c_{2^l}^3 2^{4l}}2^{2l}l^4\leq C\sum_{l=1}^{\infty}\frac{l^4}{2^{\frac{1}{3}l}}<\infty,
\end{aligned}$$

as the set $\{t\in c_{2^l}\mathbb{Z},\ |t-t_p|\leq C\}$ has at most $C\,c_{2^l}^{-1}$ elements. Using the Borel-Cantelli lemma, we conclude that the third summand on the r.h.s. of (5) is of size $o(c_n)$ almost surely. The last summand can be treated in the same way and so in total we have proved the order $O(n^{-\frac{5}{9}})$ almost surely. □

Lemma 5.4. *Unter the assumptions (A1) and (A3),*

$$\sup_{n_2>n_1}|U_{n_1,n_2}(t_p)-p|=O\left(\sqrt{\frac{\log\log n_1}{n_1}}\right)$$

almost surely.

Proof. We use the Hoeffding decomposition

$$\sup_{n_2>n_1} |U_{n_1,n_2}(t_p) - p| \leq \left| \frac{1}{n_1} \sum_{i=1}^{n_1} h_1(X_i, t_p) \right| + \left| \frac{1}{n_2} \sum_{i=1}^{n_1} h_2(X_i, t_p) \right| + \sup_{n_2>n_1} \frac{1}{n_2} \left| \sum_{j=1}^{n_1+n_2} h_2(X_j, t_p) \right| + \sup_{n_2>n_1} \frac{1}{n_1 n_2} \left| \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n_2} g(X_i, X_j, t_p) \right|.$$

For the first two summands, we use Proposition 3.7 of Wendler [34], which leads to

$$\left| \frac{1}{n_1} \sum_{i=1}^{n_1} h_k(X_i, t_p) \right| = O\left(\sqrt{\frac{\log\log n_1}{n_1}} \right)$$

for $k = 1, 2$ almost surely. Furthermore

$$\sup_{n_2>n_1} \frac{1}{n_2} \left| \sum_{j=1}^{n_1+n_2} h_2(X_j, t_p) \right| \leq \sup_{n_2>n_1} \frac{2}{n_1+n_2} \left| \sum_{j=1}^{n_1+n_2} h_2(X_j, t_p) \right| = O\left(\sqrt{\frac{\log\log n_1}{n_1}} \right).$$

For the last summand, we use Lemma 5.2 to obtain

$$E\left(\max_{\substack{0\leq m_1\leq n_1\leq 2^{l_1} \\ n_1\leq m_2\leq n_2\leq 2^{l_2}}} |Q_{m_1,n_1,m_2,n_2}| \right)^2 \leq C l_1^2 l_2^2 2^{l_1} 2^{l_2}.$$

Now by the Chebyshev inequality, we obtain

$$\sum_{l_2=1}^{\infty} \sum_{l_1=1}^{l_2} P\left(\sqrt{\frac{1}{2^{l_1}\log l_1}} \frac{1}{2^{l_2}} \max_{\substack{0\leq m_1\leq n_1\leq 2^{l_1} \\ n_1\leq m_2\leq n_2\leq 2^{l_2}}} |Q_{m_1,n_1,m_2,n_2}| \geq \epsilon \right)$$
$$\leq \frac{1}{\epsilon^2} \sum_{l_2=1}^{\infty} \sum_{l_1=1}^{l_2} \frac{1}{2^{l_1}\log l_1 2^{2l_2}} E\left(\max_{\substack{0\leq m_1\leq n_1\leq 2^{l_1} \\ n_1\leq m_2\leq n_2\leq 2^{l_2}}} |Q_{m_1,n_1,m_2,n_2}| \right)^2 \leq C \sum_{l_2=1}^{\infty} \sum_{l_1=1}^{l_2} \frac{l_1^2 l_2^2}{\log l_1 2^{l_2}} < \infty,$$

so we can conclude that the last summand is of the required order almost surely. □

Proposition 5.5. *Under the assumptions (A1), (A2) and (A3), the process*

$$\sqrt{n} \left(\lambda(1-\lambda)(U_{\lfloor \lambda n \rfloor, n-\lfloor \lambda n \rfloor}(t_p) - p \right)_{\lambda\in[0,1]}$$

converges weakly to

$$\left((1-\lambda)W_1(\lambda) + \lambda(W_2(1) - W_2(\lambda)) \right)_{\lambda\in[0,1]},$$

where $W = (W_1, W_2)$ is a two-dimensional Brownian motion with covariance structure

$$\mathrm{Cov}(W_i(\mu), W_j(\lambda)) = (\mu \wedge \lambda) \sum_{k\in\mathbb{Z}} E(h_i(X_0; Q(p)), h_j(X_k; Q(p))).$$

This is Theorem 2.4 of Dehling et al. (2013).

Lemma 5.6. *Under the assumptions (A1), (A2) and (A3) for any bandwidth $b = b_n$ with $b + \frac{1}{nb^3} = o(1)$*

$$\hat{u}(0) = \frac{2}{n(n-1)b} \sum_{1\leq i<j\leq n} K\left(\frac{X_i - X_j}{b}\right) \to u(0)$$

in probability.

Proof. First note that $\hat{u}$ is a one-sample U-statistic with symmetric kernel $k_n(x,y) = \frac{1}{b}K(\frac{x-y}{b})$ depending on n. We use the Hoeffding decomposition

$$\begin{aligned}\tilde{u}_n &= Ek_n(X,Y),\\ k_{1,n}(x) &= Ek_n(x,X) - \tilde{u}_n,\\ k_{2,n}(x,y) &= k_n(x,y) - k_{1,n}(x) - k_{1,n}(y) - \tilde{u}_n,\end{aligned}$$

where X, Y are independent with the same distribution as X_1. We obtain

$$\hat{u}(0) = \tilde{u}_n + \frac{2}{n}\sum_{i=1}^{n} k_{1,n}(X_i) + \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} k_{2,n}(X_i,Y_i)$$

By our assumptions, K has a bounded support, so let $K(x) = 0$ for $|x| > M$. Because the density u is continuos and K integrates to 1, we can conclude that

$$\tilde{u}_n - u(0) = \int \frac{1}{b}K(\frac{x}{b})u(x)dx - u(0) \leq \sup_{|x|\leq Mb} |u(x) - u(0)| \xrightarrow{n\to\infty} 0,$$

since $b_n \to 0$. As K is Lipschitz continous, i.e. $|K(x) - K(y)| \leq L_1|x-y|$, for some constant L_1, we have that $k_{1,n}(x)$ is Lipschitz continous with constant L_1/b^2. By Proposition 2.11 of Borovkova et al. [5], it follows that $(k_{1,n}(X_i))_{i\in\mathbb{N}}$ is near epoch dependent with approximation constants $a'_k = 3\sqrt{a_k}/b_n^2$. Let $C_1 = C/b$ be the upper bound of $k_{1,n}(X_i)$, then by Lemma 2.18 of Borovkova et al.

$$|E\left(k_{1,n}(X_i)k_{1,n}(X_j)\right)| \leq 4C_1 a'_{|i-j|/3} + 2C_1^2\beta(|i-j|/3) \leq C\frac{1}{b^3}(\sqrt{a_{|i-j|/3}} + \beta(|i-j|/3)),$$

so we obtain by stationarity that

$$\begin{aligned}E\Big(\frac{2}{n}\sum_{i=1}^{n} k_{1,n}(X_i)\Big)^2 &\leq \frac{4}{n}\sum_{i=1}^{\infty} |E\left(k_{1,n}(X_1)k_{1,n}(X_i)\right)|\\ &\leq C\frac{1}{nb^3}\sum_{i=1}^{\infty}(\sqrt{a_{|i-j|/3}} + \beta(|i-j|/3)) \to 0,\end{aligned}$$

because $nb^3 \to \infty$, so the second summand converges to 0. For the third summand, we use Lemma 4.3 of Borovkova et al. and the fact that $k_{2,n}(x,y)$ is a degenerate kernel bounded by $4C_1/b$ and that the product $k_{2,n}(x_1,x_2)k_{2,n}(x_3,x_4)$ is P-Lipschitz with constant $4(4\frac{C_1}{b}\frac{L_1}{b^2}) = Cb^{-3}$. We get the inequality

$$|E\left(k_{2,n}(X_{i_1},X_{i_2})k_{2,n}(X_{i_3},X_{i_4})\right)| \leq \frac{C}{b^2}\left(A_{m/3} + \beta(m/3)\right) + C\frac{1}{b^3}A_{m/3}$$

with $A_i = \sqrt{2\sum_{n=i}^{\infty} a_n}$ and $m = \max\left\{i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)}\right\}$, where $i_{(1)} \leq i_{(2)} \leq i_{(3)} \leq i_{(4)}$ are the order statistics of the indices i_1, i_2, i_3, i_4. Thus, we obtain

$$\begin{aligned}E\Big(\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n} k_{2,n}(X_i,Y_i)\Big)^2 &\leq C\frac{1}{n^4}\sum_{i_1,i_2,i_3,i_4=1}^{n} |E\left(k_{2,n}(X_{i_1},X_{i_2})k_{2,n}(X_{i_3},X_{i_4})\right)|\\ &= C\frac{1}{n^4}\sum_{m=0}^{n}\sum_{\substack{i_1,i_2,i_3,i_4\\ \max\{i_{(2)}-i_{(1)},i_{(4)}-i_{(3)}\}=m}} |E\left(k_{2,n}(X_{i_1},X_{i_2})k_{2,n}(X_{i_3},X_{i_4})\right)|\\ &\leq C\frac{1}{n^4b^3}\sum_{m=0}^{n}\sum_{\substack{i_1,i_2,i_3,i_4\\ \max\{i_{(2)}-i_{(1)},i_{(4)}-i_{(3)}\}=k}} \Big(A_{\frac{m}{3}} + \beta(\frac{m}{3})\Big).\end{aligned}$$

A this point, we have to calculate the number of quadrubles (i_1, i_2, i_3, i_4) such that $\max\{i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)}\} = m$. First note that there are at most 6 quadrubles wich lead to the same ordered numbers $i_{(1)}, i_{(2)}, i_{(3)}, i_{(4)}$. There are at most n^2 possibilities to choose $i_{(1)}$ and $i_{(4)}$. If $i_{(2)} - i_{(1)} = \max\{i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)}\} = m$, then $i_{(2)}$ is already fixed and there are k possibilities $i_{(3)}$. The same argument applies if $i_{(4)} - i_{(3)} = \max\{i_{(2)} - i_{(1)}, i_{(4)} - i_{(3)}\} = m$, so we finally obtain

$$E\Big(\frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} k_{2,n}(X_i, Y_i)\Big)^2 \leq C\frac{1}{n^2b^3} \sum_{m=0}^{n} m\Big(A_{\frac{m}{3}} + \beta(\frac{m}{3})\Big) \to 0,$$

as the $mA_{\frac{m}{3}}$ and $m\beta(\frac{m}{3})$ are summable by assumption (A1), and $n^2b^3 \to \infty$. □

Lemma 5.7. *Let G be a non-decreasing function, $c, l > 0$ constants and $[C_1, C_2] \subset \mathbb{R}$. If for all $t, t' \in [C_1, C_2]$ with $|t - t'| \leq l + 2c$*

$$|G(t) - G(t') - (t - t')| \leq c,$$

then for all $p, p' \in \mathbb{R}$ with $|p - p'| \leq l$ and $G^{-1}(p), G^{-1}(p') \in (C_1 + 2c + l, C_2 - 2c - l)$

$$|G^{-1}(p) - G^{-1}(p') - (p - p')| \leq c$$

where $G^{-1}(p) := \inf\{t \big| G(t) \geq p\}$ is the generalized inverse.

Proof. This is Lemma 3.5 of Wendler [35]. □

5.2. Proof of the Main Theorems.

Proof of Theorem 2.6. Without loss of generality, we can assume that $u(t_p) = 1$, otherwise replacing $h(x, y, t)$ by $h(x, y, \frac{t}{u(t_p)})$. We will first concentrate on the first half, that means we will investigate

$$\begin{aligned}
&\sup_{\lambda\in[0,\frac{1}{2}]} \lambda(1-\lambda)\left|U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(p) - t_p + U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p) - p\right| \\
&\leq \sup_{\lambda\in[0,\frac{1}{2}]} \lambda(1-\lambda)\left|U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(p) - U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p)) + U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p) - p\right| \\
&\qquad + \sup_{\lambda\in[0,\frac{1}{2}]} \lambda(1-\lambda)\left|U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p)) - t_p\right|
\end{aligned}$$

By Lemma 5.4, we can choose $C_1 > 0$, such that for all n

$$P\left(\sup_{\lambda\in[0,\frac{1}{2}]} \left|U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p) - p\right| / \sqrt{\frac{\log\log(\lambda n)}{\lambda n}} \geq C_1\right) \leq \epsilon.$$

Hence, using Lemma 5.3 and 5.7, there exists a constant C_2 such that

$$P\Big(\sup_{\lambda\in[0,\frac{1}{2}]}\lambda(1-\lambda)\big|U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(p)-U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p))+U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p)-p\big|>C_2n^{-5/9}\Big)$$

$$\leq P\left(\sup_{\substack{\lambda\in[0,\frac{1}{2}]\\ |p-p'|\leq C_1\sqrt{\frac{\log\log(\lambda n)}{\lambda n}}}}\lambda(1-\lambda)\left|U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(p)-U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(p')+p'-p\right|>C_2n^{-5/9}\right)$$

$$+P\left(\sup_{\lambda\in[0,\frac{1}{2}]}\left|U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p)-p\right|/\sqrt{\frac{\log\log(\lambda n)}{\lambda n}}\geq C_1\right)$$

$$\leq P\left(\sup_{\substack{\lambda\in[0,\frac{1}{2}]\\ |t-t_p|\leq C_1\sqrt{\frac{\log\log(\lambda n)}{\lambda n}}}}\lambda(1-\lambda)\left|U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t)-U(t)-U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p)+p\right|>C_2n^{-5/9}\right)+\epsilon$$

$$\leq 2\epsilon$$

Thus, the first summand is of order $n^{-5/9}$. It remains to show the convergence of the second summand $U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p))-t_p$. By the definition of the generalized inverse, $U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p))-t_p\leq 0$. Furthermore, if $U_{n_1,n_2}(t)<U_{n_1,n_2}(t_p)$ by the monotonicity of h, we have for all $n_2'\geq n_2$ that $U_{n_1,n_2'}(t)<U_{n_1,n_2'}(t_p)$. As $U^{-1}_{n_1,n_2}(U_{n_1,n_2}(t_p))$ is the supremum of all t such that $U_{n_1,n_2}(t)<U_{n_1,n_2}(t_p)$, it follows that $U^{-1}_{n_1,n_2}(U_{n_1,n_2}(t_p))$ is nondecreasing in n_2.

For every $c>0$ it holds that $(U^{-1}_{n_1,n_1}(U_{n_1,n_1}(t_p))-t_p)<-c$ only if $U_{n_1,n_1}(t_p-c)-U_{n_1,n_1}(t_p)\geq 0$, which is equivalent to

$$U_{n_1,n_1}(t_p-c)-U_{n_1,n_1}(t_p)-U(t_p-c)+p\geq -U(t_p-c)+p.$$

By Lemma 5.3, there a constant C_3 such that

$$P\left(\sup_{n_1\in\mathbb{N}}n_1^{\frac{5}{9}}\sup_{|t-t_p|\leq\sqrt{\frac{\log\log(n_1)}{n_1}}}|U_{n_1,n_1}(t)-U(t)-(U_{n_1,n_1}(t_p)-p)|>C_3\right)<\epsilon.$$

As U is differentiable, we have that $U(t_p-C_4n_1^{-\frac{5}{9}})+p>C_3n_1^{-\frac{5}{9}}$ for some constant C_4 and consequently for all $n_2\geq n_1$

$$U^{-1}_{n_1,n_2}(U_{n_1,n_2}(t_p))-t_p\geq -C_4n_1^{-\frac{5}{9}}.$$

Finally we have that $\lambda(1-\lambda)\lfloor\lambda n\rfloor^{-\frac{5}{9}}\leq n^{-\frac{5}{9}}$, and so we arrive at

$$P\left(\sup_{\lambda\in[0,\frac{1}{2}]}\lambda(1-\lambda)\left|U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p))-t_p\right|>C_4n^{-5/9}\right)$$

$$\leq P\left(\sup_{n_1\leq n/2}n_1^{5/9}\left|U^{-1}_{n_1,n_1}(U_{n_1,n_1}(t_p))-t_p\right|>C_4\right)$$

$$\leq P\left(\sup_{n_1\in\mathbb{N}}n_1^{5/9}\left|U_{n_1,n_1}(t_p-C_4n_1^{-5/9})-U((t_p-C_4n_1^{-5/9})-(U_{n_1,n_1}(t_p)-p)\right||>C_3\right)$$

$$\leq P\left(\sup_{n_1\in\mathbb{N}}n_1^{\frac{5}{9}}\sup_{|t-t_p|\leq\sqrt{\frac{\log\log(n_1)}{n_1}}}|U_{n_1,n_1}(t)-U(t)-(U_{n_1,n_1}(t_p)-p)|>C_3\right)<\epsilon.$$

So we have shown the convergence in probability for λ restricted to $[0, \frac{1}{2}]$. For the second half ($\lambda \in [1/2, 1]$), note that

$$\sup_{\lambda\in[\frac{1}{2},1]} \lambda(1-\lambda)\left|U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(p) - t_p + U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p) - p\right|$$
$$= \sup_{\lambda\in[0,\frac{1}{2}]} \lambda(1-\lambda)\left|\tilde{U}^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(p) - t_p + \tilde{U}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p) - p\right|,$$

where $\tilde{U}_{n_1,n_2}$ is the two sample U-statistics with kernel $\tilde{h}(x,y,t) = h(y,x,t)$ calculated for the stochastic process $(\tilde{X}_i)_{i\in\mathbb{Z}}$ with $\tilde{X}_i = X_{n-i}$. Because of the stationarity, the probability distribution of this does not change if we insert the random variables $\tilde{X}'_i = X_{-i}$ instead. The process $(X_{-i})_{i\in\mathbb{Z}}$ inherits the near epoch properties of $(X_i)_{i\in\mathbb{Z}}$. And with the same arguments as above, it follows that

$$\sup_{\lambda\in[0,\frac{1}{2}]} \lambda(1-\lambda)\left|\tilde{U}^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(p) - t_p + \tilde{U}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p) - p\right| = O_P(n^{-5/9})$$

□

Proof of Theorem 2.5. We decompose the stochastic process into two parts:

$$\sqrt{n}\left(\lambda(1-\lambda)(U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(p) - t_p)\right)_{\lambda\in[0,1]}$$
$$= \sqrt{n}\left(\lambda(1-\lambda)\frac{1}{u(t_p)}(p - U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p))\right)_{\lambda\in[0,1]}$$
$$+ \sqrt{n}\left(\lambda(1-\lambda)\left(U^{-1}_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(p) - t_p + \frac{U_{\lfloor\lambda n\rfloor,n-\lfloor\lambda n\rfloor}(t_p) - p}{u(t_p)}\right)\right)_{\lambda\in[0,1]}.$$

By Theorem 2.6, the second part converges to zero in supremums norm. As a consequence of Proposition 5.5, the first part converges weakly to

$$((1-\lambda)W_1(\lambda) + \lambda(W_2(1) - W_2(\lambda)))_{\lambda\in[0,1]},$$

where $W = (W_1, W_2)$ is a two-dimensional Brownian motion with covariance structure

$$\mathrm{Cov}(W_i(\mu), W_j(\lambda)) = (\mu\wedge\lambda)\frac{1}{u^2(Q(p))}\sum_{k\in\mathbb{Z}} E(h_i(X_0; Q(p)), h_j(X_k; Q(p))).$$

□

Proof of Theorem 1.1. By Theorem 2.5,

$$\lambda(1-\lambda)\,\mathrm{median}\left\{X_i - X_j \middle| 1 \le i \le \lfloor n\lambda\rfloor, \lfloor n\lambda\rfloor + 1 \le j \le n\right\}_{\lambda\in(0,1)}$$

converges to

$$((1-\lambda)W_1(\lambda) + \lambda(W_2(1) - W_2(\lambda)))_{\lambda\in[0,1]},$$

where $W = (W_1, -W_1)$ and W_1 is a Brownian motion, as $h_1(x,0) = -h_2(x,0)$. The variance is $\mathrm{Var}(W_1(1)) = \frac{\sigma^2}{u^2(0)}$. Now

$$\frac{u(0)}{\sigma}((1-\lambda)W_1(\lambda) + \lambda(-W_1(1) + W_1(\lambda))) = \frac{u(0)}{\sigma}W_1(\lambda) - \lambda\frac{u(0)}{\sigma}W_1(1)$$

is a Brownian Bridge. Finally, by Lemma 5.6 and Theorem 1.2 of Dehling et al. [15], $\frac{\hat{u}}{\hat{\sigma}} \to \frac{u(0)}{\sigma}$ in probability, which completes the proof. □

Acknowledgement

The research was supported by the DFG Collaborative Research Center 823 *Statistical Modelling of Nonlinear Dynamic Processes*.

References

[1] D.W.K. Andrews (1984): Non-Strong Mixing Autoregressive Process, *Journal of Applied Probability* **21**, 930–934.
[2] M.A. Arcones (1996): The Bahadur-Kiefer representation for U-quantiles, *Annals of Statistics* **24**, 1400–1422.
[3] R.R. Bahadur (1966): A note on quantiles in large samples, *Annals of Mathematical Statistics* **37**, 577–580.
[4] P. Billingsley (1968): *Convergence of Probability Measures*, J. Wiley, New York.
[5] S. Borovkova, R. Burton and H. Dehling (2001): Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation, *Transactions of the American Mathematical Society* **353**, 4261–4318.
[6] R.C. Bradley (2007): *Introduction to Strong Mixing Conditions*, vol. 1–3, Kendrick Press, Heber City.
[7] E. Carlstein (1986): The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics* **14**, 1171–1179.
[8] J. Choudhury and R.J. Serfling (1988) Generalized order statistics, Bahadur representations, and sequential nonparametric fixed-width confidence intervals, *J. Statist. Plann. Inference* **19**, 269–282.
[9] M. Csörgő and L. Horváth (1988): Invariance principles for changepoint problems, *J. Multivariate Anal.* **27**, 151–168.
[10] M. Csörgő and L. Horváth (1997): *Limit theorems in change point analysis*, J. Wiley, Chichester.
[11] J. Dedecker, P. Doukhan, G. Lang, J. Leon, S. Louhichi and C. Prieur (2007): *Weak Dependence: With Examples and Applications*, Springer, New York.
[12] H. Dehling, M. Denker and W. Philipp (1987): The almost sure invariance principle for the empirical process of U-statistic structure, *Annales de l'I.H.P.* **23**, 121–134.
[13] H. Dehling and R. Fried (2012): Asymptotic distribution of two-sample empirical U-quantiles with applications to robust tests for shifts in location, *J. Multivariate Anal.* **105**, 124–140.
[14] H. Dehling, R. Fried, I. García, M. Wendler (2013): Change-Point Detection under Dependence Based on Two-Sample U-Statistics, *preprint* arXiv:1304.2479.
[15] H. Dehling, R. Fried, D. Vogel, O. Sh. Sharipov and M. Wornowizki (2013): Estimation of the variance of partial sums of dependent processes, *Statistics and Probability Letters* **83**, 141–147.
[16] H. Dehling and M. Wendler (2010): Central limit theorem and the bootstrap for U-statistics of strongly mixing data, *J. Multivariate Anal.* **101**, 126–137.
[17] M. Denker and G. Keller (1986): Rigorous statistical procedures for data from dynamical systems, *J. Stat. Phys.* **44**, 67–93.
[18] P. Doukhan (1994): *Mixing: Properties and Examples*, Springer, New York.
[19] R. Fried and H. Dehling (2011): Robust nonparametric tests for the two-sample location problem, *Stat. Methods Appl.* **20**, 409–422.
[20] J.C. Geertsema (1970): Sequential confidence intervals based on rank test, *Ann. Math. Stat.* **41**, 1016–1026.
[21] B.E. Hansen (1991): GARCH(1,1) processes are near epoch dependent, *Econom. Lett.* **36**, 181–186.
[22] J.L. Hodges and E.L. Lehmann (1963): Estimates of location based on rank tests, *Ann. Math. Stat.* **34**, 598–611.
[23] J.L. Hodges and E.L. Lehmann (1956): The efficiency of some nonparametric competitors of the t-Test, *Ann. Math. Stat.* **27**, 324–335.
[24] W. Hoeffding (1948): A class of statistics with asymptotically normal distribution, *Ann. Math. Stat.* **19**, 293–325.
[25] F. Hofbauer and G. Keller (1982): Ergodic properties of invariant measures for piecewise monotonic transformations, *Math. Z.* **180**, 119–142.
[26] I.A. Ibragimov and Y.V. Linnik (1961): *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.
[27] N. Inagaki (1973): The asymptotic representation of the Hodges-Lehmann estimator based on Wilcoxon two-sample statistic, *Ann. Inst. Statist. Math.* **25**, 457–466.

[28] E.L. Lehmann (1951): Consistency and unbiasedness of certain nonparametric statistics, *Ann. Math. Stat.* **22**, 165–179.
[29] E.L. Lehmann (1963): Nonparametric confidence intervals for a shift parameter, *Ann. Math. Stat.* **34**, 1507–1512.
[30] E. Rio (2000): *Théorie asymptotique des processus aléatoires faiblement dépendants*, Springer Verlag, Paris.
[31] M. Rosenblatt (1956): A Central Limit Theorem and a Strong Mixing Condition, *Proceedings of the National Academy of Sciences* **42**, 43–47.
[32] R.J. Serfling (1968): The Wilcoxon two-sampe statistic on strongly mixing processes, *Ann. Math. Stat.* **39**, 1202–1209.
[33] R.J. Serfling (1984): Generalized L-, M-, and R-Statistics, *Ann. Prob.* **12**, 76–86.
[34] M. Wendler (2011): Bahadur representation for U-quantiles of dependent data, *J. Multivariate Anal.* **102**, 1064–1079.
[35] M Wendler (2012): U-processes, U-quantile-processes and gerneralized linear statistics of dependent data, *Stochastic Process. Appl.* **122**, 787–807.
[36] K. Yoshihara (1976): Limiting behavior of U-statistics for stationary, absolutely regular processes, *Probab. Theory Related Fields* **35**, 237–252.

Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany
E-mail address: herold.dehling@rub.de

Fakultät Statistik, TU Dortmund, 44221 Dortmund, Germany
E-mail address: fried@statistik.tu-dortmund.de

Institut für Mathematik und Informatik, Ernst-Moritz-Arndt-Universität Greifswald, 17487 Greifswald, Germany
E-mail address: martin.wendler@uni-greifswald.de