

Change-Point Detection Under Dependence Based on Two-Sample U -Statistics

Herold Dehling, Roland Fried, Isabel Garcia, and Martin Wendler

1 Introduction

Change-point tests address the question whether a stochastic process is stationary during the entire observation period or not. In the case of independent data, there is a well-developed theory; see the book by Csörgő and Horváth [6] for an excellent survey. When the data are dependent, much less is known. The CUSUM statistic has been intensely studied, even for dependent data; see again Csörgő and Horváth [6]. The CUSUM test, however, is not robust against outliers in the data. In the present paper, we study a robust test which is based on the two-sample Wilcoxon test statistic. Simulations show that this test outperforms the CUSUM test in the case of heavy-tailed data.

In order to derive the asymptotic distribution of the test, we study the stochastic process

$$\sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n h(X_i, X_j), \quad 0 \leq \lambda \leq 1, \quad (1)$$

H. Dehling (✉) • I. Garcia
Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany
e-mail: herold.dehling@rub.de; isabel.garciaarboleda@rub.de

R. Fried
Fakultät für Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany
e-mail: fried@statistik.tu-dortmund.de

M. Wendler
Institut für Mathematik und Informatik, Ernst-Moritz-Arndt-Universität,
17487 Greifswald, Germany
e-mail: martin.wendler@rub.de

where $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a kernel function. In the case of independent data, the asymptotic distribution of this process has been studied by Csörgő and Horváth [5]. In the present paper, we extend their result to short range dependent data $(X_i)_{i \geq 1}$. Similar results have been obtained for long range dependent data by Dehling, Rooch and Taqqu [10], albeit with completely different methods.

U -statistics have been introduced by Hoeffding [14], where the asymptotic normality was established both for the one-sample as well as the two-sample U -statistic in the case of independent data. The asymptotic distribution of one-sample U -statistics of dependent data was studied by Sen [18, 19], Yoshihara [22], Denker and Keller [12, 13] and by Borovkova, Burton and Dehling [3] in the so-called non-degenerate case, and by Babbel [1] and Leucht [16] in the degenerate case. For two-sample U -statistics, Dehling and Fried [8] established the asymptotic normality of $\sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n_1+n_2} h(X_i, X_j)$ for dependent data, when $n_1, n_2 \rightarrow \infty$. The main theoretical result of the present paper is a functional version of this limit theorem.

In our paper, we focus on data that can be represented as functionals of a mixing process. In this way, we cover most examples from time series analysis, such as ARMA and ARCH processes, but also data from chaotic dynamical systems. For a survey of processes that have a representation as functional of a mixing process, see e.g. Borovkova, Burton and Dehling [3]. Earlier references can be found in Ibragimov and Linnik [15], Denker [11] and Billingsley [2].

2 Definitions and Main Results

Given the samples X_1, \dots, X_{n_1} and Y_1, \dots, Y_{n_2} , and a kernel $h(x, y)$, we define the two-sample U -statistic

$$U_{n_1, n_2} := \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} h(X_i, Y_j). \quad (2)$$

More generally, one can define U -statistics with multivariate kernels $h : \mathbb{R}^k \times \mathbb{R}^l \rightarrow \mathbb{R}$. In the present paper, for the ease of exposition, we will restrict attention to bivariate kernels $h(x, y)$. The main results, however, can easily be extended to the multivariate case.

Assuming that $(X_i)_{i \geq 1}$ and $(Y_i)_{i \geq 1}$ are stationary processes with one-dimensional marginal distribution functions F and G , respectively, we can test the hypothesis $H : F = G$ using the two-sample U -statistic. E.g., the kernel $h(x, y) = y - x$ leads to the U -statistic

$$U_{n_1, n_2} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (Y_j - X_i) = \frac{1}{n_2} \sum_{j=1}^{n_2} Y_j - \frac{1}{n_1} \sum_{i=1}^{n_1} X_i, \quad (3)$$

and thus to the familiar two-sample Gauß-test. Similarly, the kernel $h(x, y) = 1_{\{x \leq y\}}$ leads to the U -statistic

$$U_{n_1, n_2} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} 1_{\{X_i \leq X_j\}}, \tag{4}$$

and thus to the 2-sample Mann-Whitney-Wilcoxon test.

In the present paper, we investigate tests for a change-point in the mean of a stochastic process $(X_i)_{i \geq 1}$. We consider the model

$$X_i = \mu_i + \xi_i, \quad i \geq 1, \tag{5}$$

where $(\mu_i)_{i \geq 1}$ are unknown constants and where $(\xi_i)_{i \geq 1}$ is a stochastic process. We want to test the hypothesis $H : \mu_1 = \dots = \mu_n$ against the alternative that there exists $1 \leq k \leq n - 1$ such that $\mu_1 = \dots = \mu_k \neq \mu_{k+1} = \dots = \mu_n$.

Tests for the change-point problem are often derived from 2-sample tests applied to the samples X_1, \dots, X_k and X_{k+1}, \dots, X_n , for all possible $1 \leq k \leq n - 1$. For two-sample tests based on U -statistics with kernel $h(x, y)$, this leads to the test statistic $\sum_{i=1}^k \sum_{j=k+1}^n h(X_i, X_j)$, $1 \leq k \leq n$, and thus to the process

$$U_n(\lambda) = \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n h(X_i, X_j), \quad 0 \leq \lambda \leq 1. \tag{6}$$

In this paper, we will derive a functional limit theorem for the process $(U_n(\lambda))_{0 \leq \lambda \leq 1}$, $n \geq 1$. Specifically, we will show that under certain technical assumptions on the kernel h and on the process $(X_i)_{i \geq 1}$, a properly centered and renormalized version of $(U_n(\lambda))_{0 \leq \lambda \leq 1}$ converges to a Gaussian process.

In our paper, we will assume that the process $(\xi_i)_{i \geq 0}$ is weakly dependent. More specifically, we will assume that $(\xi_i)_{i \geq 0}$ can be represented as a functional of an absolutely regular process.

Definition 1. (i) Given a stochastic process $(X_n)_{n \in \mathbb{Z}}$ we denote by \mathcal{A}_1^k the σ -algebra generated by (X_k, \dots, X_1) . The process is called absolutely regular if

$$\beta(k) = \sup_n \left\{ \sup_{j=1}^J \sum_{i=1}^I |P(A_i \cap B_j) - P(A_i)P(B_j)| \right\} \rightarrow 0, \tag{7}$$

as $k \rightarrow \infty$, where the last supremum is over all finite \mathcal{A}_1^n -measurable partitions (A_1, \dots, A_I) and all finite \mathcal{A}_{n+k}^∞ -measurable partitions (B_1, \dots, B_J) .

(ii) The process $(X_n)_{n \geq 1}$ is called a two-sided functional of an absolutely regular sequence if there exists an absolutely regular process $(Z_n)_{n \in \mathbb{Z}}$ and a measurable function $f : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}$ such that

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

Analogously, $(X_n)_{n \geq 1}$ is called a one-sided functional if $X_i = f((Z_{i+n})_{n \geq 0})$.

(iii) The process $(X_n)_{n \geq 1}$ is called 1-approximating functional with coefficients $(a_k)_{k \geq 1}$ if

$$E |X_i - E(X_i | Z_{i-k}, \dots, Z_{i+k})| \leq a_k. \tag{8}$$

In addition to weak dependence conditions on the process $(X_i)_{i \geq 1}$, the asymptotic analysis of the process (6) requires some continuity assumptions on the kernel functions $h(x, y)$. We use the notion of 1-continuity, which was introduced by Borovkova, Burton and Dehling [3]. Alternative continuity conditions have been used by Denker and Keller [13].

Definition 2. The kernel $h(x, y)$ is called 1-continuous, if there exists a function $\phi : (0, \infty) \rightarrow (0, \infty)$ with $\phi(\epsilon) = o(1)$ as $\epsilon \rightarrow 0$ such that for all $\epsilon > 0$

$$E(|h(X', Y) - h(X, Y)| 1_{\{|X-X'| \leq \epsilon\}}) \leq \phi(\epsilon) \tag{9}$$

$$E(|h(X, Y') - h(X, Y)| 1_{\{|Y-Y'| \leq \epsilon\}}) \leq \phi(\epsilon) \tag{10}$$

for all random variables X, X', Y and Y' having the same marginal distribution as X_1 , and such that X, Y are either independent or have joint distribution $P_{(X_1, X_k)}$, for some integer k .

The most important technical tool in the study of U -statistics is Hoeffding's decomposition, originally introduced by Hoeffding [14]. If $E|h(X, Y)| < \infty$ for two independent random variables X and Y with the same distribution as X_1 , we can write

$$h(x, y) = \theta + h_1(x) + h_2(y) + g(x, y), \tag{11}$$

where the terms on the right-hand side are defined as follows:

$$\theta = \int \int h(x, y) dF(x) dF(y)$$

$$h_1(x) = \int h(x, y) dF(y) - \theta$$

$$h_2(y) = \int h(x, y) dF(x) - \theta$$

$$g(x, y) = h(x, y) - h_1(x) - h_2(y) - \theta.$$

Here, F denotes the distribution function of the random variables X_i . Observe that, by Fubini's theorem,

$$E(h_1(X)) = E(h_2(X)) = 0.$$

In addition, the kernel $g(x, y)$ is degenerate in the sense of the following definition.

Definition 3. Let $(X_i)_{i \geq 1}$ be a stationary process, and let $g(x, y)$ be a measurable function. We say that $g(x, y)$ is degenerate if

$$E(g(x, X_1)) = E(g(X_1, y)) = 0, \tag{12}$$

for all $x, y \in \mathbb{R}$.

The following theorem, a functional central limit theorem for two-sample U -statistics of dependent data, is the main theoretical result of the present paper.

Theorem 1. Let $(X_n)_{n \geq 1}$ be a 1-approximating functional with constants $(a_k)_{k \geq 1}$ of an absolutely regular process with mixing coefficients $(\beta(k))_{k \geq 1}$, and let $h(x, y)$ be a 1-continuous bounded kernel, satisfying

$$\sum_{k=1}^{\infty} k^2(\beta(k) + \sqrt{a_k} + \phi(a_k)) < \infty, \tag{13}$$

Then, as $n \rightarrow \infty$, the $D[0, 1]$ -valued process

$$T_n(\lambda) := \frac{1}{n^{3/2}} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n (h(X_i, X_j) - \theta), \quad 0 \leq \lambda \leq 1, \tag{14}$$

converges in distribution towards a mean-zero Gaussian process with representation

$$Z(\lambda) = (1 - \lambda)W_1(\lambda) + \lambda(W_2(1) - W_2(\lambda)), \quad 0 \leq \lambda \leq 1, \tag{15}$$

where $(W_1(\lambda), W_2(\lambda))_{0 \leq \lambda \leq 1}$ is a two-dimensional Brownian motion with mean zero and covariance function $\text{Cov}(W_k(s), W_l(t)) = \min(s, t)\sigma_{kl}$, where

$$\sigma_{kl} = E(h_k(X_0)h_l(X_0)) + 2 \sum_{j=1}^{\infty} \text{Cov}(h_k(X_0), h_l(X_j)), \quad k, l = 1, 2. \tag{16}$$

Remark 1. (i) In the case of i.i.d. data, Theorem 1 was established by Csörgő and Horváth [5]. In the case of long-range dependent data, weak convergence of the process $(T_n(\lambda))_{0 \leq \lambda \leq 1}$ has been studied by Dehling, Rooch and Taqqu [10] and by Rooch [17], albeit with a normalization different from $n^{3/2}$.

(ii) Using the representation (15), one can calculate the autocovariance function of the process $(Z(\lambda))_{0 \leq \lambda \leq 1}$. We obtain

$$\begin{aligned} \text{Cov}(Z(\lambda), Z(\mu)) &= \sigma_{11}[(1 - \lambda)(1 - \mu) \min\{\lambda, \mu\}] \\ &\quad + \sigma_{22}[\lambda\mu(1 - \mu - \lambda + \min\{\lambda, \mu\})] \\ &\quad + \sigma_{12}[\mu(1 - \lambda)(\lambda - \min\{\lambda, \mu\}) + \lambda(1 - \mu)(\mu - \min\{\lambda, \mu\})]. \end{aligned}$$

- (iii) We conjecture that a similar theorem also holds for unbounded kernels under some moments conditions and faster mixing rates (similar to Theorem 2.7 of Sharipov, Wendler [20]). As our main application is the Wilcoxon test, where the kernel is bounded, we restrict the theorem to the case of bounded kernels.
- (iv) For the kernel $h(x, y) = y - x$, we can analyze the asymptotic behavior of the process $T_n(\lambda)$ using the functional central limit theorem (FCLT). Note that, since $X_j - X_i = (X_j - E(X_j)) - (X_i - E(X_i))$, we may assume without loss of generality that X_i has mean zero. Then we get the representation

$$\begin{aligned} T_n(\lambda) &= \frac{1}{n^{3/2}} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n (X_j - X_i) \\ &= \frac{[n\lambda]}{n} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i - \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\lambda]} X_i. \end{aligned} \quad (17)$$

Thus, weak convergence of $(T_n(\lambda))_{0 \leq \lambda \leq 1}$ can be derived from the FCLT for the partial sum process $\frac{1}{\sqrt{n}} \sum_{i=1}^{[n\lambda]} X_i$. Such FCLTs have been proved under a wide range of conditions, e.g. for functionals of uniformly mixing data in Billingsley [2].

We finally want to state an important special case of Theorem 1, namely when the kernel is anti-symmetric, i.e. when $h(x, y) = -h(y, x)$. Kernels that occur in connection with change-point tests usually have this property. For anti-symmetric kernels, the limit process has a much simpler structure; moreover one can give a simpler direct proof in this case. Note that for independent random variables X, Y we have by anti-symmetry that $Eh(X, Y) = -Eh(Y, X) = -Eh(X, Y)$ and so $\theta = Eh(X, Y) = 0$.

Theorem 2. *Let $(X_n)_{n \geq 1}$ be a 1-approximating functional with constants $(a_k)_{k \geq 1}$ of an absolutely regular process with mixing coefficients $(\beta(k))_{k \geq 1}$, and let $h(x, y)$ be a 1-continuous bounded anti-symmetric kernel, such that (13) holds. Then, as $n \rightarrow \infty$, the $D[0, 1]$ -valued process*

$$T_n(\lambda) := \frac{1}{n^{3/2}} \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n h(X_i, X_j), \quad 0 \leq \lambda \leq 1, \quad (18)$$

converges in distribution towards the mean-zero Gaussian process $\sigma W^{(0)}(\lambda)$, $0 \leq \lambda \leq 1$, where $(W^0(\lambda))_{0 \leq \lambda \leq 1}$ is a standard Brownian bridge and

$$\sigma^2 = \text{Var}(h_1(X_1)) + 2 \sum_{i=2}^{\infty} \text{Cov}(h_1(X_1), h_1(X_i)). \quad (19)$$

3 Application to Change Point Problems

In this section, we will apply Theorem 1 in order to derive the asymptotic distribution of two change-point test statistics. Specifically, we wish to test the hypothesis

$$H_0 : \mu_1 = \dots = \mu_n \tag{20}$$

against the alternative of a level shift at an unknown point in time, i.e.

$$H_A : \mu_1 = \dots = \mu_k \neq \mu_{k+1} = \dots = \mu_n, \text{ for some } k \in \{1, \dots, n - 1\}. \tag{21}$$

We consider the following two test statistics,

$$T_{1,n} = \max_{1 \leq k < n} \left| \frac{1}{n^{3/2}} \sum_{i=1}^k \sum_{j=k+1}^n (1_{\{X_i < X_j\}} - 1/2) \right| \tag{22}$$

$$T_{2,n} = \max_{1 \leq k < n} \left| \frac{1}{n^{3/2}} \sum_{i=1}^k \sum_{j=k+1}^n (X_j - X_i) \right|. \tag{23}$$

Theorem 3. *Let $(X_n)_{n \geq 1}$ be a 1-approximating functional with constants $(a_k)_{k \geq 1}$ of an absolutely regular process with mixing coefficients $(\beta(k))_{k \geq 1}$, satisfying (13), and assume that X_1 has a distribution function $F(x)$ with bounded density. Then, under the null hypothesis H_0 ,*

$$T_{1,n} \rightarrow \sigma_1 \sup_{0 \leq \lambda \leq 1} |W^{(0)}(\lambda)|, \tag{24}$$

where $(W^{(0)}(\lambda))_{0 \leq \lambda \leq 1}$ denotes the standard Brownian bridge process, and where

$$\sigma_1^2 = \text{Var}(F(X_1)) + 2 \sum_{k=2}^{\infty} \text{Cov}(F(X_1), F(X_k)). \tag{25}$$

Assuming that $E|X_1|^{2+\delta} < \infty$, $\beta(k) = O(k^{-(2+\delta)/\delta})$ and $a_k = O(k^{-(1+\delta)/2\delta})$, and under the null hypothesis H_0 ,

$$T_{2,n} \rightarrow \sigma_2 \sup_{0 \leq \lambda \leq 1} |W^{(0)}(\lambda)|, \tag{26}$$

where

$$\sigma_2^2 = \text{Var}(X_1) + 2 \sum_{k=2}^{\infty} \text{Cov}(X_1, X_k). \tag{27}$$

Proof. We will establish weak convergence of $T_{1,n}$. In order to do so, we will apply Theorem 1 to the kernel $h(x, y) = 1_{\{x < y\}}$. Borovkova, Burton and Dehling [3] showed that this kernel is 1-continuous. By continuity of the distribution function of X_1 , we get that $\theta = \iint 1_{\{x < y\}} dF(x) dF(y) = 1/2$. Moreover, we get

$$h_1(x) = P(x < X_1) - \frac{1}{2} = \frac{1}{2} - F(x)$$

$$h_2(x) = P(X_1 < x) - \frac{1}{2} = F(x) - \frac{1}{2}.$$

Note that $h_2(x) = -h_1(x)$. Hence $W_2(\lambda) = -W_1(\lambda)$, and thus the limit process in Theorem 1 has the representation

$$Z(\lambda) = (1 - \lambda)W_1(\lambda) + \lambda(W_2(1) - W_2(\lambda)) = W_1(\lambda) - \lambda W_1(1).$$

Here $W_1(\lambda)$ is a Brownian motion with variance σ_1^2 . Weak convergence of $T_{2,n}$ can be shown directly from the functional central limit theorem for the partial sum process; see Corollary 3.2 of Wooldridge and White [21]. We have to check the L_2 -near epoch dependence. Note that by our assumptions

$$\begin{aligned} & E |X_0 - E[X_0|Z_{-l}, \dots, Z_l]|^2 \\ &= E \left[|X_0 - E[X_0|Z_{-l}, \dots, Z_l]|^2 1_{\{|X_0 - E[X_0|Z_{-l}, \dots, Z_l]| \leq a_l^{-\frac{1}{1+\delta}}\}} \right] \\ &\quad + E \left[|X_0 - E[X_0|Z_{-l}, \dots, Z_l]|^2 1_{\{|X_0 - E[X_0|Z_{-l}, \dots, Z_l]| > a_l^{-\frac{1}{1+\delta}}\}} \right] \\ &\leq a_l^{-\frac{1}{1+\delta}} E |X_0 - E[X_0|Z_{-l}, \dots, Z_l]| + a_l^{\frac{\delta}{1+\delta}} E |X_0 - E[X_0|Z_{-l}, \dots, Z_l]|^{2+\delta} \\ &\leq C a_l^{\frac{\delta}{1+\delta}} = O(l^{-1/2}), \end{aligned} \tag{28}$$

so the condition of Corollary 3.2 of Wooldridge and White [21] holds. Hence, the partial sum process $(\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i)_{0 \leq t \leq 1}$ converges in distribution to $(\sigma_2 W(t))_{0 \leq t \leq 1}$, where W is standard Brownian motion. Convergence in distribution of $T_{2,n}$ follows by an application of the continuous mapping theorem.

Remark 2. (i) The distribution of $\sup_{0 \leq \lambda \leq 1} |W(\lambda)|$ is the well-known Kolmogorov-Smirnov distribution. Quantiles of the Kolmogorov-Smirnov distribution can be found in most statistical tables.

(ii) In order to apply Theorem 3, we need to estimate the variances σ_1^2 and σ_2^2 . Regarding σ_2^2 given in expression (27), we apply the non-overlapping subsampling estimator

$$\hat{\sigma}_2^2 = \frac{1}{[n/l_n]} \sum_{i=1}^{[n/l_n]} \frac{1}{l_n} \left(\sum_{j=(i-1)l_n+1}^{il_n} X_j - \frac{l_n}{n} \sum_{j=1}^n X_j \right)^2 \tag{29}$$

investigated by Carlstein [4] for α -mixing data. In case of AR(1)-processes, Carlstein derives

$$l_n = \max(\lceil n^{1/3}(2\rho/(1-\rho^2))^{2/3} \rceil, 1) \tag{30}$$

as the choice of the block length which minimizes the MSE asymptotically, with ρ being the autocorrelation coefficient at lag 1.

Regarding σ_1^2 given in (25), one faces the additional challenge that the distribution function F is unknown. This problem has been addressed, e.g. in Dehling, Fried, Sharipov, Vogel and Wornowizki [9], for the case of functionals of absolutely regular processes and F being estimated by the empirical distribution function F_n . The authors find the subsampling estimator for σ_1^2

$$\hat{\sigma}_1 = \frac{1}{[n/l_n]} \sqrt{\frac{\pi}{2}} \sum_{i=1}^{[n/l_n]} \frac{1}{\sqrt{l_n}} \left| \sum_{j=(i-1)l_n+1}^{il_n} F_n(X_j) - \frac{l_n}{n} \sum_{j=1}^n F_n(X_j) \right|, \tag{31}$$

employing non-overlapping subsampling, to give smaller biases, but somewhat larger MSEs than the corresponding overlapping subsampling estimator. The adaptive choice of the block length l_n proposed by Carlstein worked well in their simulations if the data were generated from a stationary ARMA(1,1) model and an estimate of ρ was plugged in. In the next section, we will explore this and other proposals in situations with level shifts and normally or heavy-tailed innovations.

4 Simulation Results

The assumptions regarding the underlying process (X_i) in Theorem 1 are satisfied by a wide range of time series, such as AR and ARMA processes. To illustrate the results and to investigate the finite sample behavior and the power of the tests based on $T_{1,n}$ and $T_{2,n}$, we will give some simulation results. We study the underlying change-point model

$$X_i = \begin{cases} \xi_i & \text{if } i = 1, \dots, [n\lambda] \\ \mu + \xi_i & \text{if } i = [n\lambda] + 1, \dots, n. \end{cases} \tag{32}$$

Within this model, the hypothesis of no change is equivalent to $\mu = 0$. We assume that the noise follows an AR(1) process, i.e. that

$$\xi_i = \rho \xi_{i-1} + \epsilon_i, \tag{33}$$

Table 1 Empirical level of the tests based on $T_{1,n}$ and $T_{2,n}$, for $n = 200$, with fixed or adaptive subsampling block length l_n and overlapping (ol) or non-overlapping (nol) subsampling. The results are for AR(1) observations with different lag-one autocorrelations ρ and different t_3 -distributed innovations, and based on 4000 simulation runs each

ν	ρ	$T_{1,n}$						$T_{2,n}$					
		Unadj.	l_n fixed		Adaptive		Unadj.	l_n fixed		Adaptive			
			ol	nol	ol	nol		ol	nol	ol	nol		
∞	0.0	2.8	2.0	2.9	2.0	2.2	4.5	2.9	3.9	3.7	3.8		
∞	0.4	24.5	2.5	3.1	3.5	3.9	34.2	3.9	4.9	5.5	6.0		
∞	0.8	81.6	6.2	6.5	1.9	2.5	91.5	10.5	10.6	3.4	4.0		
3	0.0	3.1	2.2	2.9	2.2	2.9	3.8	2.5	3.5	3.1	3.1		
3	0.4	26.9	2.4	3.0	3.2	3.0	32.0	3.3	3.8	4.3	4.9		
3	0.8	82.7	6.9	7.0	2.0	2.8	90.6	10.2	10.5	3.2	3.9		

where $-1 < \rho < 1$, and where the innovations ϵ_i are i.i.d. random variables with mean zero, bounded density and finite second moments. The innovations ϵ_i are generated from a standard normal or a t_ν -distribution with $\nu = 3$ degrees of freedom, scaled to have the same 84.13 % percentile as the standard normal, which is 1. The autoregression coefficient is varied in $\rho = \{0.0, 0.4, 0.8\}$, corresponding to zero, moderate or strong positive autocorrelation, and the sample size is $n = 200$. For the choice of the block length we used Carlstein’s adaptive rule outlined above, or a fixed block length of $l_n = 9$, which is in good agreement with the empirical findings of Dehling et al. [10] for larger sample sizes, and their theoretical result that l_n should be chosen as $o(\sqrt{n})$ to achieve consistency. For comparison, we also include tests employing overlapping subsampling for estimation of the asymptotical variance, applying the same block lengths as the non-overlapping versions.

Table 1 contains the empirical levels (i.e. the fraction of rejections) of the tests with an asymptotic level of 5 %, obtained from 4000 simulation runs for each situation. Note that the tests developed under the assumption of independence, not adjusting for autocorrelation, become strongly oversized with an increasingly positive autocorrelation, i.e. they reject a true null hypothesis far too often, and are practically useless already for $\rho = 0.4$. The performance of the adjusted tests is much better in this respect and in a good agreement with the asymptotic results. Only if the autocorrelation is strong ($\rho = 0.8$), the tests with a fixed block length become somewhat anti-conservative (oversized), and even more so for the CUSUM-test. Longer block lengths are needed for stronger positive autocorrelations, and Carlstein’s adaptive block length (30) adjusts for this. There is little difference between the tests employing overlapping and non-overlapping subsampling here.

In order to investigate the powers of the tests under the alternative, we consider shifts of increasing height μ , generating 400 data sets for each situation. The sample size is again $n = 200$, and the change point is at observation number $\tau = \lfloor \lambda n \rfloor = 100$.

Figure 1 illustrates the powers of the different versions of the tests in case of Gaussian or t_3 -distributed innovations and several autocorrelation coefficients ρ . Under normality, the CUSUM test $T_{2,n}$ is somewhat more powerful than the test $T_{1,n}$ based on the Wilcoxon statistic, while under the t_3 -distribution it is the other way round. The CUSUM test with the fixed block length considered here becomes strongly oversized if ρ is large, while this effect is less severe for the test based on the Wilcoxon statistic. Carlstein's adaptive choice of the block length increases the power if ρ is small and improves the size of the test substantially if ρ is large. The tests employing overlapping subsampling (not shown here) perform even slightly more powerful in case of zero or moderate autocorrelations, but much less powerful in case of strong autocorrelations. We have also considered the case of negative autocorrelation ($\rho = -0.4$, not shown here). We obtained similar results for the power of the test based on the Wilcoxon statistic relatively to that of the CUSUM test, and little difference between using a fixed or the adaptive block length.

The tests with Carlstein's adaptive choice of the block length could be improved further by using a more sophisticated estimate of ρ than the ordinary sample autocorrelation used here. The latter is positively biased in the presence of a shift, which leads to too 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 l_n increases more rapidly if $\hat{\rho}$ is close to 1, while it is rather stable for moderate and small values of $\hat{\rho}$. In our study, for $\rho = 0$ the average value chosen for l_n increases from about 2 to about 3, only, as the height of the shift increases, while it increases from about 6 to about 9 if $\rho = 0.4$, and even from about 16 to about 24 if $\rho = 0.8$. An estimate of the autocorrelation coefficient which resists shifts could be used, e.g. by applying a stepwise procedure which estimates the possible time of occurrence of a shift before calculating $\hat{\rho}$ from the corrected data, but this will not be pursued here.

5 Data Example

For illustration we apply the tests to time series data representing the monthly average daily minimum temperatures in Potsdam, Germany, measured between January 1893 and December 1992. The 1200 data points for these 100 years have been deseasonalized by subtracting the median value from each calendar month, see Fig. 2. Our interest is in whether the level of this time series is constant or whether there is a monotonic change. Such a systematic change is likely to show a trend-like behavior and not a sharp shift, but nevertheless we would like a change-point test to detect such a change if its null hypothesis is a constant level.

The empirical autocorrelation and partial autocorrelation functions suggest a first order autoregressive model with lag-one autocorrelation about 0.25 for the deseasonalized data. The test statistics take their maximum values after time point 595, i.e. rather in the middle of the time series. The resulting p-values are 0.23 and 0.16 for the CUSUM test with the fixed and the adaptive block length, respectively.

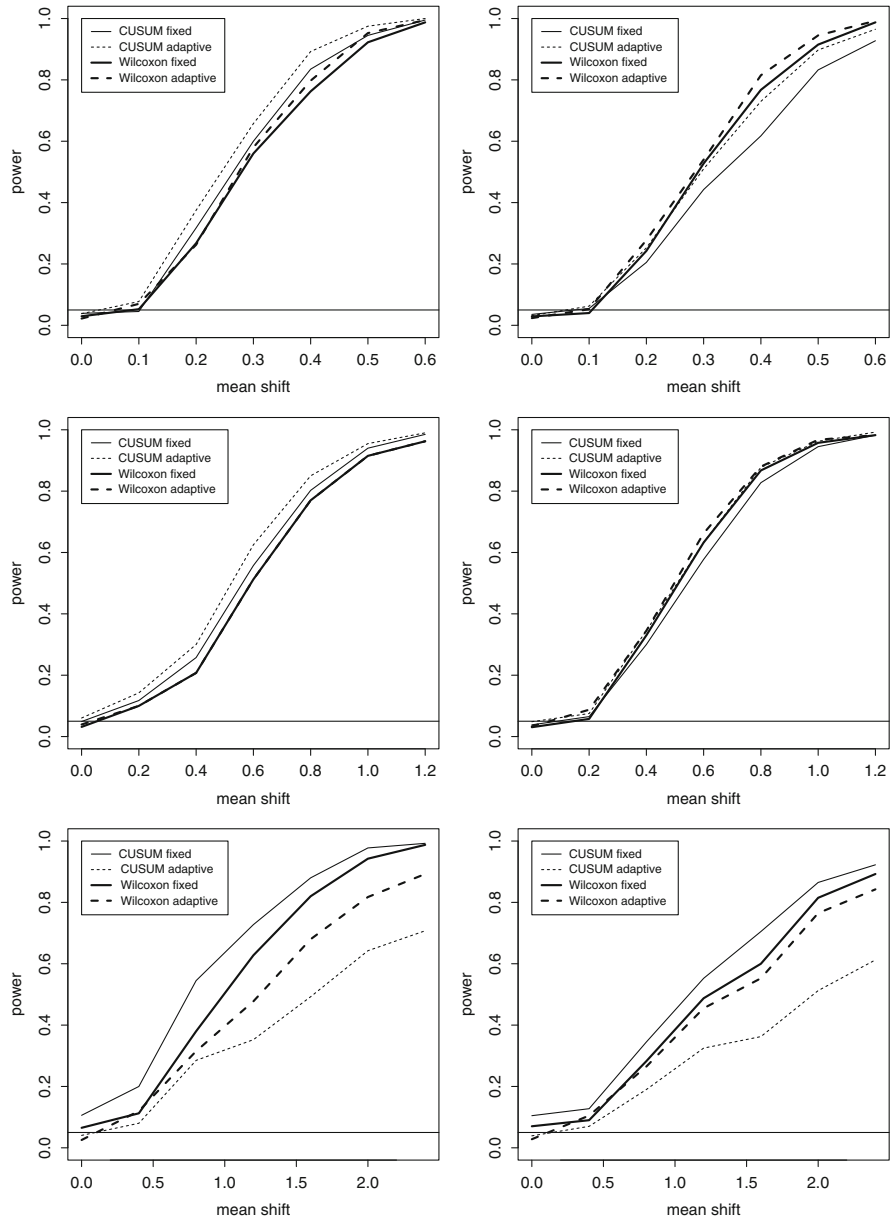


Fig. 1 Power of the tests in case of a shift in the middle of an AR(1) process with Gaussian (*left*) or t_3 -innovations (*right*) and different lag one correlations $\rho = 0.0$ (*top*), $\rho = 0.4$ (*middle*) or $\rho = 0.8$ (*bottom*), $n = 200$. Wilcoxon test $T_{n,1}$ (*bold lines*) and CUSUM test $T_{n,2}$ (*thin lines*). Adjustment by non-overlapping subsampling with fixed (*black*) or adaptive block length (*dashed*)

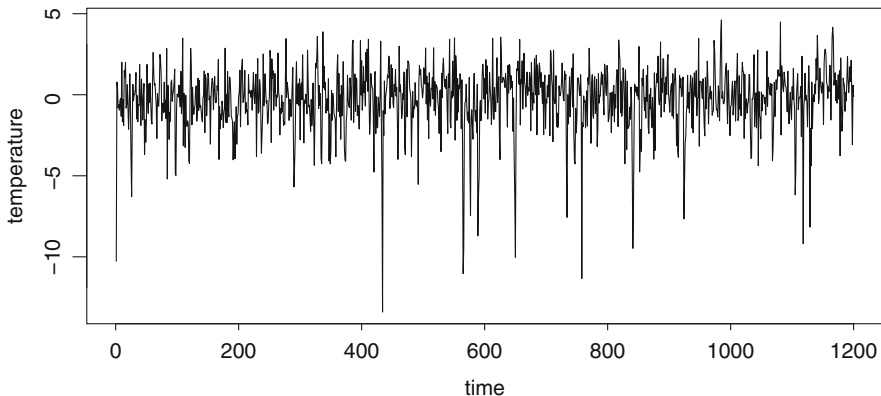


Fig. 2 Deseasonalized time series representing the monthly average daily minimum temperatures in Potsdam, Germany

As opposed to this, both versions of the Wilcoxon based test become significant as the corresponding p -values are 0.04 and 0.015, respectively. The differences between the results agree with the better power behavior of the Wilcoxon based test relatively to the CUSUM test in case of the (left-)skewed distributions of minimum temperatures, and the better power of the versions employing the adaptive block length over those with the fixed block length considered here in case of small positive autocorrelations. The sample median of the second time period is about 0.4 degrees larger than that of the first period.

6 Auxiliary Results

In this section, we will prove some auxiliary results which will play a crucial role in the proof of Theorem 1. The main result of this section is the following proposition, which essentially shows that the degenerate part in the Hoeffding decomposition of the U -statistic $T_n(\lambda)$ is uniformly negligible.

Proposition 1. *Let $(X_n)_{n \geq 1}$ be a 1-approximating functional with constants $(a_k)_{k \geq 1}$ of an absolutely regular process with mixing coefficients $(\beta(k))_{k \geq 1}$, satisfying*

$$\sum_{k=1}^{\infty} k(\beta(k) + \sqrt{a_k} + \phi(a_k)) < \infty. \tag{34}$$

Moreover, let $g(x, y)$ be a 1-continuous bounded degenerate kernel. Then, as $n \rightarrow \infty$,

$$\frac{1}{n^{3/2}} \sup_{0 \leq \lambda \leq 1} \left| \sum_{i=1}^{\lfloor n\lambda \rfloor} \sum_{j=\lfloor n\lambda \rfloor + 1}^n g(X_i, X_j) \right| \rightarrow 0 \tag{35}$$

in probability.

The proof of Proposition 1 requires some moment bounds for increments of U -statistics of degenerate kernels, which we will now state as separate lemmas.

Lemma 1. *Let $(X_n)_{n \geq 1}$ be a 1-approximating functional with constants $(a_k)_{k \geq 1}$ of an absolutely regular process with mixing coefficients $(\beta(k))_{k \geq 1}$, satisfying*

$$\sum_{k=1}^{\infty} k(\beta(k) + \sqrt{a_k} + \phi(a_k)) < \infty. \tag{36}$$

Moreover, let $g(x, y)$ be a 1-continuous bounded degenerate kernel. Then, there exists a constant C_1 such that

$$E \left(\sum_{i=1}^{\lfloor n\lambda \rfloor} \sum_{j=\lfloor n\lambda \rfloor+1}^n g(X_i, X_j) \right)^2 \leq C_1 \lfloor n\lambda \rfloor (n - \lfloor n\lambda \rfloor). \tag{37}$$

Proof. We can write

$$\begin{aligned} E \left(\sum_{i=1}^{\lfloor n\lambda \rfloor} \sum_{j=\lfloor n\lambda \rfloor+1}^n g(X_i, X_j) \right)^2 &= \sum_{i=1}^{\lfloor n\lambda \rfloor} \sum_{j=\lfloor n\lambda \rfloor+1}^n E(g(X_i, X_j))^2 \\ &+ 2 \sum_{1 \leq i_1 \neq i_2 \leq \lfloor n\lambda \rfloor} \sum_{\lfloor n\lambda \rfloor+1 \leq j_1 \neq j_2 \leq n} E(g(X_{i_1}, X_{j_1})g(X_{i_2}, X_{j_2})) \end{aligned} \tag{38}$$

The elements of the first sum all are bounded, hence

$$\sum_{i=1}^{\lfloor n\lambda \rfloor} \sum_{j=\lfloor n\lambda \rfloor+1}^n E(g(X_i, X_j))^2 \leq C \lfloor n\lambda \rfloor (n - \lfloor n\lambda \rfloor). \tag{39}$$

Concerning the second sum, by Lemma 5, we get

$$\begin{aligned} &\sum_{1 \leq i_1 < i_2 \leq \lfloor n\lambda \rfloor} \sum_{\lfloor n\lambda \rfloor+1 \leq j_1 < j_2 \leq n} E(g(X_{i_1}, X_{j_1})g(X_{i_2}, X_{j_2})) \\ &\leq 4S \sum_{1 \leq i_1 < i_2 \leq \lfloor n\lambda \rfloor} \sum_{\lfloor n\lambda \rfloor+1 \leq j_1 \leq j_2 \leq n} \phi(a_{\lfloor k/3 \rfloor}) \\ &+ 8S^2 \sum_{1 \leq i_1 < i_2 \leq \lfloor n\lambda \rfloor} \sum_{\lfloor n\lambda \rfloor+1 \leq j_1 \leq j_2 \leq n} (\sqrt{a_{\lfloor k/3 \rfloor}} + \beta(\lfloor k/3 \rfloor)) \end{aligned} \tag{40}$$

with $k = \max\{|i_2 - i_1|, |j_2 - j_1|\}$. We will first treat the summands with $k = i_2 - i_1$. Suppose for one moment that k is fixed and we will bound the number of indices that appear in the sum. Observe that in this case we have $\lfloor n\lambda \rfloor$ ways to choose i_1 ,

once i_1 is chosen we have one way to pick i_2 because $i_2 = i_1 + k$. For j_1 we have as before $n - [n\lambda]$ ways to pick this index and then for each j_1, j_2 need to be in the interval $[j_1, j_1 + k]$ and there are exactly k integers in such interval.

$$\begin{aligned} & \sum_{1 \leq i_1 < i_2 \leq [n\lambda]} \sum_{[n\lambda]+1 \leq j_1 < j_2 \leq n} (4S\phi(a_{[k/3]}) + 8S^2\sqrt{a_{[k/3]}} + 8S^2\beta([k/3])) \\ & \leq C[n\lambda](n - [n\lambda]) \left(\sum_{k=1}^n k\phi(a_k) + \sum_{k=1}^n k\sqrt{a_k} + \sum_{k=1}^n k\beta(k) \right) \leq C[n\lambda](n - [n\lambda]) \end{aligned} \tag{41}$$

Analogously we can find the bounds for the terms with $k = i_1 - i_2, k = j_2 - j_1$ and $k = j_1 - j_2$ using the conditions of summability.

We now define the process $G(\lambda), 0 \leq \lambda \leq 1$, by

$$G_n(\lambda) := n^{-3/2} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^n g(X_i, X_j), \quad 0 \leq \lambda \leq 1. \tag{42}$$

Lemma 2. *Under the conditions of Lemma 1, there exists a constant C such that*

$$E(|G_n(\eta) - G_n(\mu)|^2) \leq \frac{C}{n}(\eta - \mu), \tag{43}$$

for all $0 \leq \mu \leq \eta \leq 1$.

Proof. We can write

$$\begin{aligned} & E(|G_n(\eta) - G_n(\mu)|^2) \tag{44} \\ & \leq \frac{2}{n^3} E \left(\sum_{i=1}^{[n\mu]} \sum_{j=[n\mu]+1}^{[n\eta]} g(X_i, X_j) \right)^2 + \frac{2}{n^3} E \left(\sum_{i=[n\mu]+1}^{[n\eta]} \sum_{j=[n\eta]+1}^n g(X_i, X_j) \right)^2 \\ & = \frac{2}{n^3} E \left(\sum_{i=1}^{[n\mu]} \sum_{j=[n\mu]+1}^{[n\eta]} g(X_i, X_j) \right)^2 + \frac{2}{n^3} E \left(\sum_{i=1}^{[n\eta]-[n\mu]} \sum_{j=[n\eta]-[n\mu]+1}^{n-[n\mu]} g(X_i, X_j) \right)^2 \\ & \leq C \frac{1}{n^3} ([n\mu]([n\eta] - [n\mu]) + ([n\eta] - [n\mu])(n - [n\eta])) \leq \frac{C}{n}(\eta - \mu) \end{aligned}$$

using the stationarity of the process $(X_n)_{n \in \mathbb{N}}$ and Lemma 1.

Proof of Proposition 1. From Lemma 2 we obtain, using Chebyshev’s inequality,

$$P(|G_n(\eta) - G_n(\mu)| \geq \epsilon) \leq \frac{1}{\epsilon^2} \frac{C}{n} (\eta - \mu), \tag{45}$$

for all $\epsilon > 0$. Thus we get for $0 \leq k \leq m \leq n$ with $k, m, n \in \mathbb{N}$

$$\begin{aligned} P\left(\left|G_n\left(\frac{m}{n}\right) - G_n\left(\frac{k}{n}\right)\right| \geq \epsilon\right) &\leq \frac{1}{\epsilon^2} E\left(G_n\left(\frac{m}{n}\right) - G_n\left(\frac{k}{n}\right)\right)^2 \\ &\leq \frac{1}{\epsilon^2} \frac{C}{n^2} (m - k) \leq \frac{1}{\epsilon^2} \frac{C}{n^{5/3}} (m - k)^{4/3} \end{aligned} \tag{46}$$

as $m - k \leq n$. Now consider the variables

$$\zeta_i = \begin{cases} G_n\left(\frac{i}{n}\right) - G_n\left(\frac{i-1}{n}\right) & \text{if } i = 1, \dots, n-1 \\ 0 & \text{else} \end{cases} \tag{47}$$

and suppose that $S_i = \zeta_1 + \zeta_2 + \dots + \zeta_i$ with $S_0 = 0$, then $S_i = G_n\left(\frac{i}{n}\right)$. In consequence the inequality (46) is equivalent to

$$P(|S_m - S_k| \geq \epsilon) \leq \frac{1}{\epsilon^2} \left[\frac{C^{3/4}}{n^{5/4}} (m - k) \right]^{4/3} \quad \text{for } 0 \leq k \leq m \leq n. \tag{48}$$

So the assumption of Theorem 7 are satisfied with the variables (47) in the role of the ξ_i , $\beta = 1/2$, $\alpha = 2/3$ and $u_l = C^{3/4}/n^{5/4}$, $u_o = 0$ and hence

$$P\left(\max_{1 \leq i \leq n-1} |S_i| \geq \epsilon\right) \leq \frac{K}{\epsilon^2} \left[\frac{C^{3/4}}{n^{5/4}} (n - 1) \right]^{4/3} \leq \frac{KC}{\epsilon^2 n^{1/3}} \tag{49}$$

where K depends only of α and β . Thus, (35) holds as $n \rightarrow \infty$. □

7 Proof of Main Results

In this section, we will prove Theorems 1 and 2. Note that Theorem 2 is a direct consequence of Theorem 1, applied to anti-symmetric kernels. We will nevertheless present a direct proof of Theorem 2, since this proof is much simpler than the proof in the general case. Moreover, Theorem 2 covers those cases that are most relevant in applications.

The first part of the proof is identical for both Theorems 1 and 2. Note that, for each $\lambda \in [0, 1]$, the statistic $T_n(\lambda)$ is a two-sample U -statistic. Thus, using the Hoeffding decomposition (11), we can write $T_n(\lambda)$ as

$$\begin{aligned}
 T_n(\lambda) &= \frac{1}{n^{3/2}} \left(\sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n (h_1(X_i) + h_2(X_j) + g(X_i, X_j)) \right) \\
 &= \frac{1}{n^{3/2}} \left((n - [\lambda n]) \sum_{i=1}^{[\lambda n]} h_1(X_i) + [\lambda n] \sum_{j=[\lambda n]+1}^n h_2(X_j) + \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n g(X_i, X_j) \right)
 \end{aligned} \tag{50}$$

By Proposition 1, we know that

$$\frac{1}{n^{3/2}} \sup_{0 \leq \lambda \leq 1} \left| \sum_{i=1}^{[\lambda n]} \sum_{j=[\lambda n]+1}^n g(X_i, X_j) \right| \rightarrow 0$$

in probability. Thus, by Slutsky’s lemma, it suffices to show that the sum of the first two terms, i.e.

$$\left(\frac{n - [\lambda n]}{n^{3/2}} \sum_{i=1}^{[\lambda n]} h_1(X_i) + \frac{[\lambda n]}{n^{3/2}} \sum_{j=[\lambda n]+1}^n h_2(X_j) \right)_{0 \leq \lambda \leq 1} \tag{51}$$

converges in distribution to the desired limit process.

Proof of Theorem 2. It remains to show that (51) converges in distribution to $\sigma W^{(0)}(\lambda)$, $0 \leq \lambda \leq 1$, where $(W^{(0)}(\lambda))_{0 \leq \lambda \leq 1}$ is standard Brownian bridge on $[0, 1]$, and where σ^2 is defined in (19). By antisymmetry of the kernel $h(x, y)$, we obtain that $h_2(x) = -h_1(x)$. Hence, in this case, (51) can be rewritten as

$$\frac{n - [\lambda n]}{n^{3/2}} \sum_{i=1}^{[\lambda n]} h_1(X_i) - \frac{[\lambda n]}{n^{3/2}} \sum_{i=[\lambda n]+1}^n h_1(X_i) = \frac{1}{n^{1/2}} \sum_{i=1}^{[\lambda n]} h_1(X_i) - \frac{[\lambda n]}{n^{3/2}} \sum_{i=1}^n h_1(X_i).$$

By Proposition 2.11 and Lemma 2.15 of Borovkova, Burton and Dehling [3], the sequence $(h_1(X_i))_{i \geq 1}$ is a 1-approximating functional with approximating constant $C\sqrt{a_k}$. Since $h_1(X_i)$ is bounded, the L_2 -near epoch dependence in the sense of Wooldridge and White [21] also holds, with the same constants. Moreover, the underlying process $(Z_n)_{n \geq 1}$ is absolutely regular, and hence also strongly mixing. Thus we may apply the invariance principle in Corollary 3.2 of Wooldridge and White [21], and obtain that the partial sum process

$$\left(\frac{1}{n^{1/2}} \sum_{i=1}^{[\lambda n]} h_1(X_i) \right)_{0 \leq \lambda \leq 1} \tag{52}$$

converges weakly to Brownian motion $(W(\lambda))_{0 \leq \lambda \leq 1}$ with $\text{Var}(W(1)) = \sigma^2$. The statement of the Theorem follows with the continuous mapping theorem for the mapping $x(t) \mapsto x(t) - tx(1)$, $0 \leq t \leq 1$.

The proof of Theorem 1 requires an invariance principle for the partial sum process of \mathbb{R}^2 -valued dependent random variables; see Proposition 2 below. For mixing processes, such invariance principles have been established even for partial sums of Hilbert space valued random vector, e.g. by Dehling [7]. In this paper, we provide an extension of these results to functionals of mixing processes.

Proposition 2. *Let $(X_n)_{n \in \mathbb{N}}$ be a 1-approximating functional of an absolutely regular process with mixing coefficients $(\beta(k))$ and let $h_1(\cdot), h_2(\cdot)$ be bounded1-continuous functions with mean zero, such that*

$$\sum_k k^2(\beta(k) + a_k + \phi(a_k)) < \infty. \tag{53}$$

Then, as $n \rightarrow \infty$,

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \begin{pmatrix} h_1(X_i) \\ h_2(X_i) \end{pmatrix} \right)_{0 \leq t \leq 1} \rightarrow \begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix}_{0 \leq t \leq 1} \tag{54}$$

where $(W_1(t), W_2(t))_{0 \leq t \leq 1}$ is a two-dimensional Brownian motion with mean zero and covariance $E(W_k(s) W_l(t)) = \min(s, t)\sigma_{kl}$, where $\sigma_{k,l}$ as defined in (16).

Proof. To prove (54), we need to establish finite dimensional convergence and tightness. Concerning finite-dimensional convergence, by the Cramér-Wold device it suffices to show the convergence in distribution of a linear combination of the coordinates of the vector

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{[nt_1]} h_1(X_i), \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt_1]} h_2(X_i), \dots, \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt_j]} h_1(X_i), \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt_j]} h_2(X_i), \dots, \frac{1}{\sqrt{n}} \sum_{i=1}^n h_1(X_i), \frac{1}{\sqrt{n}} \sum_{i=1}^n h_2(X_i) \right), \tag{55}$$

for $0 = t_0 < t_1 < \dots < t_j < \dots < t_k = 1$. Any such linear combination can be expressed as

$$\sum_{j=1}^k \frac{1}{\sqrt{n}} \sum_{i=[nt_{j-1}]+1}^{[nt_j]} (a_j h_1(X_i) + b_j h_2(X_i)), \tag{56}$$

for $(a_j, b_j)_{j=1}^k \in \mathbb{R}^{2k}$. By using the Cramér-Wold device again, the weak convergence of this sum is equivalent to the weak convergence of the vector

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt_1 \rfloor} (a_1 h_1(X_i) + b_1 h_2(X_i)), \dots, \frac{1}{\sqrt{n}} \sum_{i=\lfloor nt_{j-1} \rfloor + 1}^{\lfloor nt_j \rfloor} (a_j h_1(X_i) + b_j h_2(X_i)), \dots, \frac{1}{\sqrt{n}} \sum_{i=\lfloor nt_{k-1} \rfloor + 1}^n (a_k h_1(X_i) + b_k h_2(X_i)) \right) \quad (57)$$

to

$$(a_1(W_1(t_1) - W_1(t_0)) + b_1(W_2(t_1) - W_2(t_0)), \dots, a_k(W_1(t_k) - W_1(t_{k-1})) + b_k(W_2(t_k) - W_2(t_{k-1}))). \quad (58)$$

Since $(X_n)_{n \geq 1}$ is a 1-approximating functional, it can be coupled with a process consisting of independent blocks. Given integers $L := L_n = \lceil n^{3/4} \rceil$ and $l_n = \lceil n^{1/2} \rceil$, we introduce the (l, L) blocking $(B_m)_{m \geq 0}$ of the variables $(a_j h_1(X_i) + b_j h_2(X_i))$ with $i = \lfloor nt_{j-1} \rfloor + 1, \dots, \lfloor nt_j \rfloor, j = 0, \dots, k$ and

$$B_m := \sum_{i=(m-1)(L_n+l_n)+1}^{m(L_n+(m-1)l_n)} (a_j h_1(X_i) + b_j h_2(X_i)) \quad (59)$$

and separating blocks

$$\tilde{B}_m := \sum_{i=mL_n+(m-1)l_n+1}^{m(L_n+l_n)} (a_j h_1(X_i) + b_j h_2(X_i)). \quad (60)$$

By Theorem 5 there exists a sequence of independent blocks (B'_m) with the same blockwise marginal distribution as (B_m) and such that

$$P(|B_m - B'_m| \leq 2\alpha_l) \geq 1 - \beta(l) - 2\alpha_l,$$

where $\alpha_l := (2 \sum_{k=\lfloor l_n/3 \rfloor}^{\infty} a_k)^{1/2}$. We can express the components of our vector (57) as a sum of blocks

$$\begin{aligned} & \sum_{i=\lfloor nt_j \rfloor + 1}^{\lfloor nt_{j+1} \rfloor} (a_j h_1(X_i) + b_j h_2(X_i)) \\ &= \sum_{m=\lfloor \frac{nt_j}{L+l} \rfloor + 1}^{\lfloor \frac{nt_{j+1}}{L+l} \rfloor} B_m + \sum_{m=\lfloor \frac{nt_j}{L+l} \rfloor + 1}^{\lfloor \frac{nt_{j+1}}{L+l} \rfloor} \tilde{B}_m + \sum_{R_j} (a_j h_1(X_i) + b_j h_2(X_i)), \quad (61) \end{aligned}$$

where R_j denotes the set of indices not contained in the blocks. Observe that by the Lemma 3 for any set $A \subset \{1, \dots, n\}$

$$E \left(\sum_{i \in A} (a_j h_1(X_i) + b_j h_2(X_i)) \right)^2 \leq C \#A \tag{62}$$

and hence

$$E \left(\sum_{m=\lfloor \frac{n_j}{L+l} \rfloor + 1}^{\lfloor \frac{n_j+1}{L+l} \rfloor} \tilde{B}_m \right)^2 \leq C \frac{n}{L_n + l_n} l_n \leq C n^{3/4}, \tag{63}$$

so it follows with the Chebyshev inequality that this term is negligible. For the last summand, we have that

$$E \left(\sum_{R_j} (a_j h_1(X_i) + b_j h_2(X_i)) \right)^2 \leq C 2(L_n + l_n) \leq C n^{3/4}. \tag{64}$$

Furthermore, we need to show that we can replace the blocks B_m by the independent coupled blocks B'_m :

$$\begin{aligned} P \left(\left| \frac{1}{\sqrt{n}} \sum_{m=\lfloor \frac{n_j}{L+l} \rfloor + 1}^{\lfloor \frac{n_j+1}{L+l} \rfloor} (B_m - B'_m) \right| > \epsilon \right) &\leq \sum_{m=\lfloor \frac{n_j}{L+l} \rfloor + 1}^{\lfloor \frac{n_j+1}{L+l} \rfloor} P \left(|B_m - B'_m| > \frac{\epsilon \sqrt{n}}{n^{1/4}} \right) \\ &\leq n^{\frac{1}{4}} \left(\beta \left(\lfloor \frac{l_n}{3} \rfloor \right) + \alpha_{\lfloor \frac{l_n}{3} \rfloor} \right) \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$ by our conditions on the mixing coefficients and approximation constants. Here we used that fact that $\alpha_n \rightarrow 0$ and thus, for almost all $n \in \mathbb{N}$,

$$P \left(|B_m - B'_m| > \epsilon n^{1/4} \right) \leq P \left(|B_m - B'_m| > 2\alpha_{l_n} \right). \tag{65}$$

With the above arguments the result holds if we show the convergence of

$$\frac{1}{\sqrt{n}} \left(\sum_{m=\lfloor \frac{n_0}{L+l} \rfloor + 1}^{\lfloor \frac{n_1}{L+l} \rfloor} B'_m, \dots, \sum_{m=\lfloor \frac{n_k}{L+l} \rfloor + 1}^{\lfloor \frac{n_{k+1}}{L+l} \rfloor} B'_m \right). \tag{66}$$

Since this vector has independent components, we only need to show the one-dimensional convergence, which is a consequence of Theorem 4, using the summability condition (53).

We now turn to the question of tightness and show that, for each ϵ and η , there exist a δ , $0 < \delta < 1$, and an integer n_0 such that, for $0 \leq t \leq 1$,

$$\frac{1}{\delta} P \left(\sup_{t \leq s \leq t+\delta} |Y_n(s) - Y_n(t)| \geq \epsilon \right) \leq \eta, \quad n \geq n_0 \tag{67}$$

with

$$Y_n(t) = \frac{1}{\sigma \sqrt{n}} \sum_{i=1}^{[nt]} h_1(X_i) + (nt - [nt]) \frac{1}{\sigma \sqrt{n}} h(X_{[nt]+1}) \tag{68}$$

(h_2 can be treated in the same way) and by Theorem 8, this condition reduces to: For each positive ϵ there exist a $\alpha > 1$ and an integer n_0 , s. t.

$$P \left(\max_{i \leq n} \left| \sum_{j=1}^i h_1(X_j) \right| \geq \lambda \sqrt{n} \right) \leq \frac{\epsilon}{\lambda^2}, \quad n \geq n_0. \tag{69}$$

Let $t \geq s, s, t \in [0, 1]$. By Lemma 4 we get

$$\begin{aligned} E \left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} h_1(X_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^{[ns]} h_1(X_i) \right|^4 \right) &= \frac{1}{n^2} E \left(\sum_{i=[ns]+1}^{[nt]} h_1(X_i) \right)^4 \\ &\leq \frac{1}{n^2} (([nt] - [ns])^2 C) \end{aligned} \tag{70}$$

and this implies

$$P \left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^m h_1(X_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^k h_1(X_i) \right| \geq \epsilon \right) \leq \frac{1}{\epsilon^4} \left(\frac{C^{1/2}}{n} (m - k) \right)^2. \tag{71}$$

By Theorem 7

$$P \left(\max_{i \leq n} \left| \sum_{j=1}^i h_1(X_j) \right| \geq \epsilon \sqrt{n} \right) \leq \frac{K}{\epsilon^4} \left(\frac{C^{1/2}}{n} (n - 1) \right)^2 \tag{72}$$

and we get the assertion. Thus we have established tightness of each of the two coordinates of the partial sum process, which implies tightness of the vector-valued process.

Proof of Theorem 1. From Proposition 2 we obtain that

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{[n\lambda]} \begin{pmatrix} h_1(X_i) \\ h_2(X_i) \end{pmatrix} \right)_{0 \leq \lambda \leq 1} \longrightarrow \begin{pmatrix} W_1(\lambda) \\ W_2(\lambda) \end{pmatrix}_{0 \leq \lambda \leq 1}, \tag{73}$$

in distribution on the space $(D([0, 1]))^2$. We consider the functional given by

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \mapsto (1 - t)x_1(t) + t(x_2(1) - x_2(t)), \quad 0 \leq t \leq 1. \tag{74}$$

This is a continuous mapping from $(D[0, 1])^2$ to $D[0, 1]$, so we may apply the continuous mapping theorem to (73), and obtain

$$\begin{aligned} & \left(\frac{n - [n\lambda]}{n^{3/2}} \sum_{i=1}^{[n\lambda]} h_1(X_i) + \frac{[n\lambda]}{n^{3/2}} \sum_{j=[n\lambda]+1}^n h_2(X_j) \right)_{0 \leq \lambda \leq 1} \\ & \longrightarrow ((1 - \lambda)W_1(\lambda) + \lambda(W_2(1) - W_2(\lambda)))_{0 \leq \lambda \leq 1}. \end{aligned}$$

Together with the remarks at the beginning of this section, this proves Theorem 1.

Appendix: Some Auxiliary Results from the Literature

In this section, we collect some known lemmas and theorems for weakly dependent data. We start with some results on the behaviour of partials sums:

Lemma 3 (Borovkova, Burton, Dehling [3], Lemma 2.23). *Let $(X_k)_{k \in \mathbb{Z}}$ be a 1-approximating functional with constants $(a_k)_{k \geq 0}$ of an absolutely regular process with mixing coefficients $(\beta(k))_{k \geq 0}$. Suppose moreover that $EX_i = 0$ and that one of the following two conditions holds:*

1. X_0 is bounded a.s. and $\sum_{k=0}^{\infty} (a_k + \beta(k)) < \infty$.
2. $E|X_0|^{2+\delta} < \infty$ and $\sum_{k=0}^{\infty} (a_k^{\frac{\delta}{1+\delta}} + \beta^{\frac{\delta}{1+\delta}}(k)) < \infty$.

Then, as $N \rightarrow \infty$,

$$\frac{1}{N} ES_N^2 \rightarrow EX_0^2 + 2 \sum_{j=1}^{\infty} E(X_0 X_j) \tag{75}$$

and the sum on the r.h.s. converges absolutely.

Lemma 4 (Borovkova, Burton, Dehling [3], Lemma 2.24). *Let $(X_k)_{k \in \mathbb{Z}}$ be a 1-approximating functional with constants (a_k) of an absolutely regular process with mixing coefficients $(\beta(k))_{k \geq 0}$. Suppose moreover that $EX_i = 0$ and that one of the following two conditions holds:*

1. X_0 is bounded a.s. and $\sum_{k=0}^{\infty} k^2(a_k + \beta(k)) < \infty$.
2. $E|X_0|^{4+\delta} < \infty$ and $\sum_{k=0}^{\infty} k^2(a_k^{\frac{\delta}{3+\delta}} + \beta^{\frac{\delta}{4+\delta}}(k)) < \infty$.

Then there exists a constant C such that

$$ES_N^4 \leq CN^2. \quad (76)$$

Theorem 4 (Borovkova, Burton, Dehling [3], Theorem 4). *Let $(X_k)_{k \in \mathbb{Z}}$ be a 1-approximating functional with constants $(a_k)_{k \geq 0}$ of an absolutely regular process with mixing coefficients $(\beta(k))_{k \geq 0}$. Suppose moreover that $EX_i = 0$, $E|X_0|^{4+\delta} < \infty$ and that*

$$\sum_{k=0}^{\infty} k^2(a_k^{\frac{\delta}{3+\delta}} + \beta^{\frac{\delta}{4+\delta}}(k)) < \infty, \quad (77)$$

for some $\delta > 0$. Then, as $n \rightarrow \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow \mathcal{N}(0, \sigma^2), \quad (78)$$

where $\sigma^2 = EX_0^2 + 2 \sum_{j=1}^{\infty} E(X_0 X_j)$. In case $\sigma^2 = 0$, $\mathcal{N}(0, 0)$ denotes the point mass at the origin. If X_0 is bounded, the CLT continues to hold if (77) is replaced by the condition that $\sum_{k=0}^{\infty} k^2(a_k + \beta(k)) < \infty$.

An important tool to derive asymptotic results for weakly dependent data are coupling methods. We will apply this method in the proof of Proposition 2.

Theorem 5 (Borovkova, Burton, Dehling [3], Theorem 3). *Let $(X_n)_{n \in \mathbb{N}}$ be a 1-approximating functional with summable constants $(a_k)_{k \geq 0}$ of an absolutely regular process with mixing rate $(\beta(k))_{k \geq 0}$. Then given integers K, L and N , we can approximate the sequence of $(K + 2L, N)$ -blocks $(B_s)_{s \geq 1}$ by a sequence of independent blocks $(B'_s)_{s \geq 1}$ with the same marginal distribution in such a way that*

$$P(\|B_s - B'_s\| \leq 2\alpha_L) \geq 1 - \beta(K) - 2\alpha_L, \quad (79)$$

where $\alpha_L := (2 \sum_{l=L}^{\infty} a_l)^{1/2}$.

In statistical application, the question of how to estimate σ^2 is important. In the situation when the observations are a functional of α -mixing process, Dehling et al. [10] propose the estimation of the variance of partial sums of dependent processes by the subsampling estimator

$$\hat{D}_n = \frac{1}{[n/l_n]} \sqrt{\frac{\pi}{2}} \sum_{i=1}^{[n/l_n]} \frac{|\hat{T}_i(l_n) - l_n \tilde{U}_n|}{\sqrt{l_n}} \tag{80}$$

with $\hat{T}_i(l) = \sum_{j=(i-1)l+1}^{il} F_n(X_j)$ and $\tilde{U}_n = \frac{1}{n} \sum_{j=1}^n F_n(X_j)$, where $F_n(\cdot)$ is the empirical distribution function.

Theorem 6 (Dehling, Fried, Sharipov, Vogel, Wornowizki [9], Theorem 1.2). *Let $(X_k)_{k \geq 1}$ be a stationary, 1-approximating functional of an α -mixing processes. Suppose that for some $\delta > 0$, $E|X_1|^{2+\delta} < \infty$, and that the mixing coefficients $(\alpha_k)_{k \geq 1}$ and the approximation constants $(a_k)_{k \geq 1}$ satisfy*

$$\sum_{k=1}^{\infty} (\alpha_k)^{\frac{2}{2+\delta}} < \infty, \quad \sum_{k=1}^{\infty} (a_k)^{\frac{1+\delta}{2+\delta}} < \infty. \tag{81}$$

In addition, we assume that F is Lipschitz-continuous, that $\alpha_k = O(n^{-8})$ and that $a_m = O(m^{-12})$. Then, as $n \rightarrow \infty$, $l_n \rightarrow \infty$ and $l_n = o(\sqrt{n})$, we have $\hat{D}_n \rightarrow \sigma$ in L_2 .

To deal with the degenerate kernel g , we need to find upper bounds for the expectations $E(g(X_{i_1}, X_{j_1})g(X_{i_2}, X_{j_2}))$, in terms of the maximal distance among the indices. Since $1 \leq i_1 < i_2 \leq [n\lambda]$ and $[n\lambda] + 1 \leq j_1 < j_2 \leq n$, we get $i_1 < i_2 < j_1 < j_2$.

Lemma 5 (Dehling, Fried [8], Proposition 6.1). *Let $(X_n)_{n \geq 1}$ be a 1-approximating functional with constants $(a_k)_{k \geq 1}$ of an absolutely regular process with mixing coefficients $(\beta(k))_{k \geq 1}$ and let $g(x, y)$ be a 1-continuous bounded degenerate kernel. Then we have*

$$|E(g(X_{i_1}, X_{j_1})g(X_{i_2}, X_{j_2}))| \leq 4S\phi(a_{[k/3]}) + 8S^2(\sqrt{a_{[k/3]}} + \beta([k/3])) \tag{82}$$

where $S = |\sup_{x,y} g(x, y)|$ and $k = \max \{i_2 - i_1, j_1 - i_2, j_2 - j_1\}$.

The following two results are useful for proving tightness of a stochastic process. The first one is used to control the fluctuation of maximum. Let ξ_1, \dots, ξ_n be random variables, and define $S_k = \xi_1 + \dots + \xi_k$ ($S_0 = 0$), and $M_n = \max_{0 \leq k \leq n} |S_k|$.

Theorem 7 (Billingsley [2], Theorem 10.2). *Suppose that $\beta \geq 0$ and $\alpha > 1/2$ and that there exist nonnegative numbers u_1, \dots, u_n such that for all positive λ*

$$P(|S_j - S_i| \geq \lambda) \leq \frac{1}{\lambda^{4\beta}} \left(\sum_{i < l \leq j} u_l \right)^{2\alpha}, \quad 0 \leq i \leq j \leq n, \tag{83}$$

then for all positive λ

$$P(M_n \geq \lambda) \leq \frac{K_{\beta,\alpha}}{\lambda^{4\beta}} \left(\sum_{0 < l \leq n} u_l \right)^{2\alpha}, \quad (84)$$

where $K_{\beta,\alpha}$ is a constant depending only on β and α .

Theorem 8 (Billingsley [2], Theorem 8.4). *The sequence $\{Y_n\}$, defined by*

$$Y_n(t) = \frac{1}{\sigma\sqrt{n}} S_{[nt]} + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1} \quad (85)$$

is tight if for each $\epsilon > 0$ there exist a $\lambda > 1$ and a $n_0 \in \mathbb{N}$ such that for $n \geq n_0$

$$P\left(\max_{i \leq n} |S_{k+i} - S_k| \geq \lambda\sigma\sqrt{n}\right) \leq \frac{\epsilon}{\lambda^2}. \quad (86)$$

Acknowledgements The authors wish to thank the referees for their very careful reading of an earlier version of this manuscript, and for their many thoughtful comments that helped to improve the presentation of the paper. This research was supported by the Collaborative Research Center 823, Project C3 *Analysis of Structural Change in Dynamic Processes*, of the German Research Foundation DFG.

References

1. Babel, B.: Invariance principles for U -statistics and von Mises functionals. *J. Stat. Plan. Inference* **22**, 337–354 (1989)
2. Billingsley, P.: *Convergence of Probability Measures*. 2nd edn. Wiley, New York (1999)
3. Borovkova, S.A., Burton, R.M., Dehling, H.G.: Limit theorems for functionals of mixing processes with applications to U -statistics and dimension estimation. *Trans. Am. Math. Soc.* **353**, 4261–4318 (2001)
4. Carlstein, E.: The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Stat.* **14**, 1171–1179 (1986)
5. Csörgő, M., Horvath, L.: Invariance principles for changepoint problems. *J. Multivar. Anal.* **27**, 151–168 (1988)
6. Csörgő, M., Horvath, L.: *Limit Theorems in Change Point Analysis*. Wiley, New York (1997)
7. Dehling, H.: Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. für Wahrscheinlichkeitstheorie und verwandte Gebiete* **63**, 393–432 (1983)
8. Dehling, H., Fried, R.: Asymptotic distribution of two-sample empirical U -quantiles with applications to robust tests for shifts in location. *J. Multivar. Anal.* **105**, 124–140 (2012)
9. Dehling, H., Fried, R., Sharipov, O.Sh., Vogel, D., Wornowizki, M.: Estimation of the variance of partial sums of dependent processes. *Stat. Probab. Lett.* **83**, 141–147 (2013)
10. Dehling, H., Rooch, A., Taqqu, M.S.: Nonparametric change-point tests for long-range dependent data. *Scand. J. Stat.* **40**, 153–173 (2013)
11. Denker, M.: *Asymptotic Distribution Theory in Nonparametrics Statistics*. Vieweg Verlag, Braunschweig/Wiesbaden (1985)

12. Denker, M., Keller, G.: On U -statistics and v. Mises' statistics for weakly dependent processes. *Z. für Wahrscheinlichkeitstheorie und verwandte Gebiete* **64**, 505–522 (1983)
13. Denker, M., Keller, G.: Rigorous statistical procedures for data from dynamical systems. *J. Stat. Phys.* **44**, 67–93 (1986)
14. Hoeffding, W.: A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19**, 293–325 (1948)
15. Ibragimov, I.A., Linnik, Yu.V.: *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen (1971)
16. Leucht, A.: Degenerate U - and V -statistics under weak dependence: asymptotic theory and bootstrap consistency. *Bernoulli* **18**, 552–585 (2012)
17. Rooch, A.: *Change-point tests for long-range dependent data*. Dissertation, Ruhr-Universität Bochum (2012)
18. Sen, P.K.: On the properties of U -statistics when the observations are not independent. I. Estimation of non-serial parameters in some stationary stochastic processes. *Calcutta Stat. Assoc. Bull.* **12**, 69–92 (1963)
19. Sen, P.K.: Limiting behavior of regular functionals of empirical distributions for stationary $*$ -mixing processes. *Z. für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **25**, 71–82 (1972)
20. Sharipov, O.Sh., Wendler, M.: Bootstrap for the sample mean and for U -statistics of mixing and near-epoch dependent processes. *J. Nonparametric Stat.* **24**, 317–342 (2012)
21. Wooldridge, J.M., White, H.: Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econ. Theory* **4**, 210–230 (1988)
22. Yoshihara, K.-I.: Limiting behavior of U -statistics for stationary, absolutely regular processes. *Z. für Wahrscheinlichkeitstheorie und verwandte Gebiete* **35**, 237–252 (1976)